



# Genomic signatures of relaxed disruptive selection associated with speciation reversal in whitefish

Hudson *et al.*

RESEARCH ARTICLE

Open Access

# Genomic signatures of relaxed disruptive selection associated with speciation reversal in whitefish

Alan G Hudson<sup>1,2,3\*</sup>, Pascal Vonlanthen<sup>1,2</sup>, Etienne Bezault<sup>1,2,4</sup> and Ole Seehausen<sup>1,2</sup>

## Abstract

**Background:** Speciation reversal: the erosion of species differentiation via an increase in introgressive hybridization due to the weakening of previously divergent selection regimes, is thought to be an important, yet poorly understood, driver of biodiversity loss. Our study system, the Alpine whitefish (*Coregonus* spp.) species complex is a classic example of a recent postglacial adaptive radiation: forming an array of endemic lake flocks, with the independent origination of similar ecotypes among flocks. However, many of the lakes of the Alpine radiation have been seriously impacted by anthropogenic nutrient enrichment, resulting in a collapse in neutral genetic and phenotypic differentiation within the most polluted lakes. Here we investigate the effects of eutrophication on the selective forces that have shaped this radiation, using population genomics. We studied eight sympatric species assemblages belonging to five independent parallel adaptive radiations, and one species pair in secondary contact. We used AFLP markers, and applied  $F_{ST}$  outlier (BAYESCAN, DFDIST) and logistic regression analyses (MATSAM), to identify candidate regions for disruptive selection in the genome and their associations with adaptive traits within each lake flock. The number of outlier and adaptive trait associated loci identified per lake were then regressed against two variables (historical phosphorus concentration and contemporary oxygen concentration) representing the strength of eutrophication.

**Results:** Whilst we identify disruptive selection candidate regions in all lake flocks, we find similar trends, across analysis methods, towards fewer disruptive selection candidate regions and fewer adaptive trait/candidate loci associations in the more polluted lakes.

**Conclusions:** Weakened disruptive selection and a concomitant breakdown in reproductive isolating mechanisms in more polluted lakes has led to increased gene flow between coexisting Alpine whitefish species. We hypothesize that the resulting higher rates of interspecific recombination reduce either the number or extent of genomic islands of divergence surrounding loci evolving under disruptive natural selection. This produces the negative trend seen in the number of selection candidate loci recovered during genome scans of whitefish species flocks, with increasing levels of anthropogenic eutrophication: as the likelihood decreases that AFLP restriction sites will fall within regions of heightened genomic divergence and therefore be classified as  $F_{ST}$  outlier loci. This study explores for the first time the potential effects of human-mediated relaxation of disruptive selection on heterogeneous genomic divergence between coexisting species.

**Keywords:** *Coregonus*, Population genomics, Adaptive radiation, Speciation reversal, Relaxed selection, Heterogeneous genomic divergence, Ecological speciation, Speciation-with-gene-flow, Eutrophication, Divergence hitchhiking

\* Correspondence: alan.g.hudson@gmail.com

<sup>1</sup>Division of Aquatic Ecology & Macroevolution, Institute of Ecology & Evolution, University of Bern, Baltzerstrasse 6, CH 3012 Bern, Switzerland

<sup>2</sup>Department of Fish Ecology & Evolution, EAWAG Centre of Ecology, Evolution and Biogeochemistry, Seestrasse 79, CH 6047 Kastanienbaum, Switzerland

Full list of author information is available at the end of the article

## Background

Understanding the evolutionary processes that control the diversification, coexistence and persistence of species has taken on a heightened urgency as human-induced environmental changes and stressors have increased in prevalence and the augurs of a new, anthropogenic mass extinction become more evident [1-3]. Changes in abiotic and biotic environmental conditions will alter the direction, strength and form of selection that generates and maintains species differences and species coexistence, particularly among recently diverged species. How populations of such species respond to environmental change will depend on the severity, rate and duration of the change and the associated selection pressures experienced. Negative population growth is just one way species can go extinct; another scenario for species loss in the face of environmental change is speciation reversal [4]. Here, a reduction in environmental heterogeneity increases gene-flow and/or weakens disruptive selection, resulting in a break down in pre-zygotic and/or extrinsic post-zygotic reproductive isolation: leading to the fusion of previously distinct species through introgressive hybridization [5-8].

The effects of human-induced habitat alteration on the levels of intra- and interspecific diversity within adaptive radiations have generally been studied through the measurement of changes in the frequencies and distributions of neutral genetic markers and/or putative adaptive trait values among and within the affected species. Examples from vertebrate radiations indicate a weakening of disruptive selection during the course of environmental perturbations and concomitantly an increase in hybridization between coexisting species, as the distribution of feeding resources and/or habitats for breeding becomes homogenized and prezygotic and extrinsic postzygotic reproductive isolation weakens. This process is accompanied by the complete or near complete fusion of previously discrete gene pools into single, albeit genetically diverse populations, and trait distributions losing multimodality [5,6,9].

During ecological speciation gene flow between diverging populations is expected to be restricted mainly at loci underpinning traits involved in divergent adaptation and areas of the genome in physical linkage with these loci; so called "genomic islands of divergence" [10,11]. Among recently diverged species, the balance between disruptive selection and homogenizing gene flow will determine the extent of genetic differentiation achieved and its genomic distribution [12,13]. This heterogeneous genomic divergence will be at its most extreme during moderate to high gene flow and strong exogenous disruptive selection, predicted to characterize the early stages of sympatric/parapatric speciation scenarios [14,15]. As speciation progresses, genomic islands may

increase both in number and extent, as inter-specific recombination rates decrease, potentially enhancing reproductive isolation via divergence hitchhiking (DH) [10,13,16]. At later stages of ecological speciation, as many loci diverge among incipient species, genome-wide effective gene flow will be reduced even at unlinked loci: divergence hitchhiking being superseded by genome hitchhiking (GH) [16]. Many studies have employed population genomic methods to identify genomic regions with unusually strong differentiation as holding potential candidate genes for disruptive selection between diverging taxa *e.g.* [17-21]. However little effort has been devoted to investigating how changes in natural selection regimes through anthropogenic perturbation affect the occurrence and size of genomic islands of divergence between diverging (now converging) species, and nothing is published about this to our knowledge.

The Alpine whitefish (*Coregonus* spp.) radiation is native to around 40 lakes in the headwaters of three major European river drainages: the Rhine, Rhône and Danube [22]. Colonising following the glacial retreat (15,000 – 20,000 years BP), this monophyletic group has diverged in-situ into at least seven species flocks, centred on single large lakes or previously connected sister lakes [23-25]. The constituent species of these lake or super-lake flocks are generally each other's closest relatives, indicating intralacustrine speciation [25]. Historically, *Coregonus* species richness varied from one to six endemic species per lake. Coexisting species differ in traits linked to feeding and reproductive ecology, with very similar ecotypes having evolved repeatedly in several lakes [23,25,26].

During the 20th Century, many of the lakes in which the Alpine *Coregonus* radiation occurs have seen unprecedented anthropogenic impacts. Run-off from agriculture and insufficient sewage treatment massively increased the nutrient input into many lakes, leading to eutrophication of these naturally oligotrophic lakes. The increase in biologically available phosphorus had several effects on lake ecology, changing the community structure of phytoplankton and their zooplankton grazers, therefore affecting the prey resource base for whitefish species [8,27]. Increased productivity also resulted in increased bacterial decomposition, reducing levels of available oxygen in the hypolimnion [28,29]. This de-oxygenation is most severe at the sediment-water interface, where whitefish egg development occurs [29]. A consequence of this is the reduction of the available depth gradient partitioned between whitefish species during spawning and concomitantly an increase in rates of hybridization between shallow and deep spawning species [8,30,31]. In more extreme cases de-oxygenation renders deeper areas of the lake anoxic, potentially reducing the available area and diversity of foraging habitat

[29]. During this period of unprecedented environmental change, the Alpine whitefish radiation experienced a loss of around 38% of its pre-eutrophication species richness [8]. This diversity loss appears to have been driven by a combination of negative population growth combined with the genetic collapse of coexisting species through speciation reversal. Indeed by comparing historical population samples taken before large-scale eutrophication with samples taken from the same species post-eutrophication, we see a clear signal of lower levels of neutral genetic and phenotypic differentiation between coexisting species within lakes that had high levels of eutrophication, in comparison to species from lakes that have never been strongly eutrophic [8].

Here we present the first genome scan analysis that explores the impact of anthropogenic environmental change on heterogeneous genomic divergence in an adaptive radiation. We studied whitefish species from nine Alpine lakes, comprising five phylogenetically independent *Coregonus* adaptive radiations and one species assemblage arising from recent anthropogenic secondary contact. We employ genome scans of 183 individuals comprising 851 AFLP markers. We subject these AFLP loci to  $F_{ST}$  outlier detection to identify candidate regions in the genome that likely evolved under disruptive selection, and logistic regression analyses methods to link candidate regions to adaptive traits that they potentially influence. For all analysis methods, we consistently find a trend for more outlier and more trait-associated loci in those lakes that experienced lower levels of eutrophication. Here we provide the first genetic evidence for the weakening of the genomic signature of disruptive selection and divergent adaptation that are predicted to cause speciation reversal following human-induced habitat modification.

## Methods

### Study system & sampling

We sampled *Coregonus* species from nine different lakes, each containing between two and five coexisting species. Individuals were collected between 2004 and 2006 on their respective spawning grounds using overnight gill-netting. Information about the species and lakes sampled can be found in Table 1 and Figure 1. Species were assigned to feeding ecotype using their mean gill-raker number, a correlate of feeding ecotype [32]: very low gill-raker numbers (VLGR <25 gill-rakers), low (LGR 25–30), medium (MGR 31–35) and high (HGR >35). We used both the historical maximum total dissolved phosphorus concentration ( $P_{max}$  ( $P_{tot}$ ) ( $\mu\text{g l}^{-1}$ )) and contemporary minimum oxygen concentration at depth (Min.  $O_2$  ( $\text{mg l}^{-1}$ )) as indicators of the intensity of anthropogenic nutrient pollution: both variables potentially show different aspects of the speciation reversal process,

relating to the duration and intensity of habitat disruption and its effects on genomic divergence.  $P_{max}$  was used as a surrogate for the maximum level of pollution experienced by each lake system and therefore to measure the impact on interspecific genomic divergence when pollution was at its most severe. Min.  $O_2$  was used to gauge the extent to which lake systems have currently returned to their pre-eutrophication state, and therefore represents the potential ongoing effects of relaxed selection on genomic divergence. Both  $P_{max}$  and Min.  $O_2$  have been shown to strongly correlate (positively and negatively respectively) with a loss in intralacustrine *Coregonus* species diversity [8].

