

# Genomic signatures of divergent selection and speciation patterns in a 'natural experiment', the young parallel radiations of Nicaraguan crater lake cichlid fishes

ANDREAS F. KAUTT,\*† KATHRYN R. ELMER\* and AXEL MEYER\*†‡

\*Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, Universitätsstrasse 10, 78457, Konstanz, Germany, †International Max Planck Research School for Organismal Biology, University of Konstanz, Universitätsstrasse 10, 78457, Konstanz, Germany, ‡Konstanz Research School Chemical Biology (KoRS-CB), University of Konstanz, Universitätsstrasse 10, 78457, Konstanz, Germany

## Abstract

Divergent selection is the main driving force in sympatric ecological speciation and may also play a strong role in divergence between allopatric populations. Characterizing the genome-wide impact of divergent selection often constitutes a first step in unravelling the genetic bases underlying adaptation and ecological speciation. The Midas cichlid fish (*Amphilophus citrinellus*) species complex in Nicaragua is a powerful system for studying evolutionary processes. Independent colonizations of isolated young crater lakes by Midas cichlid populations from the older and great lakes of Nicaragua resulted in the repeated evolution of adaptive radiations by intralacustrine sympatric speciation. In this study we performed genome scans on two repeated radiations of crater lake species and their great lake source populations (1030 polymorphic AFLPs,  $n \sim 30$  individuals per species). We detected regions under divergent selection (0.3% in the crater lake Xiloá flock and 1.7% in the older crater lake Apoyo radiation) that might be responsible for the sympatric diversifications. We find no evidence that the same genomic regions have been involved in the repeated evolution of parallel adaptations across crater lake flocks. However, there is some genetic parallelism apparent (seven out of 51 crater lake to great lake outlier loci are shared; 13.7%) that is associated with the allopatric divergence of both crater lake flocks. Interestingly, our results suggest that the number of outlier loci involved in sympatric and allopatric divergence increases over time. A phylogeny based on the AFLP data clearly supports the monophyly of both crater lake species flocks and indicates a parallel branching order with a primary split along the limnetic-benthic axis in both radiations.

**Keywords:** AFLP, genome scan, limnetic-benthic, parallel evolution, population genomics, sympatric speciation

Received 4 May 2012; revision received 4 July 2012; accepted 10 July 2012

## Introduction

The role of divergent selection in driving population divergence and speciation has recently received wide attention, as evidenced by the increasing number of studies investigating ecological speciation (reviewed in Nosil 2012). In ecological speciation populations

might become reproductively isolated because of ecologically based divergent natural selection (Schluter 2000) and this can occur in any geographical context (Rundle & Nosil 2005). Still, whereas phenotypic divergence and speciation in allopatry is not controversial, sympatric speciation has long been a matter of debate (reviewed in Coyne & Orr 2004). Although theoretical models have shown that the evolution of reproductive barriers in an initially panmictic population is possible (Dieckmann & Doebeli 1999;

Correspondence: Axel Meyer, Fax: +49 7531 883018; E-mail: axel.meyer@uni-konstanz.de

Kondrashov & Kondrashov 1999; Bolnick 2006; Gavrillets & Vose 2007; Gavrillets *et al.* 2007), so far few empirical case studies for sympatric divergence are widely accepted (reviewed in Via 2001; Bolnick & Fitzpatrick 2007). Yet, the investigation of divergence-with-gene-flow, including sympatric speciation, is currently an active topic of research (Nosil 2008; Pinho & Hey 2010; Smadja & Butlin 2011; Feder *et al.* 2012) and the empirical evidence for sympatric speciation may thus grow within the next few years. Theoretical models assign an essential role to disruptive (divergent) selection in most cases of sympatric speciation (Gavrillets 2003; Gavrillets *et al.* 2007), but still, many questions remain unanswered about the qualitative and quantitative nature of this type of selection on the genome. It remains unclear, for example, if multifarious selection on many loci, or stronger selection on only a few loci, is more likely to facilitate divergence and speciation (Nosil & Feder 2012). Moreover, more empirical data are needed to investigate whether one should typically expect a positive correlation between the number and/or size of diverging regions and the degree of phenotypic and overall neutral genetic population differentiation (e.g. Nadeau *et al.* 2012; Renaut *et al.* 2012; Roesti *et al.* 2012). It also remains to be discovered how frequently the same or different loci are responsible for parallel phenotypic evolution in closely related species (Elmer & Meyer 2011) and whether the genetic changes underlying adaptations mainly stem from standing variation or are acquired *de-novo* (Barrett & Schluter 2008; Stapley *et al.* 2010).