Lakes Neuchâtel and Biel are connected by *ca.* 8 Km of river and their three endemic *Coregonus* species form part of a monophyletic group with other nearby single-species lakes [23,25]. The five *Coregonus* species inhabiting connected (*ca.* 5 km of river) lakes Thun and Brienz, form a monophyletic clade too, except *C. fatioi*, which appears to have had some genetic input from Lake Constance *Coregonus* species [23,25]. The endemic species flock of Lake Lucerne consists of potentially five coexisting species [Lundsgaard *et al.* unpublished observations, 26]. For this study no samples of two recently discovered forms, intermediate in size and gill-raker counts, were available. Lake Constance consists of two major basins, Obersee (476 Km<sup>2</sup>) and Untersee (63 Km<sup>2</sup>). Previously with up to six reported species, heavy eutrophication has led to the extinction of at least one species, the VLGR *C. gutturosus* [8,22]. Four species coexist in the Obersee, with the *C. sp.* “weissfelchen” found in the Untersee basin [23,25]. Lakes Walen and Zuerich are connected by *ca.* 17 Km of river, the two *Coregonus* species endemic to these lakes, together with whitefish from two nearby, single-species lakes, form an independent, monophyletic radiation. Additionally from Walen and Zuerich, one or two poorly known species, may also have become extinct [8,26]. The two coexisting whitefish species of Lake Maggiore are introduced; *C. sp.* “lavarello” is most closely related to *C. sp.* “Zugbalchen” from Lake Zug [23,25], while *C. sp.* “bondella”, groups closest to *C. zugensis* from Lake Lucerne [25]. This species pair represents a case of recent secondary contact, *C. sp.* “lavarello” and *C. sp.* “bondella” being introduced *ca.* 1860 and 1950 respectively [33].

### AFLP analysis

Total genomic DNA was extracted using the Wizard Genomic DNA purification kit (Promega, Madison, WI). A genome scan of putatively independent loci was then carried out using AFLPs, following a protocol modified from [34], with combined restriction/ligation steps, using *EcoRI* and *MseI* restriction enzymes. Selective PCRs were performed with 14 different primer-pairs

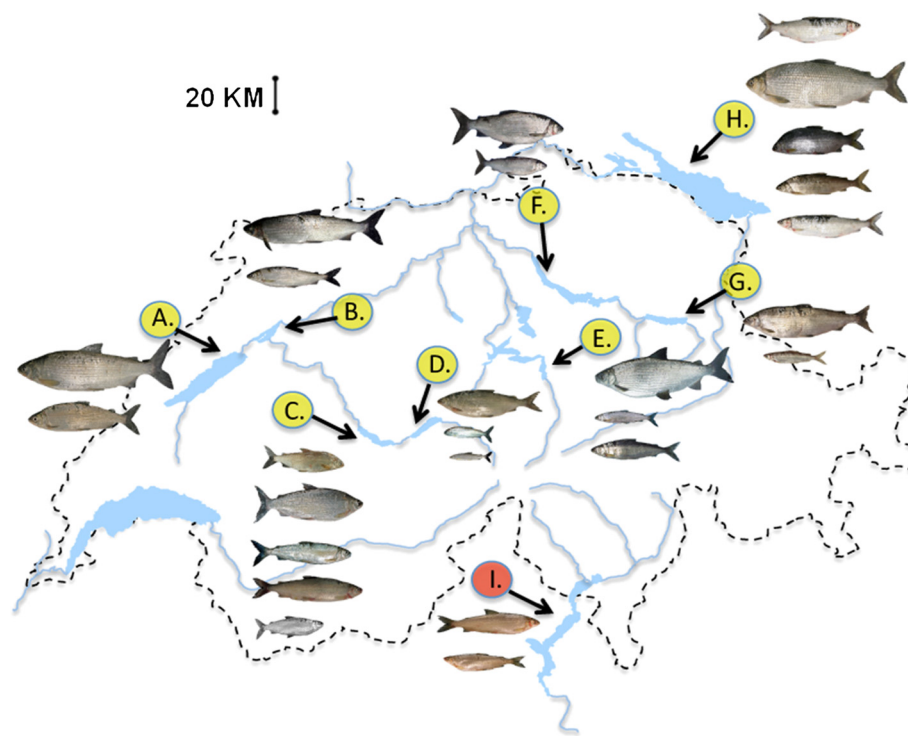
**Table 1 Geographical location, lake surface area, lake trophic state, taxonomy, spawning/feeding ecotype designation and the number of individuals of each species included, for all species flocks analysed**

Lake	Co-ordinates	Surface Area (km <sup>2</sup> )	P <sub>max</sub> (P <sub>tot</sub> ) (µg l <sup>-1</sup> )	Min. O <sub>2</sub> (mg l <sup>-1</sup> )	Species	Ecotype	# Individuals
Neuchâtel	46°53 N; 6°5 E	217.9	50	3.2	<i>C. palaea</i>	LGR, W	7
					<i>C. candidus</i>	MGR, W	7
							<b>14</b>
Biel	47°04 N; 7°1 E	39.3	132	0.5	<i>C. palaea</i>	LGR, W	6
					<i>C. confusus</i>	MGR, W	7
							<b>13</b>
Thun	46°41 N; 7°42 E	48.4	21	7.6	<i>C. alpinus</i>	VLGR, S	7
					<i>C. sp. "balchen"</i>	LGR, W	6
					<i>C. fatioi</i>	MGR, W	8
					<i>C. sp. "felchen"</i>	MGR, W (Deep)	7
					<i>C. albellus</i>	HGR, S	7
							<b>35</b>
Brienz	46°43 N; 7°58 E	29.8	17	9.3	<i>C. sp. "balchen"</i>	LGR, W	4
					<i>C. sp. "felchen"</i>	MGR, W	7
					<i>C. albellus</i>	HGR, S	5
							<b>16</b>
Lucerne	47°01 N; 8°22 E	114	34	3.4	<i>C. sp. "bodenbalchen"</i>	LGR, W	10
					<i>C. zugensis</i>	HGR, S/W	9
					<i>C. nobilis</i>	HGR, S	7
							<b>26</b>
Constance	47°36 N; 9°26 E	539	87	6.0	<i>C. sp. "weissfelchen"</i>	LGR, W	7
					<i>C. arenicolus</i>	LGR, W	6
					<i>C. sp. "Alpenrhein"</i>	MGR, W	5
					<i>C. macrophthalmus</i>	HGR, W	6
					<i>C. wartmanni</i>	HGR, W (Pelagic)	6
							<b>30</b>
Walen	47°07 N; 9°12 E	24.1	26	6.1	<i>C. duplex</i>	LGR, W	8
					<i>C. heglingus</i>	HGR, S/W	9
							<b>17</b>
Zuerich	47°13 N; 8.43 E	88.4	119	1.1	<i>C. duplex</i>	LGR, W	11
					<i>C. heglingus</i>	HGR, W	7
							<b>18</b>
Maggiore	45°58 N; 8°4 E	212.5	37	n/a	<i>C. sp. "lavarello"</i>	MGR, W	8
					<i>C. sp. "bondella"</i>	HGR, W	6
							<b>14</b>

Feeding ecotype: VLGR = Very low mean gill-raker number (<25 gill-rakers), LGR = low (25–30), MGR = medium (31–35) and HGR = high (>35). Reproductive ecotype: W = winter spawner (October-January), S = summer spawner (July-September). P<sub>max</sub> = peak historical levels of phosphorus enrichment. Min. O<sub>2</sub> = contemporary minimum oxygen concentration at depth.

(ACG-CAC, ACT-CAC, ACC-CAG, ACG-CAG, ACT-CAG, ACT-CAT, ACA-CGC, ACC-CGC, ACT-CGC, ACT-CGG, ATG-CTC, ATG-CTG, AAG-CTT, ACG-CTT). Fragment separation was carried out by electrophoresis on a CEQ-8000 automatic sequencer (Beckman-Coulter). AFLP traces were then screened and

scored for peak presence/absence using GENEMARKER 1.85 (SoftGenetics, LLC, State College, PA). Individual traces were automatically scored and then bins were visually inspected and manually adjusted. Only AFLP loci between 55–400 base pairs were scored. To check levels of marker reproducibility 30 individuals were



**Figure 1 Geographic locations of the species flocks studied within Switzerland.** Map showing the number of species included, their sampling location and their representative phenotypes (species names listed from above to below). (A) Lake Neuchâtel: (*C. palaea*, *C. candidus*), (B) Lake Biel: (*C. palaea*, *C. confusus*), (C) Lake Thun: (*C. alpinus*, *C. sp.* "balchen", *C. fatioi*, *C. sp.* "felchen", *C. albellus*), (D) Lake Brienz: (*C. sp.* "balchen", *C. sp.* "felchen", *C. albellus*), (E) Lake Lucerne: (*C. sp.* "bodenbalchen", *C. zugensis*, *C. nobilis*), (F) Lake Zuerich: (*C. duplex*, *C. heglingus*), (G) Lake Walen: (*C. duplex*, *C. heglingus*), (H) Lake Constance: (*C. sp.* "weissfelchen", *C. arenicolus*, *C. sp.* "Alpenrhein", *C. macrophthalmus*, *C. wartmanni*), (I) Lake Maggiore: (*C. sp.* "lavarello", *C. sp.* "bondella"). Yellow circles denote native species flocks. Red circle denotes anthropogenic origin and secondary contact.

repeated (16% of the total analysed samples). The level of reproducibility per locus was then calculated for each of the 14 genotyped primer-pairs [35].

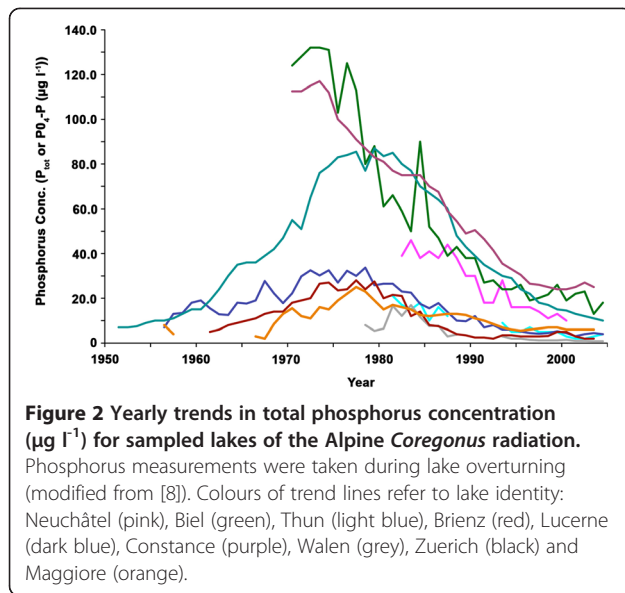
#### Genomic outlier analysis

All outlier analyses and logistic regressions were run on the nine lake flock data sets separately. This avoided potential violations of the demographic models employed in the outlier detection methods that could be caused by hierarchical structuring of genetic variation [36,37], present in the Alpine whitefish radiation [25]. Additionally, with this analysis design, we were able to compare paired lake flock data sets with similar evolutionary histories and patterns of ecological diversification, but differing in their nutrient enrichment history: Lake Walen (very low level of nutrient enrichment) vs. Lake Zuerich (high), Lake Biel (high) vs. Lake Neuchâtel (moderate to high), and Lake Thun (low) vs. Lake Brienz (very low) (Figure 2).

Two different  $F_{ST}$ -based approaches to search for selection candidate loci, suitable for dominant markers were used: DFDIST\_D [38] and BAYESCAN [39]. DFDIST estimates the probability of a locus to be under

selection according to its observed  $F_{ST}$  and  $H_e$ , compared to simulated neutral distributions computed for an average neutral  $F_{ST}$  observed in the empirical data set (the trimmed mean  $F_{ST}$ ), effective population size and minimum levels of polymorphism [38]. In contrast, BAYESCAN directly estimates the probability of each locus being subject to selection by comparing the likelihoods of neutral and selective models [39]. As the power to detect stabilising selection will be low among recently diverged species, only candidate loci showing unusually high levels of genetic differentiation, implicating disruptive natural selection, were counted [38].  $F_{ST}$  outlier analyses were first run on each multi-species lake data set (global comparisons among all species sampled in a lake), followed by all possible pairwise comparisons between individual species for lakes with more than two species.

In the DFDIST analysis the mean  $F_{ST}$ , used for simulating a null distribution of  $F_{ST}$  values for loci evolving under neutrality, was estimated using a trimmed mean of 0.3 [38]. The null distribution was generated using 50,000 simulated loci. In the empirical AFLP data set the maximum allele frequency was set to 0.98. AFLP loci



with  $F_{ST}$  values above the upper 95th quantile of the null distribution were considered candidate loci for being under disruptive natural selection.

In the BAYESCAN analysis, model parameter estimation was automatically tuned during the pilot runs (ten runs, length 2,000). A 10,000 iteration burn-in was found to be sufficient for convergence between MCMC chains. Sample sizes of 10,000 and a thinning ratio of 50 were used as suggested by [39]. Prior to the analysis all non-polymorphic loci (fixed presence or absence alleles) were removed as this could bias the analysis (M. Foll pers. comm.). Loci were considered as candidates for being under divergent selection if their  $\log_{10}$  Bayes Factor scores were  $\geq 0.5$  ("Substantial" on Jeffrey's scale of evidence) [39].

To understand the effects of loci under selection on the phylogenetic relationships within the Alpine whitefish radiation, three different sets of Neighbour-joining (NJ) population-based trees were created: (1) using all genotyped AFLP loci (835 loci), (2) a subset comprising only outlier loci identified by DFDIST and BAYESCAN in either global or pairwise comparisons (96 loci), (3) a subset comprising only neutral loci remaining following removal of outlier loci (739 loci). Nei's genetic distances between populations were calculated in AFLP-SURV 1.0 [40]. NJ trees were reconstructed from these distances using the program NEIGHBOR in the software package PHYLIP 3.67, with 1,000 bootstrap replicates [41]. To quantify the levels of parallel ecotypic trends among lakes in the different loci subsets, following Allender *et al.* [42], Mantel randomizations were performed for neutral and outlier loci subsets for a matrix of Nei's genetic distances and a design matrix encoding feeding ecotype. In the design matrix, species belonging to the same ecotype were encoded  $D = 0$ ,  $D = 1$  encoding assignment to

different ecotypes. As the Maggiore species are non-native, they were excluded from this analysis. Mantel randomizations of 10,000 permutations were carried out in ARLEQUIN 3.5 [43].

To assess the statistical independence of individual outlier loci from each other, levels of linkage disequilibrium (LD) were estimated for different AFLP loci subsets. For each species, genotyped AFLP loci were split into exclusive subsets: neutral and outlier. The outlier subsets comprised all loci identified as outliers by DFDIST analysis, during global-level comparisons within their respective species flock. LD ( $\bar{r}_d$ ) values were calculated in the program MULTILOCUS 1.3 [44]. Pairwise  $\bar{r}_d$  values were calculated for all possible pairs of loci within each neutral and outlier loci subset, for each species. The significance of differences in mean pairwise LD between neutral and outlier loci subsets, across populations, was tested using paired  $t$ -tests.

#### Divergent trait measurement

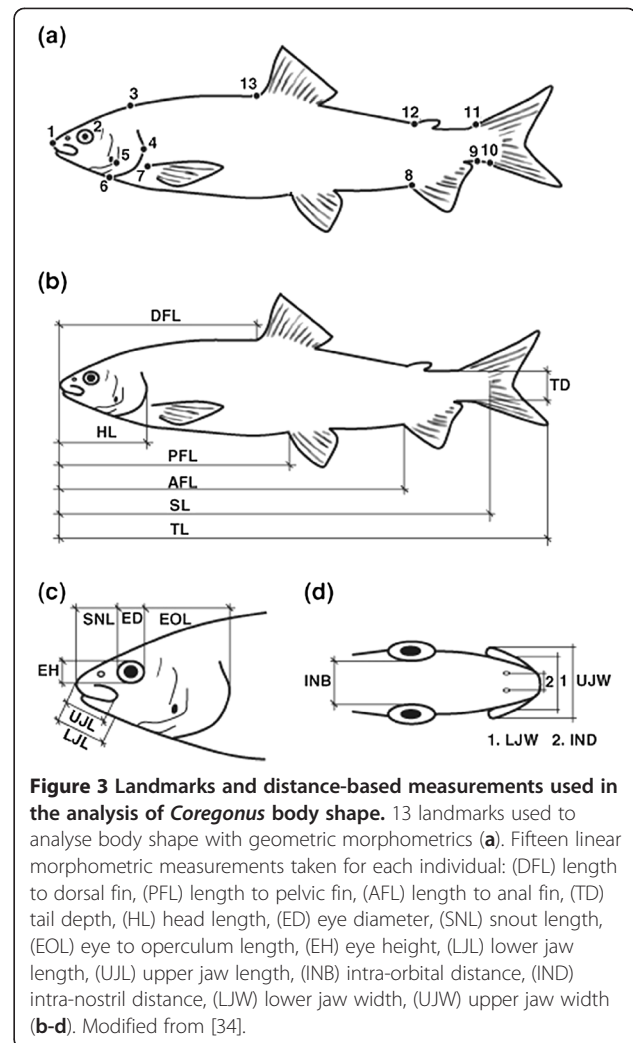
The first two trait categories measured, spawning depth and time, are related to reproductive ecology [26,31]. (1) Spawning date influences the time of emergence of the fry and therefore may influence survival through different mechanisms such as food availability, predation or over-wintering mortality [45-48] and has been shown to transgressively segregate in whitefish hybrids [48,49]. Importantly, spawning date also influences temporal reproductive isolation between species [26]. The spawning date for each individual was taken as the day of capture on spawning site, adjusted so that the earliest caught fish (summer spawning) were caught on day 1, therefore those fish caught  $x$  days after the first fish were given the spawning date of  $x$ . Fish caught on the same date of different calendar years would be given the same spawning date. (2) Spawning depth is likely related to physiological or behavioural adaptations to depth [50], to different predation pressures [51], or to differences in egg size and development [48]. Also spawning depth influences spatial reproductive isolation between species [8,26,31]. Spawning depth for each individual was estimated by finding the average depth over which the gill nets were set during each sampling event.

Finally we measured three trait categories primarily related to feeding ecology: gill-raker number, linear morphometric distances and geometric shape analyses. (3) Gill-raker number has been shown to be correlated with feeding niche in Scandinavian whitefish [52] and to have a high heritability [32]. Gill-raker numbers were counted on the first gill arch from the left side of each fish. (4-5) Overall morphological variation among fish was measured using (4) linear morphological distances (LM) and (5) geometric morphometrics of body shape (GM). Shape aspects are potentially targets of disruptive

selection, as revealed in  $Q_{ST}$ - $F_{ST}$  comparisons in lakes of the Alpine radiation [31] and show significant phenotype-environment correlations in Scandinavian *Coregonus* species flocks [53]. (4) Fifteen different LM measurements, based on those described in Chouinard, Pigeon & Bernatchez [54], were taken using digital callipers on 1090 individuals, comprising 38 whitefish populations/species from 22 lakes in the Alpine region: to capture the major axes of morphological divergence across the radiation, (Figure 3b-d). Size correction of the measured traits was performed using common principal components analysis (CPCA), followed by Burnaby's back projection method (BBPM) through the CPCBP package [55] in R (R Development Core Team, 2004). Principal component analysis (PCA) was then performed on the transformed data in SPSS 16 (SPSS Inc., Chicago Illinois, USA). The resulting PC 1–3 scores of the AFLP-genotyped individuals from the nine lake flocks were taken for logistic regression analysis. For PC 1 (30% of variation) traits with positive loadings were mainly body length traits (AFL, PFL, DFL, HL), whereas head measures particularly jaw width traits (LJW, UJW, UJL, ED, SEL) had negative loadings. PC 2 (24%) had positive loadings for head measures (HL, ED, SEL, EH, LJL, UJL) and negative loadings for traits: AFL, PFL, DFL, TD, LJW, UJW. PC 3 (10%) had strongly positive loadings for eye shape (ED, EH) and negative loadings for: SEL, UJL, INB, SNW. (5) GM landmark-based analysis of body shape variation was carried out on digital images of 1545 fish comprising 41 populations/species from 22 lakes, using 13 homologous landmarks (Figure 3a). Digital landmark placement was carried out in TPSDIG2 (<http://life.bio.sunysb.edu/morph/>). All further shape analysis was carried out in MORPHOJ 1.02d [56]. Generalized least-squares Procrustes super-imposition was carried out on landmark configurations. Allometric size correction was then carried out by the pooling of species-specific regressions of Procrustes coordinates against individual standard length. Following PCA, PC 1 was found to be mostly artifactual (associated with dorso-ventral flexion of dead fish) and was discarded. Therefore individual PC scores for PC 2 (13%), PC 3 (9%) and PC4 (9%) were used for further analysis. Shape changes along PC 2 (13%) were associated with changes in head length. PC 3 (9%) was associated mainly with the amount of up/down rotation of the head, and changes in body depth. PC 4 (9%) was mainly associated with changes in the length of the caudal peduncle and body depth.

#### Association between genomic outlier loci and putative adaptive traits

Univariate logistic regressions were carried out in the program MATSAM 2Beta [57]. This approach analyses



**Figure 3 Landmarks and distance-based measurements used in the analysis of *Coregonus* body shape.** 13 landmarks used to analyse body shape with geometric morphometrics (a). Fifteen linear morphometric measurements taken for each individual: (DFL) length to dorsal fin, (PFL) length to pelvic fin, (AFL) length to anal fin, (TD) tail depth, (HL) head length, (ED) eye diameter, (SNL) snout length, (EOL) eye to operculum length, (EH) eye height, (LJL) lower jaw length, (UJL) upper jaw length, (INB) intra-orbital distance, (IND) intra-nostril distance, (LJW) lower jaw width, (UJW) upper jaw width (b-d). Modified from [34].

the relationship between alleles at a given AFLP locus and specific values of an explanatory variable, such as putative adaptive traits [57]. For each AFLP locus, the genotype of individuals was regressed against the individual values for each of the five putative adaptive trait variables measured: spawning date, spawning depth, gill-raker number, PC score on principal axes of morphological variation resulting from LM measurements and GM analysis of shape. Each suite of genotype/phenotype regressions were calculated separately for each of the nine species flock data sets. Individuals with missing phenotypic data were removed from the logistic regression for that specific trait. The adaptive trait values for each individual, used in logistic regressions, can be found in Additional file 1: Table S1. The significance of the association between the frequency of an allele at a locus and the value of each trait was assessed using the *G*-statistic at the 0.01, 0.05 and 0.1 significance levels, with Bonferroni correction using the number of AFLP loci used in each regression analysis.