Population genomics approaches aim to detect genomic regions under divergent selection and can thereby be applied to investigate adaptation and speciation (reviewed in Butlin 2008; Nosil *et al.* 2009). These approaches rely on the expectation that divergent selection will result in differentially altered allele frequencies of adaptive loci, and regions linked to them (via genetic hitchhiking), between diverging populations. At the same time, differentiation at neutral loci will be impeded by the homogenizing effect of gene flow or insufficient time for the random effects of genetic drift (Lewontin & Krakauer 1973; Barton 2001; Wu 2001; Gavrillets & Vose 2005; reviewed in Storz 2005). This process has been termed 'heterogeneous genomic divergence' (Nosil *et al.* 2008, 2009) or the 'genetic mosaic of speciation' (Via & West 2008).

The role of natural selection in driving phenotypic divergence and speciation seems to be clearer in sympatric than in allopatric scenarios. This is because, without gene flow, random effects might contribute to divergence between allopatric populations (Butlin *et al.* 2012). In particular founder effects, triggered by the

colonization of a new environment by a small number of individuals, have been proposed to be an important factor causing phenotypic divergence and even speciation (Mayr 1954; Carson 1975; Templeton 1980). Alternatively, a newly colonized environment may differ from the source environment, resulting in population divergence mainly because of divergent selection (Schluter 2000). The relative importance of founder effects in speciation is controversial (see Barton & Charlesworth 1984; Carson & Templeton 1984 for opposing views). However, random and selective factors are not necessarily mutually exclusive and both may contribute to evolutionary divergence (e.g. Kolbe *et al.* 2012). Whereas genomic regions of extraordinarily high differentiation between species that originated in sympatry are assumed to be attributable to divergent selection, such signatures identified between allopatric species could also arise because of strong demographic effects such as founder effects (Foll & Gaggiotti 2008). In this context, parallel phenotypic and genetic changes across biological replicates provide a strong argument for selection rather than random effects (Bonin *et al.* 2006; Nosil *et al.* 2008). The investigation of a system that exhibits such natural replicates therefore provides a 'natural experiment' that will allow us to at least partly disentangle the effects of divergent selection and founder effects. Although the role of divergent selection in allopatric speciation is theoretically uncontroversial, more studies are needed to infer its prevalence (Nosil 2012).

Being one of the most species-rich vertebrate groups, the cichlid fish (family Cichlidae) have emerged as a model system for speciation research (Salzburger & Meyer 2004). Cichlids inhabiting the East African Great Lakes are, moreover, famous for their independent evolution of parallel phenotypes (Meyer 1993; Stiassny & Meyer 1999); a feature providing a framework for testing the role of natural selection in evolution (Elmer & Meyer 2011). Yet, the complexity of these radiations of cichlids and huge spatial dimension of these lakes renders the identification of the mechanisms driving their diversification difficult. Smaller-scale radiations of cichlids in Nicaragua may represent a more promising system in this regard. The independent colonization of several isolated Nicaraguan crater lakes by Midas cichlid fishes (*Amphilophus citrinellus* species complex, Günther (1864)) from the two great lakes Nicaragua and Managua provides an exciting situation of a natural experiment that resulted in repeated evolution of similar phenotypes (reviewed in Elmer *et al.* 2010). At least six crater lake populations exhibit a distinct genetic signature (Barluenga & Meyer 2010) and are also phenotypically differentiated from each other and the source population (Elmer *et al.* 2010). Crater lake Midas cichlid

fishes are, however, most famous for their incipient and putatively parallel intralacustrine radiations. In the crater lakes Apoyo and Xiloá (Fig. 1), at least, the flocks have speciated in sympatry (Barluenga & Meyer 2004; Barluenga *et al.* 2006; Elmer *et al.* 2009; Barluenga & Meyer 2010).