### Tests of association between heterogeneous genomic divergence and eutrophication

To test if the distribution over lakes of the detected genomic signal of disruptive selection (number of outlier loci and trait-associated loci) was robust to analysis method (DFDIST, BAYESCAN, MATSAM), correlations between the numbers of loci detected by each analysis method were carried out. To test the hypothesis that eutrophication causes a decrease in the number of identified outlier/trait-associated loci, one-tailed regressions were calculated of the number of loci detected in global comparisons within each species flock, against maximum historical phosphorus concentration ( $P_{\max}$ ) and current minimum oxygen concentration at depth (Min.  $O_2$ ) of the lake the species flock inhabits. This test was repeated for each of the three methods employed (DFDIST, BAYESCAN, MATSAM).

As a comparison the effect of eutrophication on neutral markers was also tested by linear regression of  $P_{\max}$  and Min.  $O_2$  against  $F_{ST}$  values from each lake flock calculated from three different marker sets: (1) microsatellites (maximum within-lake pairwise  $F_{ST}$  value), (2) all AFLP loci (highest within-lake weighted mean  $F_{ST}$  in DFDIST from pairwise comparisons), (3) neutral AFLP loci (highest within-lake weighted mean  $F_{ST}$  in DFDIST from pairwise comparisons, following the removal of all candidate loci identified by the DFDIST, BAYESCAN, MATSAM analyses). Lake flock microsatellite  $F_{ST}$  values resulting from between ten (Maggiore only) or eleven unlinked loci were taken from [8,58], Hudson *et al.* unpublished observations. We also calculated pairwise correlations between the three independent estimates of  $F_{ST}$  to see how congruent their signals were. All statistical analyses were carried out in SPSS 16 (SPSS Inc., Chicago Illinois, USA).

### Results

The total number of AFLP loci scored per lake flock varied from between 355–405 (DFDIST) and 216–306 (BAYESCAN) (Table 2). The number of whitefish genotyped per flock ranged from 13 to 35 (Tables 1, 2), depending largely on the number of constituent species. For species-pairwise analyses the number of loci genotyped ranged from between 334–380 (DFDIST) and 177–247 (BAYESCAN), and the number of individuals from 9–19. Average per locus levels of reproducibility across the fourteen primer pairs exceeded 95%. Identity codes, the total number of loci (both polymorphic and fixed) and average reproducibility per primer pair can be found in Additional file 2: Table S2.

### $F_{ST}$ outlier scans

The DFDIST analyses detected disruptive selection candidate loci in every species flocks. The largest number of

outliers was found in the Lake Thun species assemblage (22) (Table 2, Additional file 3: Table S3a). Most loci detected (64 of 81) fell within the 95–99% quantiles. Loci with  $F_{ST}$  values above the 99% quantile were found in L. Thun (four), L. Zuerich (three), L. Brienz (two), L. Lucerne (two), L. Walen (two), L. Biel (one) and L. Constance (one). Of the four loci found to be outliers in multiple lakes, three loci showed parallel allelic trends among similar ecotypes. CAGC-235 was an outlier in the Thun, Lucerne, Zuerich and Maggiore species flocks. For this locus prevalence of the presence allele was reduced in Zuerich LGR *C. duplex*, Lucerne LGR *C. sp.* "bodenbalchen", Thun MGR *C. fatioi* and Maggiore MGR *C. sp.* "lavarello". The presence allele of CGAG-186 was at a higher prevalence in Brienz MGR *C. sp.* "felchen" and Thun/Brienz HGR *C. albellus*. At TGTG-183 there was a decreased prevalence of the presence allele in LGR *C. sp.* "balchen" from both lakes Brienz and Thun and also in Thun MGR *C. fatioi*. Pairwise DFDIST analyses (Additional file 3: Table S3b) revealed 36 additional loci not recovered in the global analyses. Six of these loci were identified as outliers in several pairwise species comparisons from different lakes, with three showing parallel trends among similar ecotypes. Locus CAGC-342 had a high prevalence of the presence allele in HGR L. Brienz *C. albellus* and HGR L. Constance *C. wartmanni*. TGTG-271 showed parallel trends in HGR Lucerne *C. zugensis* and HGR Maggiore *C. sp.* "bondella". Finally the presence allele of CGAG-355 had a reduced frequency in MGR Constance *C. sp.* "Alpenrhein" and MGR Maggiore *C. sp.* "lavarello".

From the global lake-level BAYESCAN analyses, we find disruptive selection candidate loci between the constituent species in every lake except Biel and Constance (Table 2, Additional file 4: Table S4a). 15 out of the 20 outlier loci identified had Bayes factor scores of "substantial" on Jeffrey's scale of evidence ( $\geq 0.5 - 1$ ). One locus scoring "strong" ( $\geq 1 - 1.5$ ) was found in each of Thun, Brienz and Zuerich flocks and two in Maggiore. Similar to the DFDIST analyses, CAGC-235, was detected as an outlier in multiple lakes. The presence allele for this locus had a low prevalence in the Thun MGR *C. fatioi* and Lucerne LGR *C. sp.* "bodenbalchen". Pairwise BAYESCAN analyses (Additional file 4: Table S4b) revealed five additional outlier loci not recovered in the global analyses at lake level. None of these loci were outliers in more than one species pair.

Overall the DFDIST analysis identified over four times more outlier loci than the BAYESCAN analysis (81 vs. 20). Similarity of identity of candidate loci identified with either method (global  $F_{ST}$  analysis at lake level) was lowest in lakes with low average mean  $F_{ST}$  (for Lakes Constance and Biel zero loci were found in the BAYESCAN analyses, but four and two loci respectively in DFDIST analyses) and/or with more than two species coexisting

**Table 2 Number of selection candidate loci identified using DFDIST and BAYESCAN within individual lake flocks (global comparisons)**

Species Flock	# Species	# Individuals	# Poly. Loci	# Candidate loci DFDIST		Total	# Candidate loci BAYESCAN		Total	Shared Loci	% Similarity
				≥ 0.95	≥ 0.99		≥ 0.5	≥ 1			
Neuchâtel	2	14	241	5	0	5	2	0	2	2	40
Biel	2	13	262	1	1	2	0	0	0	0	0
Thun	5	35	306	18	4	22	3	1	4	4	18
Brienz	3	16	216	9	2	11	1	1	2	2	18
Lucerne	3	26	266	9	2	11	3	0	3	2	18
Constance	5	30	279	4	1	5	0	0	0	0	0
Walen	2	17	229	2	2	4	2	0	2	2	50
Zuerich	2	18	230	5	3	8	3	1	4	4	50
Maggiore	2	14	223	11	2	13	1	2	3	3	23
<b>Total</b>				64	17	81	15	5	20		

Presented are the numbers of species, individuals and polymorphic AFLP loci comprising each species flock data set. Identified DFDIST outlier loci are grouped according to how many fell between the 95% and 99% quantiles ( $\geq 0.95$ ), and how many fell above the 99% quantile ( $\geq 0.99$ ). BAYESCAN outlier loci are grouped by their Bayesian posterior probability of being under selection ( $\text{Log}_{10}$  Bayes factor scores): “substantial” 0.5 to 1 ( $\geq 0.5$ ), “strong” 1 to 1.5 ( $\geq 1$ ). Also shown: the number of AFLP loci identified as selection candidates by both outlier detection methods and the overall similarity between sets of candidate loci recovered in DFDIST and in BAYESCAN, for each species flock.

(L. Brienz, L. Thun and L. Lucerne; only 18% of loci identified by both detection methods) (Table 2). Deviations were in all cases due to DFDIST detecting higher numbers of  $F_{ST}$  outliers.

Consensus trees from the three different subsets of AFLP loci showed high levels of congruence (Additional file 5: Figure S1). Compared to the all loci consensus tree, the outlier loci subset tree showed little obvious increase in ecotype clustering. Design matrices of feeding ecotype designation explained slightly more variance in genetic distances between species when the latter were calculated from the outlier loci subset rather than the neutral loci subset ( $r = 0.07$  and  $r = 0.04$  respectively), but none of these Mantel randomizations were significant (0.004%,  $P = 0.12$  and 0.001%,  $P = 0.22$  respectively).

The mean values of pairwise LD for neutral and outlier loci classes, across species, were not significantly different (mean difference = 0.026,  $t_{23} = 1.591$ ,  $p = 0.125$ ). Overall outlier loci subsets generally had lower mean pairwise LD within species than neutral loci subsets (0.077 vs. 0.105). The range in mean pairwise LD values across species was greater for the outlier loci class (0 – 0.333 vs. 0.077 – 0.157), however variance around population means were qualitatively similar for both outlier and neutral loci classes (Additional file 6: Table S5, Additional file 7: Figure S2).

#### Association between genomic outlier loci and putative adaptive traits

Across all lake flocks, a total of 43 significant associations between putative adaptive traits and AFLP markers were found (Table 3, Additional file 8: Table S6). The

divergent trait with the largest numbers of significant associations with AFLP alleles, across all lake flocks, was gill-raker number (19 loci), followed by LM PC 1 (jaw widths) (eight), spawning depth (six), spawning date (four), GM PC 2 (head length), LM PCs 2 (eye and head size) and 3 (skull widths)(each two). No significant associations were found for GM PCs 3 (head position) and 4 (caudal peduncle length). Within lake flocks the largest number of significant associations were found in Lake Thun (17), followed by L. Lucerne (nine), L. Walen (six), L. Brienz (five), L. Zuerich (four), L. Constance (two) and none in lakes Neuchâtel, Biel and Maggiore.

Four loci showed significant parallel trends of association with adaptive traits in multiple lake flocks. As in the candidate loci screen, CAGC-235 again showed parallel allelic trends; the presence allele was significantly associated in Lakes Lucerne and Brienz with lower gill-raker numbers. The CTAC-99 presence allele was significantly associated with increased gill-raker numbers in lakes Brienz and Thun. The presence allele of CGAG-186 was found in individuals with high gill-raker numbers in L. Brienz and individuals with more negative LM PC 3 scores (reduced skull widths) in L. Thun. The presence allele of CTAG-173 had a significant association with earlier spawning dates in L. Thun and with lower gill-raker numbers in L. Lucerne.