Apoyo is the oldest (max. 24 000 years, Kutterolf *et al.* 2007) and largest of the Nicaraguan crater lakes and has the highest species-richness of Midas cichlid fishes, with six endemic species described to date: *A. zaliosus* (Barlow & Munsey 1976), *A. astorquii*, *A. chancho*, *A. flaveolus* (Stauffer *et al.* 2008), *A. globosus* and *A. supercilius* (Geiger *et al.* 2010). The different Midas cichlid species inhabiting Apoyo and Xiloá all differ in meristic characters and exhibit significant differences in body shapes (Stauffer & McKaye 2002; Stauffer *et al.* 2008; Elmer *et al.* 2010; Geiger *et al.* 2010) and those studied to date have been found to be genetically significantly differentiated based on neutral genetic markers (Barluenga & Meyer 2010). These Midas cichlids form stable breeding pairs and mate assortatively by species in the field (Stauffer *et al.* 2008; Elmer *et al.* 2009). The taxonomic description of several of the new species, however, remains in flux and genetically fixed or apomorphic markers have not yet been identified for all newly described species.

Crater lake Xiloá (max. 6100 years, Kutterolf *et al.* 2007) contains to date three described endemic species: *A. sagittae*, *A. amarillo* and *A. xiloaensis* (Stauffer & McKaye 2002). Surprisingly, the two oldest and phenotypically most distinct Midas cichlid species, *A. citrinellus* and *A. labiatus* (Günther 1864) from the two large and old (>500 000 years) great lakes Nicaragua and Managua can hardly be genetically distinguished based on neutral markers, either within or among lakes (Barluenga & Meyer 2010). Supported by evidence from geology, biogeography and population genetics, these two species are presumed to be the

ancestral population from which all the crater lake populations originated (Barlow 1976b; Barluenga & Meyer 2010). Midas cichlids in captivity become sexually mature within about 9–12 months of age (AM, personal observation). Assuming that growth rates are likely lower in the wild, the maximum age of a lake in years directly relates to an upper bound for the number of generations after colonization.

Being shallow and turbid, the great lakes represent a rather homogeneous habitat, which differs markedly from the deep clear waters of the crater lakes (Barlow *et al.* 1976; Elmer *et al.* unpublished data). In the crater lakes, the water body is divided into a substrate-associated (benthic) and open-water (limnetic) zone (Vivas & McKaye 2001; Barluenga *et al.* 2006). Stemming from a more generalist great lake ancestor of benthic morphology, as no limnetic eco-types of species exist in the source lakes, the limnetic niche must have been filled within the last few thousand years in Apoyo by *A. zaliosus* and in Xiloá by *A. sagittae*. These two limnetic species differ in several ecologically important attributes from the other benthic species in their respective endemic radiations (Barlow & Munsey 1976; Meyer 1990a,b; Vivas & McKaye 2001; Barluenga *et al.* 2006) and demonstrate a parallel pattern of body shape differences (Elmer *et al.* 2010). The ecological differences between intralacustrine benthic species are less well studied thus far. However, a parallel difference in depth distribution, with *A. astorquii* and *A. amarillo* using the shallow and *A. chancho* and *A. xiloaensis* using the deep areas of respective crater lakes has been proposed, (Barluenga & Meyer 2010; Dittmann *et al.* 2012) though not tested.

Although those species groups across crater lakes are diverging along parallel axes, interspecific phenotypic (Elmer *et al.* 2010) and genetic (Barluenga & Meyer 2010) differentiation is more pronounced in the older crater lake Apoyo than in the younger crater lake Xiloá. While every speciation event is unique, investigating two replicated radiations at different ages, and therefore probably different stages of diversification, provides a rare circumstance that may inform on the increase of divergence during the speciation process (Nosil 2012).