The total number of trait-associated loci found using MATSAM, also detected as  $F_{ST}$  outlier loci was higher for DFDIST (26) than for BAYESCAN analyses (11). However the average percentage of loci that were both outlier and trait-associated, were very similar between the sets of loci identified with DFDIST (22%), and

**Table 3 Number of AFLP loci revealing a significant association with each putative adaptive trait, for each species flock**

Species Flock	# Significant Loci							Total
	GR	LM PC 1	LM PC 2	LM PC 3	GM PC 2	Date	Depth	
Neuchâtel	0/0/0	0/0/0	0/0/0	0/0/0	0/0/0	0/0/0	0/0/0	0
Biel	0/0/0	n/a	n/a	n/a	0/0/0	0/0/0	0/0/0	0
Thun	2/4/4	3/5/6	0/0/0	1/2/2	0/0/1	0/0/3	0/0/1	17
Brienz	0/2/4	0/0/0	0/0/0	0/0/0	0/0/0	0/0/1	0/0/0	5
Lucerne	1/2/5	0/0/0	0/0/1	0/0/0	0/0/0	0/0/0	2/3/3	9
Constance	0/0/1	0/0/1	0/0/0	0/0/0	0/0/0	0/0/0	0/0/0	2
Walen	0/1/1	0/1/1	0/0/1	0/0/0	0/0/1	0/0/0	0/0/2	6
Zuerich	0/4/4	0/0/0	0/0/0	0/0/0	0/0/0	0/0/0	0/0/0	4
Maggiore	0/0/0	0/0/0	0/0/0	0/0/0	0/0/0	0/0/0	0/0/0	0
Total	19	8	2	2	2	4	6	43

Numbers in adaptive trait columns refer to the total number of loci found at each level of significance using the  $G$ -statistic ( $p < 0.01$  /  $p < 0.05$  /  $p < 0.1$ ). Adaptive traits used: gill-raker number (GR), linear morphometric PCA scores (LM PC 1, PC 2, PC 3), geometric morphometric PCA scores (GM PC 2), spawning date and spawning depth. n/a = analyses that could not be carried out due to missing data. GM PC 3, PC 4 are not represented as no significant associations were found for these variables.

BAYESCAN (21%)(Additional file 9: Table S7). Overall the lake flock with the highest number of selection candidate loci significantly associated with adaptive trait values was Thun (12 loci). Lakes Brienz, Lucerne and Zuerich came next highest, each having 4 loci, with Walen having two such loci, the remaining species flocks having none.

#### Tests of association between heterogeneous genomic divergence and eutrophication

The number of selection candidate and trait-associated loci identified at the species flock level through the three different analyses methods were significantly positively correlated in two of the three comparisons: DFDIST versus BAYESCAN ( $r_8 = 0.717$ ,  $P < 0.05$ ) and DFDIST versus MATSAM ( $r_8 = 0.763$ ,  $P < 0.05$ ). Results from BAYESCAN and MATSAM also tended to be positively correlated ( $r_8 = 0.570$ ,  $P = 0.1$ ). All linear regressions of the effects of  $P_{max}$  on the number of selection candidate loci and significant trait-locus associations showed negative, albeit marginally non-significant relationships (DFDIST global comparisons  $r^2 = 0.315$ ,  $P = 0.058$ ; BAYESCAN global comparisons  $r^2 = 0.155$ ,  $P = 0.148$ ; MATSAM  $r^2 = 0.239$ ,  $P = 0.09$ ; all  $P$  values one-tailed: Figure 4). The regression effect of Min. O<sub>2</sub> on the number of selection candidate loci and trait-locus associations was positive, but weaker than that of  $P_{max}$  and non-significant (DFDIST global comparisons  $r^2 = 0.263$ ,  $P = 0.097$ ; BAYESCAN global comparisons  $r^2 = 0.009$ ,  $P = 0.413$ ; MATSAM  $r^2 = 0.236$ ,  $P = 0.111$ ; Additional file 10: Figure S3).

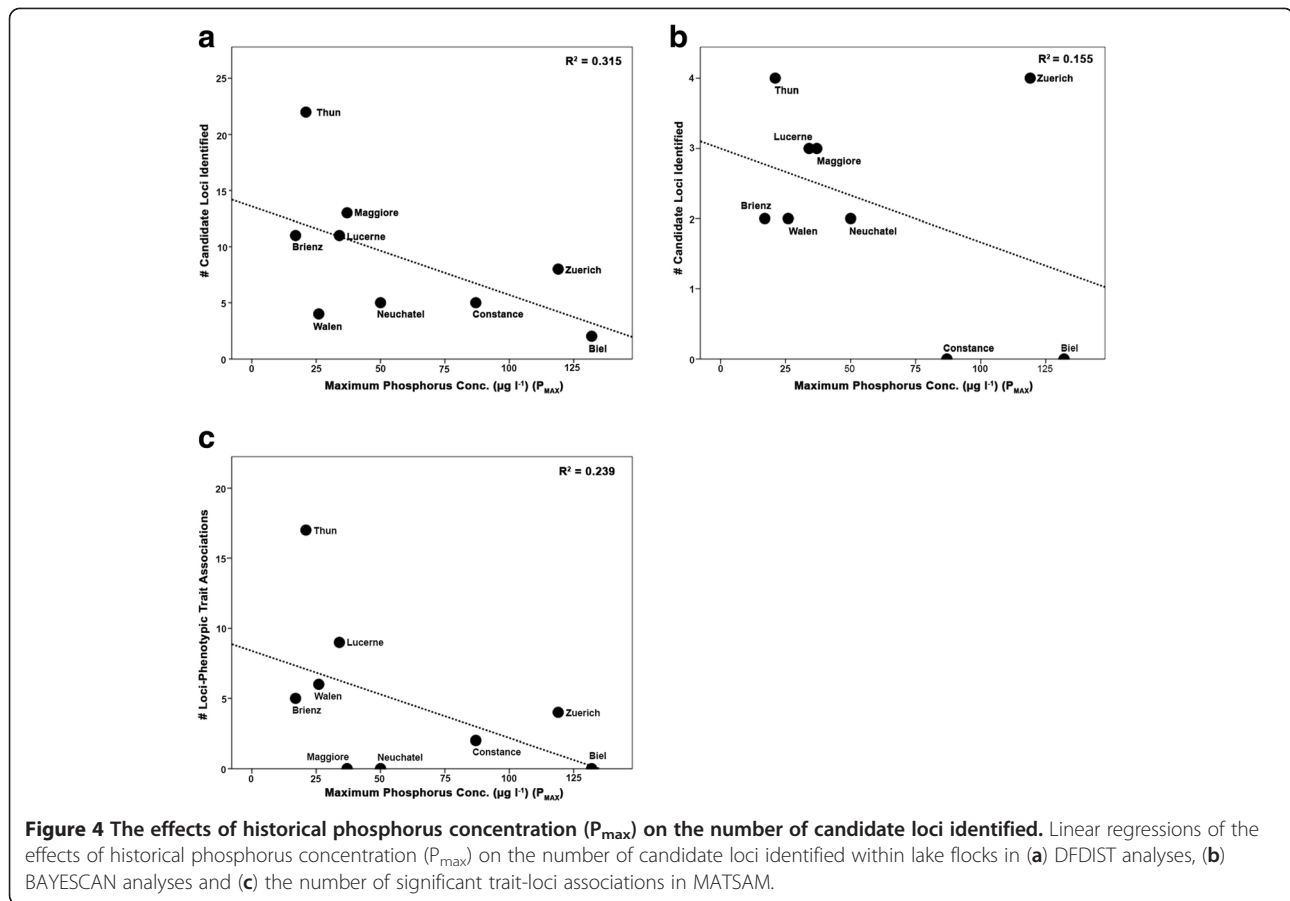
#### Tests of association between neutral genomic divergence and eutrophication

In Linear regressions,  $P_{max}$  values were strongly negatively correlated with current lake flock-wide global  $F_{ST}$

for neutral microsatellites, all AFLP markers and the neutral AFLP subsets ( $r^2 = 0.38-0.65$ ,  $P = 0.008-0.079$ ; Table 4, Figure 5). Levels of Min. O<sub>2</sub> were significantly positively correlated with lake flock-wide global  $F_{ST}$  and explained even more variation than  $P_{max}$  ( $r^2 = 0.56-0.76$ ,  $P = 0.005-0.033$ ; Additional file 11: Figure S4). Estimates of  $F_{ST}$  from all three datasets were highly positively correlated across lakes ( $r = 0.906-0.975$ , d.f. = 8,  $P < 0.001$ ).

#### Discussion

During early stages of ecological speciation, gene flow between incipient species in sympatry/parapatry is expected to be restricted mainly at loci underlying adaptive traits and associated physically linked loci: resulting in increased levels of genetic differentiation in these regions of the genome, relative to unlinked loci evolving under neutrality. In this study we find candidates for loci evolving under disruptive natural selection in each of the examined lake species flocks of the Alpine *Coregonus* adaptive radiation. We also find significant associations between many of these candidate loci and putative adaptive traits, supporting the action of disruptive natural selection in the origin and coexistence of species within this radiation. Some of the lakes have recently undergone anthropogenic homogenization of whitefish habitat through eutrophication, and all analysis methods show negative trends in the number of disruptive selection candidate loci (DFDIST, BAYESCAN) and the number of significant loci-adaptive trait associations (MATSAM) found among coexisting species, in relation to levels of historical eutrophication within the lakes the whitefish species flocks inhabit. This is congruent with patterns of genetic differentiation shown by neutral microsatellite and AFLP loci in the same species flocks, indicating the



weakening of reproductive isolation during eutrophication. We hypothesize that the pattern of reduced numbers of candidate loci found in species flocks that have experienced more severe anthropogenic eutrophication, could be due to two distinct but non-exclusive genomic processes brought about by increased interspecific hybridization and a concomitant increase in interspecific recombination. In the first, increased interspecific recombination erodes genomic islands of elevated genetic differentiation surrounding the target loci of disruptive natural selection, akin to a reversal of divergence hitchhiking, reducing the likelihood that AFLP restriction sites will fall within these regions. In the second, large chromosomal segments, including intact genomic islands of divergence, pass between the genomes of hybridizing species, equalizing allele frequencies and reducing heightened  $F_{ST}$  levels associated with genomic islands and therefore reducing the power to detect outlier candidate loci within those segments. Our results presented here provide the first genomic evidence that eutrophication may lead to the relaxation of divergent selection that in turn permits increased levels of interspecific, introgressive hybridization, driving speciation reversal.

#### Evidence for the role of disruptive natural selection in shaping intralacustrine diversity

Previous phylogenetic and population genetic analyses have shown that the Alpine *Coregonus* radiation is monophyletic and that its constituent lake flocks form a nested series of independent radiations that arose rapidly, *in-situ*, within the last 15,000 years [25]. During the course of an adaptive radiation, disruptive natural and sexual selection regimes experienced by species in contrasting environments drive adaptive divergence and speciation [59]. Given the co-occurrence of multiple, ecologically divergent species within the Alpine lake flocks and the overall youth of the radiation, heterogeneous genomic divergence between divergent species is expected [14]. Natural selection against hybrids with trait values or trait combinations maladaptive in either parental environment will maintain stronger differentiation in the face of gene flow in genomic islands of elevated interspecific genetic differentiation surrounding loci under disruptive selection than in regions invisible to selection [12]. We find candidate loci displaying unusually strong patterns of genetic differentiation between paired coexisting species in each of the species flock studied.