Although the monophyly of species flocks within lakes has been well supported (Barluenga *et al.* 2006; Elmer *et al.* 2009; Barluenga & Meyer 2010), studies making inferences about the evolutionary relationships of species within crater lakes have so far been few and were solely based on microsatellite markers (Elmer *et al.* 2009; Barluenga & Meyer 2010) and mtDNA (Barluenga & Meyer 2010). Furthermore, these studies provided discordant species relationships within the Xiloá flock (Apoyo was not assessed in the study by Elmer *et al.* (2009)). Thus, the branching order of different sympatric



**Fig. 1** Map showing the study area in Nicaragua. The two great lakes Managua and Nicaragua harbouring the source populations and the two crater lakes Apoyo and Xiloá are indicated.

species in these radiations could, so far, not be determined conclusively, although such knowledge would allow a much deeper understanding of the evolutionary history of the radiations and, hence, permit to uncover possibly common mechanisms of the formation of these adaptive radiations.

Even though considerable progress is currently being made in understanding the genetics of adaptation and speciation, many more studies are needed from diverse taxonomic groups, with different life histories and from a range of ecological settings, to develop a more complete understanding of these fundamental evolutionary processes (Stinchcombe & Hoekstra 2008; Nadeau & Jiggins 2010; Nosil 2012). This kind of knowledge is essential also for the understanding of the limits of phenotypic evolution (Wake *et al.* 2011), how fast adaptation to new environments can progress (Barrett & Schluter 2008; Elmer & Meyer 2011), why speciation rates are so variable among different taxa (Gavrilets & Losos 2009; Losos 2010, 2011; Butlin *et al.* 2012) and to understand how evolutionary diversification progresses more broadly (McNiven *et al.* 2011).

We performed AFLP-based genome scans in an effort to detect signatures of divergent natural selection in populations of Midas cichlid fishes at two hierarchical levels: among sympatric species flocks and between the crater lake flocks and their ancestral populations. At both levels, we compare the number and occurrence of outlier loci. The goals of this work were: (i) to detect genomic regions that have been involved in the sympatric divergence of both crater lake radiations and the two ancestral species, (ii) to investigate whether there are homologous parallel outlier genomic regions across flocks, (iii) to detect loci that have been involved in the evolutionary divergence of both crater lake populations and their source populations, and (iv) to investigate whether there are shared outlier loci, which provide indirect support for a role of selection rather than founder effects, (v) to examine whether there is a quantitative difference in outlier loci among differently aged radiations. Lastly, this comprehensive set of molecular markers was collected (vi) to make inferences about the evolutionary relationships of sympatric species in this complex. This phylogeny is crucial for the interpretation of all data as it also informs on the ecological and evolutionary processes that lead to the formation of repeated adaptive radiations (Harmon *et al.* 2003).

## Materials and methods

### Sampling

Fish were collected using gill nets in November – December 2005 and 2007 from great lakes Nicaragua

and Managua and crater lakes Apoyo and Xiloá. Specimens were photographed in a standardized procedure and fin clips or samples of muscle tissue were taken and stored in pure ethanol. All together, 285 samples were used for AFLP genotyping. Individuals were assigned to species based on body shape, and all have been included in previous morphological analyses (Elmer *et al.* 2010). After data generation and data refinement procedures (see below), 245 samples were retained for further analyses. Species-specific samples sizes were: 29 specimens of *A. citrinellus* and 25 specimens of *A. labiatus* from great lake Nicaragua as well as eight specimens of *A. labiatus* from great lake Managua. As populations of *A. labiatus* from the two great lakes can be hardly distinguished genetically (Barluenga & Meyer 2010), they were pooled for all subsequent analyses. Specimens from crater lake Apoyo included 35 individuals of *A. zaliosus*, 28 of *A. astorquii* and 32 of *A. chanco* (samples of *A. chanco* were augmented with eight individuals from a stock at the University of Konstanz). Samples from crater lake Xiloá included 33 specimens of *A. sagittae*, 23 of *A. amarillo* and 32 of *A. xiloaensis*. Three Apoyo species were not included: *A. flaveolus* because of insufficient number of high quality DNA samples, and *A. globosus* and *A. supercilius* because they were described after (Geiger *et al.* 2010) the data for this project were already collected. Moreover, no genetic analyses have so far been conducted to ascertain the status of these latter two species.

### AFLP procedure

DNA was extracted from fin clips or muscle tissue using a standard high-salt or a phenol-chloroform protocol. DNA concentration and quality were determined on 1% agarose gels and by measuring absorbance in a spectrophotometer.

Technical aspects of AFLPs (Vos *et al.* 1995; Mueller & Wolfenbarger 1999; Bensch & Akesson 2005; Meudt & Clarke 2007) and studies conforming their usefulness in identifying regions that are associated with phenotypic traits (Rogers & Bernatchez 2005; Via & West 2008) have been reported elsewhere. We applied a derived version of the original AFLP procedure (Vos *et al.* 1995) using modified protocols from commercial kits (AFLP Core Reagent Kit, Invitrogen). For each sample, 1  $\mu$ L of DNA (concentration 100 ng/ $\mu$ L) was digested with 0.5 U each of *EcoRI* and *MseI* restriction enzymes in a final volume of 5  $\mu$ L of 1 $\times$  reaction buffer for 2 h at 37°C, followed by heat-induced enzyme inactivation for 15 min at 70°C. Adaptors were ligated overnight at 16°C by adding 4.8  $\mu$ L of adaptor ligation solution and 0.2 U of T4 DNA ligase. Subsequently, samples were diluted 1:10

with 1× TE buffer (10 mM Tris-HCl, 0.1 mM EDTA). Pre-selective amplifications were carried out using 1 µL of diluted adaptor-bound DNA in final volumes of 10 µL 1× PCR buffer (Gennaxon) containing 1.5 mM MgCl<sub>2</sub>, 250 µM of each dNTP, 1.2 µM *Mse*I and 0.2 µM *Eco*RI pre-selective primers complementary to the restriction sites and adaptors containing one additional base, and 0.2 U of Red-*Taq* polymerase (Gennaxon). Cycling conditions for pre-selective PCRs were: denaturation for 2 min at 94°C, 20 cycles of 94°C for 30 s, 56°C for 30 s, 72°C for 90 s, followed by a 15 min final elongation at 72°C. For selective PCRs, 2.5 µL of pre-amplified DNA (diluted 1:50 with 1× TE buffer) was amplified in 10 µL volumes of 1× PCR buffer (Gennaxon) containing 1.5 mM MgCl<sub>2</sub>, 250 µM of each dNTP, 600 nM *Mse*I and 100 nM *Eco*RI selective primers and 0.2 U of Red-*Taq* polymerase (Gennaxon). Selective primers contained two additional nucleotides. *Eco*RI primers were labelled with either a JOE or a FAM fluorescent dye (Applied Biosystems). For selective amplifications, we used touchdown PCR conditions: denaturation for 2 min at 94°C, 13 cycles of 94°C for 45 s, 65°C for 45 s (decreasing 0.7°C per cycle), 72°C for 2 min, followed by 25 cycles of 94°C for 45 s, 56°C for 45 s, 72°C for 2 min and ending with elongation for 20 min at 72°C.

#### Genotyping and data refinement

After testing several primer combinations for quality and reliability, all samples were genotyped for 13 different selective primer combinations (Table S1, Supporting information). Fragments were analysed in an ABI 3130xl automatic capillary sequencer (Applied Biosystems) and sized according to an internal GeneScan 500 ROX size standard. To assess the reliability of our procedure, 95 randomly chosen specimens covering all populations were amplified and genotyped twice. To avoid observer-induced biases, fragments were scored automatically using the GeneMapper 4.0 software (Applied Biosystems). Bin sets were adjusted manually and individual electropherograms were corrected for erroneous peaks.