**Table 4 Maximum within-lake  $F_{ST}$  values among sympatric species derived from different genetic markers**

Species Flock	AFLP		Microsatellite
	All	Neutral	
Neuchâtel	0.077	0.028	0.05*
Biel	0.073	0.035	0.024†
Thun	0.153	0.107	0.252†
Brienz	0.229	0.185	0.269†
Lucerne	0.123	0.074	0.13*
Constance	0.087	0.039	0.071*
Walen	0.095	0.049	0.13*
Zuerich	0.095	0.045	0.044*
Maggiore	0.173	0.097	0.182‡

Comparison of the maximum within-lake pairwise  $F_{ST}$  values for each lake flock. All =  $F_{ST}$  value derived using all genotyped AFLP loci. Neutral =  $F_{ST}$  value derived following removal of all AFLP selection candidate loci. Data taken from: \* Vonlanthen *et al.* (2012), † Bittner (2009), ‡ Hudson *et al.* (unpublished observations).

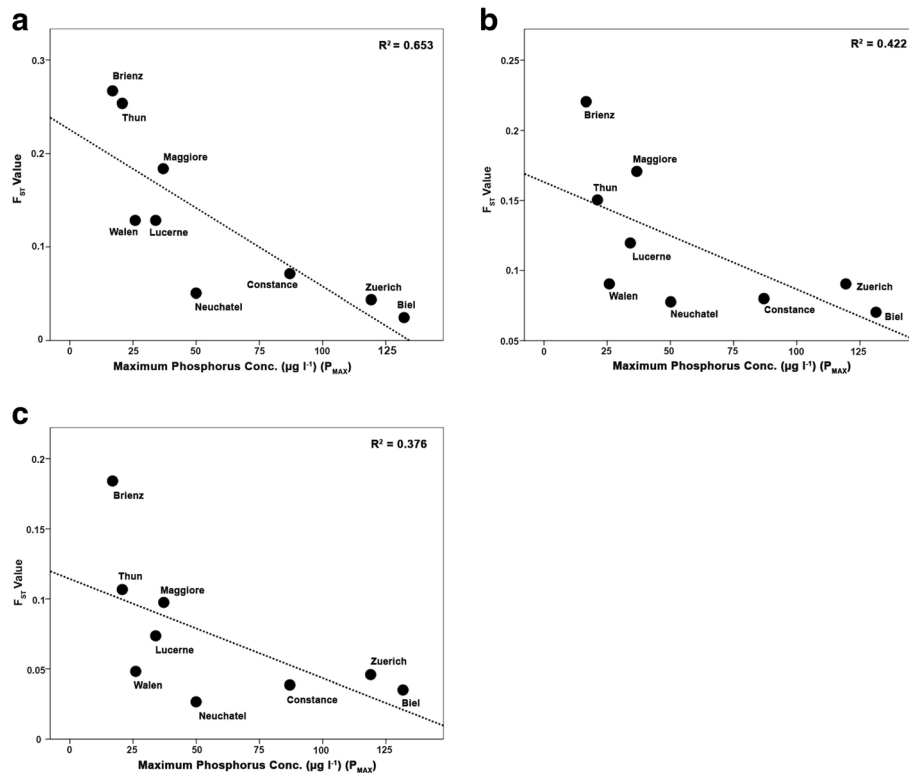
Given the recent, common origin of the Alpine radiation, parallel trends in candidate loci between similar ecotypes in different lake flocks could be expected. Shared standing genetic variation and similar levels of genetic co-variation among traits increases the likelihood that the same loci and alleles are involved in the evolution of equivalent phenotypes in different lake flocks [60]. We find support for this notion, albeit the number of loci that are repeatedly detected as outliers and/or associated with putative adaptive traits is moderate. Also there is no obvious increase in the phylogenetic grouping of species by ecotype in trees constructed solely from identified outlier loci, and matrices coding for ecotype designation in Mantel tests explained non-significant levels of variation in interspecific genetic distances generated solely from identified outlier loci: though still more than the variation explained in tests using neutral AFLP loci. The moderate number of parallel allelic trends likely reflects that many of the candidate loci we identified, are linked with varying degrees of physical proximity to loci under disruptive natural selection, rather than being the target loci of disruptive selection themselves [13,61]. During speciation within the independent whitefish species flocks, the action of recombination, eroding blocks of linkage disequilibrium along the genome, and the time since origination of the beneficial alleles, could mean that different genetic backgrounds containing different hitchhiker AFLP outlier loci are driven to, or near to fixation in similar ecotypes, among species flocks: despite the true target locus of selection, that hitchhiking loci are physically linked too, being identical in each paralleled speciation event [60]. Therefore parallel trends in candidate loci between similar ecotypes would likely be restricted to hitchhiker loci very

tightly linked to the actual loci under selection and/or parallel population divergences of recent, shared evolutionary origins.

From our study, locus CAGC-235 showed parallel allelic trends in more than two lake systems. The presence allele for this locus has a reduced prevalence in lower gill-raker ecotypes in lakes Maggiore, Thun (*C. fatioides* only), Zuerich, Lucerne and Brienz (*C. sp.* "balchen" only). Therefore there is the potential that the alternative allele for CAGC-235 is tightly linked to the actual target locus for natural selection. Five of the remaining loci support the role of recombination in breaking down the physical linkage between homologous candidate loci and the target of selection over evolutionary time: the parallel trends in these loci being shared only between the most recently separated lake flocks. Three loci were found in both lakes Thun and Brienz (CGAG-186, TGTG-183, CTAC-99), which form part of the same monophyletic super-lake radiation and given the geographical proximity of the lakes, were likely the last of the studied species flocks to separate. CGAG-355 and TGTG-271 show parallel allelic trends between introduced L. Maggiore species and their likely source populations in either lakes Lucerne and Constance [25].

#### Anthropogenic perturbations and heterogeneous genomic divergence

The genomic extent of levels of increased genetic differentiation around the target loci for disruptive natural selection will depend on the balance between the strength of selection and gene flow [12,13]. Two distinct, but to some degree complementary, mechanisms have been proposed to describe how genetic differentiation spreads through the genomes of species diverging under speciation-with-gene-flow scenarios. Under divergence hitchhiking (DH), strong disruptive selection acting on a locus reduces interspecific recombination over relatively restricted areas of the genome in physical linkage with the selective locus. In this way beneficial alleles under weaker selection can be recruited for adaptive divergence through their "uplift" onto genomic islands of divergence surrounding more strongly selected loci. This reduces the chances of their loss to interspecific recombination and facilitates the growth of genomic islands of divergence [13,62]. Under genome hitchhiking (GH), disruptive selection reduces the average effective migration across the entire genome, thereby facilitating differentiation throughout the genome irrespective of levels of physical linkage to selective loci [63,64]. Feder, Egan & Nosil [64] describe a four-phase model of speciation-with-gene-flow. In this model, DH is most important in facilitating genomic differentiation at early to intermediate stages of speciation (phase 2). At later stages, as more loci are recruited by disruptive natural selection, GH will become the dominant mechanism



**Figure 5** The effects of historical phosphorus concentration ( $P_{max}$ ) on overall lake flock genetic differentiation. Linear regressions of the effects of historical phosphorus concentration ( $P_{max}$ ) on within-lake genetic differentiation among species in different sets of genetic marker loci: (a) neutral microsatellite loci, (b) all AFLP loci, (c) neutral AFLP loci.

driving genomic differentiation (phase 3). The relatively recent origin of the species flocks of the Alpine *Coregonus* radiation (<15,000 years) suggests that DH could play an important role in the divergence seen between species. Under DH, genomic islands of divergence around selective loci are expected to increase in size as speciation progresses. This in turn would increase the likelihood that individual AFLP loci will fall within these areas of higher genetic differentiation and be classed as outlier candidate loci. Therefore as genomic divergence progresses under DH, the proportion of outlier candidate loci amongst a random genomic sample of AFLP loci should increase: relative to a similar genome scan conducted at an earlier time point in the speciation process of the same species pair [12,13]. This prediction depends on the existence of relatively extensive genomic regions of low interspecific recombination around selected loci. In dwarf/normal species pairs in the closely related *Coregonus clupeaformis* species complex, effective migration was found to be reduced for relatively large distances around selective loci ( $\approx 5$  cM) and that the extent of phenotypic divergence between coexisting species pairs was positively correlated with levels of linkage disequilibrium around these loci [48,65]. However, in the species flocks of the Alpine whitefish radiation more genomic research will be needed to

characterize the actual target loci of selection and define the effective limits of surrounding islands of divergence, taking into account factors which could potentially enhance (e.g. epistasis) or constrain (e.g. selection from standing genetic variation) the size of these genomic areas: before extensive DH can be confirmed [13,63].

In the lakes of the Alpine *Coregonus* radiation, previous research has shown that historical levels of anthropogenic nutrient enrichment negatively correlated with contemporary interspecific genetic differentiation at neutral markers, and positively correlated with the overall loss of species and functional phenotypic diversity. Direct comparisons of neutral genetic differentiation from samples taken from the same species flocks, pre- and post- (large-scale) eutrophication, corroborate this pattern, with strong declines found in a highly eutrophied lake relative to more lightly polluted lakes, despite similar levels of genetic differentiation occurring in these species flocks prior to the peak of eutrophication [8]. Our present study corroborates these results, with genetic differentiation at selectively neutral subsets of AFLP loci also being strongly negatively correlated with levels of anthropogenic nutrient enrichment. Focusing on areas of the genome evolving under the influence of disruptive natural selection, we hypothesized that the

numbers of candidate loci identified within a lake flock would be also negatively correlated with each lake's historical maximum level of eutrophication. As the lake habitats become increasingly homogeneous during eutrophication, disruptive selection between habitats would wane, reducing the efficacy of barriers to interspecific gene flow and resulting in increased interspecific recombination that would, in turn, either reduce the size of genomic islands of divergence, reversing divergence hitchhiking or result in the transfer of entire chromosomal segments containing intact genomic islands between the genomes of hybridizing species, equalizing allele frequencies at potential  $F_{ST}$  outlier loci. However, whilst we find that numbers of candidate loci in lakes with less strong eutrophication histories were indeed high (lakes Thun, Brienz, Lucerne) and were lowest in the most eutrophic lake (Biel), the relationship with phosphorus was marginally non-significant. This was due to certain lakes having fewer candidate loci (Neuchâtel, Walen) and others lakes having more (Zuerich), than would be predicted given their past levels of eutrophication.