AFLP quality was rigorously assessed. Fragments outside the range of 55–500 bp were excluded. Samples exhibiting <75% of the median fingerprint intensity (sum of all peak heights) for any single selective primer combination were removed and coded as missing data. The overall proportion of missing data was 2.5%. Individual loci that were <90% reliable were omitted from further analysis. Subsequently, we applied data refinement pipeline implemented in AFLPScore (Whitlock *et al.* 2008) to create individual genotype data in the form of a binary matrix from peak intensities. The

pipeline included data normalization step and ensured an objective way of scoring the AFLP data. Locus- and phenotype-specific thresholds (see Whitlock *et al.* 2008 for explanation) were adjusted for each primer combination, resulting in a tradeoff between reducing genotyping error while retaining a reasonable number of loci. Final overall genotyping error was 2.9% (±0.5 SD), a typical level for AFLP studies (Bonin *et al.* 2004).

#### Genetic differentiation and phylogenetic trees

Genetic differentiation between species by means of the  $F_{ST}$  fixation index (Wright 1965) and Nei's gene diversity within species (Nei 1973) were estimated using AFLP-SURV version 1.0 (Vekemans *et al.* 2002). Allele frequencies were computed using an implemented Bayesian approach (Zhitovovsky 1999) assuming HWE and non-uniform prior distributions among populations. Statistical support for genetic differentiation was determined based on 5000 permutations. Genetic differentiation was also estimated after outlier detection using the presumably neutral dataset (951 AFLPs) as well as the outlier loci alone (79 AFLPs).

Phylogenetic trees at the population level were reconstructed with FAMD vers.1.2.5 (Schlüter & Harris 2006) with the following settings: population genetic distances were calculated with the Takezaki formula adjusted from the Cavalli-Sforza & Edwards method (Cavalli-Sforza & Edwards 1967; Takezaki & Nei 1996) and employed to build Neighbour-Joining (NJ) trees (Saitou & Nei 1987). Allele frequencies were estimated based on a Bayesian approach (Zhitovovsky 1999) assuming HWE and non-uniform prior distributions among populations. Support for the tree topology was estimated by means of 1000 bootstrap replications. Missing data were randomly replaced by 50% presences and 50% absences. Majority rule consensus trees were then calculated in MEGA version 4 (Tamura *et al.* 2007). Trees were built using the complete dataset (1030 AFLPs) and, after outlier detection, using the presumably neutral markers (951 AFLPs) and the outlier loci (79 AFLPs) separately.

#### Outlier loci detection

Detection of outlier loci by means of  $F_{ST}$  genome scans was conducted in BayeScan version 2.01 (Foll & Gaggiotti 2008; Fischer *et al.* 2011) using the previously obtained dominant binary AFLP matrix as input data. BayeScan implements a Bayesian approach to estimate the posterior probability that a locus is under selection. Briefly, the program defines two models, one including only population-specific effects and one taking additionally locus-specific effects into account. It then returns a posterior odds (PO) ratio of these

alternate models for each locus. A positive value of the posterior distribution of the locus-specific parameter  $\alpha$  suggests that a locus is under divergent selection whereas a negative value indicates balancing selection. A recent comparative study (Perez-Figueroa *et al.* 2010) found BayeScan to be superior to other commonly used approaches for detection of outlier loci using dominant markers and it has become the tool of choice for this purpose (Manel *et al.* 2009; Freedman *et al.* 2010; Paris *et al.* 2010; Parisod & Joost 2010; Paun *et al.* 2010; Colbeck *et al.* 2011; Hoffman *et al.* 2011).

We performed a global analysis, lake-specific analyses and analyses between both the ancestral species and each crater lake's species flock. Following a similar approach as Deagle *et al.* (2011), we used different prior odds values for the neutral model to determine outlier loci. As analyses based on a greater number of populations have higher power (Fischer *et al.* 2011), we used a prior odds value of 10 for the global analysis and a more relaxed value of 1 for the comparisons with fewer populations. However, outliers detected in these comparisons were not taken into account for further analyses unless they also appeared as outliers in the global scan. We think of this procedure as a reasonable tradeoff because it is not only conservative but also allows us to evaluate the detected outlier loci in a more meaningful way.

Monomorphic loci were always excluded from analyses, so the number of loci screened varies across comparisons. Obviously, including more populations in an analysis increases the chance of obtaining alternate alleles at a locus, i.e. a higher number of polymorphic markers. Model parameters were automatically tuned from the default settings with 10 pilot runs, each consisting of 10 000 MCMC steps. For the actual analyses 500 000 iterations with an additional burn-in phase of 50 000 steps were applied. Outlier loci were defined by keeping the expected false discovery rate (FDR) constant ( $<0.05$ ), resulting in varying levels of the PO threshold (Deagle *et al.* 2011; Fischer *et al.* 2011). Outlier loci were further examined by means of species-specific allele frequencies as estimated from the dominant AFLP matrix by BayeScan.

Statistical difference of the number of outlier loci in different contexts was evaluated by means of Fisher's exact tests (R Development Core Team 2010); outlier loci counted as successes, and failures were defined as the total number of screened (polymorphic) loci minus the outliers. To compare the number of outlier loci between the source lakes and both the crater lakes populations, loci that also occurred as outliers in intralacustrine analyses were not taken into account.

## Results

### *AFLP generation and genotyping*

Specimens from all eight Midas cichlid species (mean  $n = 30.6 \pm 3.8$  SD) were successfully genotyped for 13 different primer combinations (Table S1, Supporting information). After refinement procedures, the final dataset consisted of 245 samples genotyped at 1179 loci. Of these, 1030 loci were polymorphic across all eight species. The number of markers was more than sufficient to obtain a unique genetic fingerprint for every individual.

### *Genetic differentiation and gene diversity*

As the evaluation of the evolutionary relationships among species is fundamental to the interpretation of the genome scans results, we will present these results first. Global genetic differentiation between species based on all 1030 AFLP markers is highly significant ( $F_{ST} = 0.154$ ,  $P < 0.0001$ ) and all pairwise comparisons based on the full dataset showed significant genetic differentiations (Table 1). Species occurring in sympatry were less genetically differentiated than any two allopatric species, the only exception being comparisons within Apoyo of *A. zalius* and *A. chancho* with *A. astorquii* (Table 1). All comparisons with *A. astorquii* showed generally high  $F_{ST}$ -values, whether in sympatry or allopatry. Species from Xiloá were generally less genetically differentiated from the ancestral species than was the Apoyo flock. The smallest degree of inter-specific genetic differentiation was found between the two ancestral species from the great lakes (Table 1).

Within Apoyo, the least genetic differentiation is found between the limnetic species *A. zalius* and the benthic species *A. chancho*. The other benthic species *A. astorquii* is more differentiated from *A. zalius* than it is from *A. chancho*.

Pairwise comparisons within Xiloá indicate that the limnetic species *A. sagittae* is more differentiated from the two benthic species, and more differentiated from *A. amarillo* than from *A. xiloensis*, than the benthic species are to each other.

Pairwise comparisons based on the presumably neutral and outlier datasets show an almost identical pattern, although the absolute values differ as expected. Global genetic differentiation ( $F_{ST} = 0.108$ ,  $P < 0.0001$ ) and all values are decreased based on the presumably neutral dataset alone (Table S2, Supporting information), whereas the global genetic differentiation ( $F_{ST} = 0.333$ ,  $P < 0.0001$ ) and all pairwise  $F_{ST}$ -values are increased when using the outlier data (Table S3, Supporting information). Estimates for Nei's gene

**Table 1** List of localities, studied species, samples sizes ( $N$ ), Nei's gene diversity ( $H_j$ ) and pairwise  $F_{ST}$ -values based on the complete dataset (1030 AFLP loci). Brackets after species names indicate limnetic (l) and benthic (b) ecotypes. Species names are abbreviated with their first three letters at the top. Statistical significance was assessed based on 5000 bootstrap replicates