Variation in the levels of ecological divergence between coexisting species, among species flocks, prior to environmental perturbations could potentially affect the relative numbers of candidate loci found among lake flocks today. If ecological divergence and therefore development of RI was impeded by high levels of gene flow, weaker disruptive selection and/or temporal fluctuations in selection regimes, this would affect the number and extent of genomic islands of divergence that occur between coexisting species naturally and how susceptible they are to environmental disturbance when it occurs and subsequently the extent of speciation reversal, relative to similarly aged species flocks from more ecologically divergent/stable lake environments [4,66]. The two pairs of coexisting species from recently connected lakes: Walen and Zuerich, and Biel and Neuchâtel, represent two distinct super-lake clades within the overall Alpine radiation [25]. Lakes Zuerich and Biel have a similar history of high eutrophication ( $P_{max} = 119 \mu\text{g/l}$  and  $132 \mu\text{g l}^{-1}$  respectively), L. Neuchâtel moderate ( $50 \mu\text{g/l}$ ) and L. Walen low ( $26 \mu\text{g/l}$ ). Using gill raker number as a surrogate for ecological diversity among coexisting species, L. Zuerich had a greater pre-eutrophication difference in average gill-raker number between species (26.7 & 35.2) than lakes Biel (27.4 & 33.2) or Neuchâtel (27.5 & 32.2), similar to L. Walen (25.8 & 35.4)[8,29]. Following heavy eutrophication, despite evidence of increased neutral gene flow and low microsatellite  $F_{ST}$  values similar to Biel and Neuchâtel [8], a higher number of candidate loci was found and relatively high phenotypic differentiation has been maintained between L. Zuerich species, despite much more severe levels of pollution

relative to Neuchâtel. Additional support for this hypothesis may come from the closely related, North American *C. clupeaformis*-complex. The coexisting *Coregonus* species of Indian pond show high levels of phenotypic differentiation and the largest number of AFLP candidate loci yet found in the *C. clupeaformis* radiation, yet have relatively low neutral genetic differentiation [18,67]. Intriguingly, Indian Pond has phosphorus concentrations three times higher than other studied North American *Coregonus* lakes and the highest post-summer decreases in oxygen with depth, suggesting this lake and its endemic *Coregonus* species may have experienced eutrophication in its recent past [68].

In pristine environments, as the number of species within a radiation increases so should the number of selection candidate loci and trait associated loci identified, as the number of axes of divergent adaptation among coexisting species increases [65]. In the Lake Thun species flock, five species coexist and the highest numbers of candidate loci/trait associated loci found in this study were identified within this species flock. In contrast, the similarly species-rich, yet heavily polluted L. Constance species flock had some of the lowest numbers of candidate loci/trait associated loci identified in the Alpine radiation. L. Constance previously held up to six coexisting species, but following eutrophication, the VLGR ecotype, *C. gutturosus* went extinct and among the remaining species a large decrease in interspecific neutral genetic differentiation occurred [8]. This may illustrate a potential trend in speciation reversal, in that young lake radiations with higher species diversity are intrinsically more vulnerable, in response to a given level of environmental change, to speciation reversal than less species-rich systems. Within evolutionarily young, species-rich radiations, ecological and reproductive niche breadths may be narrower and more tightly packed compared to related, less diverse radiations, perhaps through partial subdivision of the ancestral niche: potentially making them more sensitive to relaxation of disruptive natural selection regimes. As environmental factors imparting RI begin to break down the increased potential directions of gene flow between multiple coexisting species could in turn exacerbate speciation reversal. This hypothesis could help to explain rapid collapses in diversity within other young radiations such as the Lake Victoria cichlids [5] and Great Lake ciscoes [9].

The strategy we have adopted for the present study was to investigate all possible Alpine whitefish species flocks, including most of the endemic species present: each species flock comprising a distinct data point along a spectrum of anthropogenic habitat perturbation from low to severe. Overall, the limited number of individuals per species included in our analyses certainly places limitations on the power to identify, among the set of loci

sampled, all those that might be evolving under disruptive selection between coexisting species. Also, whilst mean values of linkage disequilibrium were not higher among identified outlier loci than among neutral loci, suggesting that detected outliers have a rather dispersed distribution in the genomes of Alpine whitefish species, the true number of separate linkage groups that our candidate loci comprise remains unknown. Additionally, we cannot rule out entirely that another, unidentified ecological factor, covarying with the extent of eutrophication, may be a stronger driver of the pattern seen in the number of outlier loci identified amongst Swiss whitefish species flocks. In this regard, our study stands as a necessary first step, allowing future studies to home in on the actual target loci of disruptive selection and their function, to uncover the genomic architecture of the phenotypic traits they affect and to measure the fitness effects of alleles at these loci within the alternate niches of the different species, and how eutrophication may affect these. Given the number and diversity of species flocks covered by our sampling scheme, which explicitly takes into account the hierarchical structuring of genetic variation within the Alpine whitefish radiation; we have the power to look at the relative numbers of candidate loci identified and formulate hypotheses as to the factors influencing variation in this number. Therefore despite the fact that while some of the individual candidate loci identified in this study, may have elevated genetic differentiation through processes other than contemporary disruptive natural selection, such as historical selective sweeps [69], demography [70] or their location in chromosomal regions with reduced recombination [21]; the recent, common origin of the species flocks of the Alpine whitefish radiation means that these potential biases are very unlikely to lead to the trends in candidate loci we find across lake flock data sets. Furthermore by using two different  $F_{ST}$  outlier detection methods, alongside logistic regressions, combining population-based and individual-based approaches respectively, not only can we identify adaptive traits important in niche differentiation and the loci potentially underpinning them, but also that the high congruence in the results produced by these different methodologies suggests that the overall patterns in candidate loci revealed by our study are real.

## Conclusions

Within each of the postglacial species flocks of the Alpine whitefish radiation we find evidence for the action of disruptive natural selection in the shaping of within-lake species diversity through the identification of  $F_{ST}$  outlier loci and/or significant adaptive trait/candidate loci associations. We also find, similar to the extent of interspecific genetic differentiation at neutral loci, a

negative correlation between the number of identified  $F_{ST}$ -outlier loci identified and significant adaptive trait/candidate loci associations, and historical levels of anthropogenic nutrient enrichment within each lake. We hypothesize that the habitat disruption caused by eutrophication weakens disruptive natural selection regimes resulting in increased gene flow between species. The resulting interspecific recombination either breaks down genomic islands of heightened genetic differentiation surrounding the target loci of disruptive natural selection or entire chromosomal segments, including intact genomic islands, pass between species' genomes, homogenizing allele frequencies. Our study presents the first genomic evidence that speciation reversal is associated with the relaxation of disruptive selection: contributing to our understanding of the mechanisms that control this important driver of biodiversity loss.

## Additional files

**Additional file 1: Table S1.** Individual phenotypic trait values.

**Additional file 2: Table S2.** AFLP primer pair information.

**Additional file 3: Table S3.** Identity of AFLP loci classified as  $F_{ST}$  outliers following DFDIST analyses.

**Additional file 4: Table S4.** Identity of AFLP loci classified as  $F_{ST}$  outliers following BAYESCAN analyses.

**Additional file 5: Figure S1.** Population-based NJ consensus trees created using different subsets of AFLP loci.

**Additional file 6: Table S5.** Mean pairwise LD values from neutral and outlier loci classes for each whitefish species.

**Additional file 7: Figure S2.** Distribution of pairwise linkage disequilibrium values for neutral and outlier loci classes within each whitefish species.

**Additional file 8: Table S6.** Identity of AFLP loci with significant associations with putative adaptive phenotypic traits identified by MATSAM.

**Additional file 9: Table S7.** Congruence between Global dataset evidence for disruptive selection and association with adaptive traits within species flocks.

**Additional file 10: Figure S3.** Linear regressions of the effects of minimum lake oxygen concentration at depth (Min.  $O_2$ ) on the number of candidate loci identified within lake flocks in (a) Dfdist analyses, (b) BayeScan analyses and (c) the number of significant trait-loci associations in MatSAM.

**Additional file 11: Figure S4.** Linear regressions of the effects of minimum oxygen concentration at maximum depth (Min.  $O_2$ ) on genetic differentiation in different sets of loci: (a) neutral microsatellite loci, (b) all AFLP loci (c) 'neutral' AFLP loci.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

AGH wrote the paper, contributed to conception and design of the study, sampled fish, generated adaptive trait and genetic data, and carried out most of the subsequent analyses. PV sampled fish, generated environmental lake data and contributed to analyses and writing. EB contributed to analyses and writing. OS conceived and designed the project, supervised the project, and contributed to data analyses and writing. All authors read and approved the final manuscript.



## Acknowledgements

Many thanks to R. Mueller, D. Bittner, B. Lundsgaard-Hansen, M. Kugler, J. Muggli, B. Polli, P. Volta, D. Gerdaux and A. Champigneulle for assisting in sampling. Also thanks to K. Lucek, S. Mwaiko, A. Sivasundar, V. Schneider, D. Joyce, C. Klingenberg and B. M. Bolker for technical advice and assistance. Constructive suggestions from D. Roy, M. Barluenga, I. Magalhaes and P. Nosil greatly improved this manuscript. This study was funded by the University of Bern and the EAWAG Handlungsfeld project: Aquadiverse.

## Author details

<sup>1</sup>Division of Aquatic Ecology & Macroevolution, Institute of Ecology & Evolution, University of Bern, Baltzerstrasse 6, CH 3012 Bern, Switzerland. <sup>2</sup>Department of Fish Ecology & Evolution, EAWAG Centre of Ecology, Evolution and Biogeochemistry, Seestrasse 79, CH 6047 Kastanienbaum, Switzerland. <sup>3</sup>Department of Ecology & Environmental Science, Umeå University, SE 901 87 Umeå, Sweden. <sup>4</sup>Department of Biology, Reed College, 3203 SE Woodstock Blvd, Portland, OR 97202, USA.

Received: 19 March 2013 Accepted: 20 May 2013

Published: 30 May 2013

## References

- Smith TB, Bernatchez L: Evolutionary change in human-altered environments. *Mol Ecol* 2008, **17**:1–8.
- Butchart SHM, Walpole M, Collen B, van Strien A, Scharlemann JPW, Almond REA, Baillie JEM, Bomhard B, Brown C, Bruno J, Carpenter KE, Carr GM, Chanson J, Chenery AM, Csirke J, Davidson NC, Dentener F, Foster M, Galli A, Galloway JN, Genovesi P, Gregory RD, Hockings M, Kapos V, Lamarque J-F, Leverington F, Loh J, McGeoch MA, McRae L, Minasyan A, Morcillo MH, Oldfield TEE, Pauly D, Quader S, Revenga C, Sauer JR, Skolnik B, Spear D, Stanwell-Smith D, Stuart SN, Symes A, Tierney M, Tyrrell TD, Vié J-C, Watson R: Global biodiversity: indicators of recent declines. *Science* 2010, **328**:1164–1168.
- Pereira HM, Leadley PW, Proença V, Alkemade R, Scharlemann JPW, Fernandez-Manjarrés JF, Araújo MB, Balvanera P, Biggs R, Cheung WWL, Chini L, Cooper HD, Gilman EL, Guénette S, Hurtt GC, Huntington HP, Mace GM, Oberdorff T, Revenga C, Rodrigues P, Scholes RJ, Sumaila UR, Walpole M: Scenarios for global biodiversity in the 21st century. *Science* 2010, **330**:1496–1501.
- Seehausen O: Conservation: Losing biodiversity by reverse speciation. *Curr Biol* 2006, **16**:334–337.
- Seehausen O, van Alphen JJM, Witte F: Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* 1997, **277**:1808–1811.
- Taylor EB, Boughman JW, Groenenboom M, Sniatynski M, Schluter D, Gow JL: Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Mol Ecol* 2006, **15**:343–355.
- Hendry AP, Farrugia TJ, Kinnison MT: Human influences on rates of phenotypic change in wild animal populations. *Mol Ecol* 2008, **17**:20–29.
- Vonlanthen P, Bittner D, Hudson AG, Young KA, Müller R, Lundsgaard-Hansen B, Roy D, Di Piazza S, Largiadier CR, Seehausen O: Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature* 2012, **482**:357–362.
- Todd TN, Stedman RN: Hybridization of ciscoes (*Coregonus* spp.) in Lake Huron. *Can J Zool* 1989, **67**:1679–1685.
- Wu CI: The genic view of the process of speciation. *J Evol Biol* 2001, **14**:851–865.
- Turner TL, Hahn MW, Nuzdin SV: Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol* 2005, **3**:e285.
- Nosil P, Funk DJ, Ortiz-Barrientos D: Divergent selection and heterogeneous genomic divergence. *Mol Ecol* 2009, **18**:375–402.
- Via S, West J: The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol Ecol* 2008, **17**:4334–4345.
- Coyne JA, Orr HA: *Speciation*. Sunderland, MA: Sinauer Associates, Inc.; 2004.
- Mallet J, Meyer A, Nosil P, Feder JL: Space, sympatry and speciation. *J Evol Biol* 2009, **22**:232–2341.
- Feder JL, Gejji R, Yeaman S, Nosil P: Establishment of new mutations under divergence hitchhiking and genome hitchhiking. *Phil. Trans. R. Soc. B* 2012, **367**:461–474.
- Wilding CS, Butlin RK, Grahame J: Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *J Evol Biol* 2001, **14**:611–619.
- Campbell D, Bernatchez L: Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Mol Biol Evol* 2004, **21**:945–956.
- Bonin A, Taberlet P, Miaud C, Pompanon F: Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Mol Biol Evol* 2006, **23**:773–783.
- Nosil P, Egan SP, Funk DJ: Heterogeneous genomic differentiation between walking-stick ecotypes: 'isolation by adaptation' and multiple roles for divergent selection. *Evolution* 2008, **62**:316–336.
- Roesti M, Hendry AP, Salzburger W, Berner D: Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Mol Ecol* 2012, **21**:2852–2862.
- Kottelat M, Freyhof J: *Handbook of European freshwater fishes*. Cornol, Switzerland: Publications Kottelat; 2007.
- Douglas MR, Brunner PC, Bernatchez L: Do assemblages of *Coregonus* (Teleostei: Salmoniformes) in the Central Alpine region of Europe represent species flocks? *Mol Ecol* 1999, **8**:589–603.
- Hudson AG, Vonlanthen P, Müller R, Seehausen O: Review: The geography of speciation and adaptive radiation in Coregonines. *Adv Limnol* 2007, **60**:111–146.
- Hudson AG, Vonlanthen P, Seehausen O: Rapid parallel adaptive radiations from a single hybridogenic ancestral population. *Proc R Soc B* 2011, **278**:58–66.
- Steinmann P: Monographie der Schweizerischen Koregonen. Beitrag zum Problem der Entstehung neuer Arten. Spezieller Teil. *Schweiz Z Hydrol* 1950, **12**:340–491.
- Burgi HR, Heller C, Gaebel S, Mookerji N, Ward JV: Strength of coupling between phyto- and zooplankton in Lake Lucerne (Switzerland) during phosphorus abatement subsequent to a weak eutrophication. *J Plankton Res* 1999, **21**:485–507.
- Gruber N, Wehrli B, Wuest A: The role of biogeochemical cycling for the formation and preservation of varved sediments in Soppensee (Switzerland). *J Paleolimnol* 2000, **24**:277–291.
- Muller R, Stadelmann P: Fish habitat requirements as the basis for rehabilitation of eutrophic lakes by oxygenation. *Fisheries Manag. Ecol.* 2004, **11**:251–260.
- Nümann W: The Bodensee: effects of exploitation and eutrophication on the Salmonid community. *J. Fish. Res. Board Can.* 1972, **29**:833–847.
- Vonlanthen P, Roy D, Hudson AG, Largiadier CR, Bittner D, Seehausen O: Divergence along a steep ecological gradient in lake whitefish (*Coregonus* sp.). *J Evol Biol* 2009, **22**:498–514.
- Bernatchez L: Ecological Theory of Adaptive Radiation. An Empirical Assessment from Coregonine Fishes (Salmoniformes). In *Evolution Illuminated*. Edited by Hendry AP, Stearns SC. Oxford: Oxford University Press; 2004:175–207.
- Berg A, Grimaldi E: Ecological relationships between planktophagic fish species in the Lago Maggiore. *Verh. Int. Ver. Theor. Angew. Limnol.* 1966, **16**:1065–1073.
- Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M: AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* 1995, **23**:4407–4414.
- Bonin A, Bellemain E, Eidesen PB, Pompanon F, Brochmann C, Taberlet P: How to track and assess genotyping errors in population genetics studies. *Mol Ecol* 2004, **13**:3261–3273.
- Thornton KR, Jensen JD: Controlling the false-positive rate in multilocus genome scans for selection. *Genetics* 2007, **175**:737–750.
- Excoffier L, Hofer T, Foll M: Detecting loci under selection in a hierarchically structured population. *Heredity* 2009, **103**:285–298.
- Beaumont MA, Balding DJ: Identifying adaptive genetic divergence among populations from genome scans. *Mol Ecol* 2004, **13**:969–980.
- Foll M, Gaggiotti O: A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* 2008, **180**:977–993.
- Vekemans X: AFLP-SURV v. 1.0. *Laboratoire de Génétique et Ecologie Végétale, Université Libre de Bruxelles*. Distributed by the author: Bruxelles, Belgium; 2002.
- Felsenstein J: PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* 1989, **5**:164–166.
- Allender CJ, Seehausen O, Knight ME, Turner GF, Maclean N: Divergent selection during speciation of Lake Malawi cichlid fishes inferred from parallel radiations in nuptial coloration. *P. Natl. Acad. Sci. USA* 2003, **100**:14074–14079.

43. Excoffier L, Lischer HEL: Arelquin suite ver 3.5: A new series of programs to perform populaion genetic analyses under Linux and Windows. *Mol Ecol Resour* 2010, **10**:564–567.
44. Agapow P-M, Burt A: Indices of multilocus linkage disequilibrium. *Mol. Ecol. Notes*. 2001, **1**:101–102.
45. Cushing DH: Plankton production and year-class strength in fish populations – an update of the match mismatch hypothesis. *Adv Mar Biol* 1990, **26**:249–293.
46. Svensson E, Sinervo B: Experimental excursions on adaptive landscapes: Density-dependent selection on egg size. *Evolution* 2000, **54**:1396–1403.
47. Sogard SM: Size-selective mortality in the juvenile stage of teleost fishes: A review. *B. Mar. Sci.* 1997, **60**:1129–1157.
48. Woods P, Müller R, Seehausen O: Intergenomic epistasis causes asynchronous hatch times in whitefish hybrids, but only when parental ecotypes differ. *J Evol Biol* 2009, **22**:2305–2319.
49. Rogers SM, Bernatchez L: 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* 2007, **24**:1423–1438.
50. Rogers SM, Gagnon V, Bernatchez L: Genetically based phenotype-environment association for swimming behavior in lake whitefish ecotypes (*Coregonus clupeaformis* Mitchill). *Evolution* 2002, **56**:2322–2329.
51. Weaver MJ, Magnuson JJ, Clayton MK: Habitat heterogeneity and fish community structure: Inferences from north temperate lakes. *Am. Fish S. S.* 1996, **16**:335–346.
52. Kahilainen K, Ostbye K: Morphological differentiation and resource polymorphism in three sympatric whitefish *Coregonus lavaretus* (L.) forms in a subarctic lake. *J Fish Biol* 2006, **68**:63–79.
53. Harrod C, Mallela J, Kahilainen KK: Phenotype-environment correlations in a putative whitefish adaptive radiation. *J Anim Ecol* 2010, **79**:1057–1068.
54. Chouinard A, Pigeon D, Bernatchez L: Lack of specialization in trophic morphology between genetically differentiated dwarf and normal forms of lake whitefish (*Coregonus clupeaformis* Mitchill) in Lac de l'Est. *Quebec. Can. J. Zool.* 1996, **74**:1989–1998.
55. McCoy MW, Bolker BM, Osenberg CW, Miner BG, Vonesh JR: Size correction: comparing morphological traits among populations and environments. *Oecologia* 2006, **148**:547–554.
56. Klingenberg CP: MorphoJ: an integrated software package for geometric morphometrics. *Mol. Ecol. Res.* 2011, **11**:353–357.
57. Joost S, Bonin A, Bruford MW, Despres L, Conord C, Erhardt G, Taberlet P: A spatial analys method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Mol Ecol* 2007, **16**:3955–3969.
58. Bittner D: *Gonad deformations in whitefish (Coregonus spp.) from Lake Thun, Switzerland - A population genetic and transcriptomic approach*, PhD Thesis. Switzerland: University of Bern; 2009.
59. Schluter D: *The ecology of Adaptive Radiation*. New York: Oxford University Press; 2000.
60. Barrett RD, Schluter D: Adaptation from standing genetic variation. *Trends Ecol Evol* 2008, **23**:38–44.
61. Butlin RK: Population genomics and speciation. *Genetica* 2010, **138**:409–418.
62. Via S, Conte G, Mason-Foley C, Mills K: Localizing  $F_{ST}$  outliers on a QTL map reveals evidence for large scale genomic regions of reduced gene exchange during speciation-with-gene-flow. *Mol Ecol* 2012, **21**:5546–5560.
63. Feder JL, Gejji R, Yeaman S, Nosil P: Establishment of new mutations under divergence and genome hitchhiking. *Philos Trans. R Soc. B: Biol Sci.* 2012, **367**:461–474.
64. Feder JL, Egan SP, Nosil P: The genomics of speciation-with-gene-flow. *Trends Genet* 2012, **28**:303–306.
65. Gagnaire P-A, Pavay SA, Normandeau E, Bernatchez B: The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD-sequencing. *Evolution* 2013. doi:10.1111/evo.12075.
66. Nosil P, Harmon L, Seehausen O: Ecological explanations for (incomplete) speciation. *Trends Ecol Evol* 2009, **24**:145–156.
67. Lu G, Bernatchez L: Correlated trophic specialization and genetic divergence in sympatric lake whitefish ecotypes (*Coregonus clupeaformis*): support for the ecological speciation hypothesis. *Evolution* 1999, **53**:1491–1505.
68. Landry L, Vincent WF, Bernatchez L: Parallel evolution of lake whitefish dwarf ecotypes in association with limnological features of the adaptive landscape. *J Evol Biol* 2007, **20**:971–984.
69. Bierne N: The distinctive footprints of local hitchhiking in a varied environment and global hitchhiking in a subdivided population. *Evolution* 2010, **64**:3254–3272.
70. Charlesworth B, Charlesworth D, Barton NH: The effects of genetic and geographic structure on neutral variation. *Annu Rev Ecol Syst* 2003, **34**:99–125.

doi:10.1186/1471-2148-13-108

**Cite this article as:** Hudson et al.: Genomic signatures of relaxed disruptive selection associated with speciation reversal in whitefish. *BMC Evolutionary Biology* 2013 **13**:108.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