Locality	Species	$N$	$H_j$	Pairwise $F_{ST}$							
				cit	lab	zal	ast	cha	sag	ama	
Great lakes	<i>Amphilophus citrinellus</i>	29	0.119								
	<i>Amphilophus labiatus</i>	33	0.134	0.021*							
Apoyo	<i>Amphilophus zaliosus</i> (l)	35	0.137	0.134**	0.122**						
	<i>Amphilophus astorquii</i> (b)	28	0.223	0.229**	0.206**	0.164**					
Xiloá	<i>Amphilophus chancho</i> (b)	32	0.109	0.150**	0.147**	0.066**	0.145**				
	<i>Amphilophus sagittae</i> (l)	33	0.123	0.089**	0.075**	0.135**	0.227**	0.151**			
	<i>Amphilophus amarillo</i> (b)	23	0.093	0.067**	0.079**	0.157**	0.254**	0.175**	0.066**		
	<i>Amphilophus xiloaensis</i> (b)	32	0.097	0.078**	0.089**	0.152**	0.259**	0.178**	0.050**	0.035**	

Probability values: \* $P < 0.05$ , \*\* $P < 0.001$ .

diversity (analogous to expected heterozygosity) within populations based on the complete dataset (1030 loci) range from 0.093 to 0.223 (Table 1), with a mean of 0.129 ( $\pm 0.041$  SD).

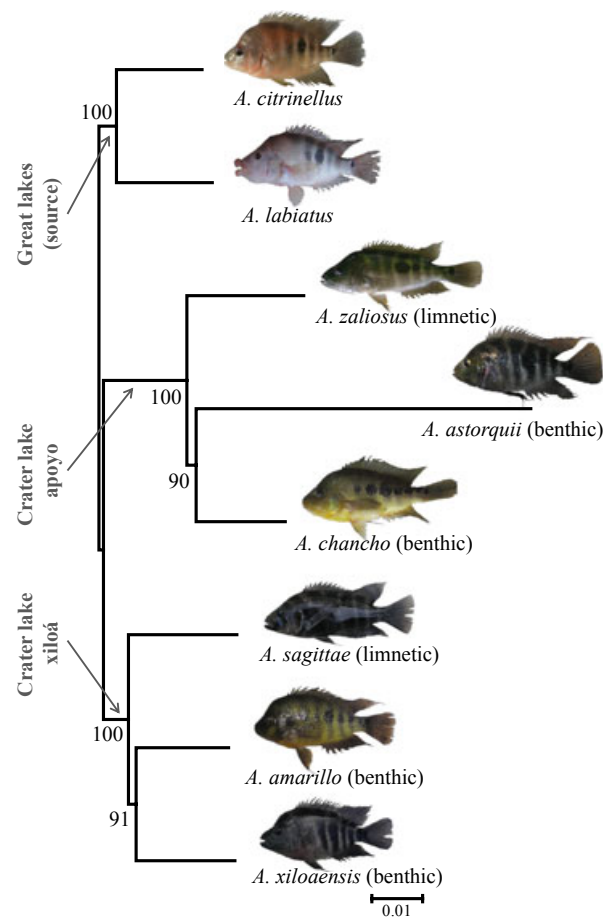
#### Phylogenetic reconstruction

Using the complete dataset (1030 AFLP markers), the population-level NJ trees correctly assign species to their respective lakes (100% of the bootstrap replicates) (Fig. 2). Within the clade of Apoyo species, the benthic species *A. astorquii* and *A. chancho* form a cluster with high support (90% bootstrap replicates). Within the clade of Xiloá species, the two benthic species *A. amarillo* and *A. xiloaensis* also resolve together (91% bootstrap replicates).

Based only on the outlier loci (79 AFLPs), we find a similar pattern as for the complete dataset (Fig. S1a, Supporting information). The tree topology is identical with high support (96% bootstrap replicates) for sorting species by lake of origin. The sister relationship of the benthic species within Apoyo is again highly supported (100% bootstrap replicates) and the benthic species within Xiloá resolve together with moderate support (66% bootstrap replicates) (Fig. S1a, Supporting information).

The NJ trees based on the presumably neutral dataset (951 AFLPs) again successfully assigns species to lake of origin (100% bootstrap replicates), though based on a slightly different topology (Fig. S1b, Supporting information). Whereas the relationships in the Xiloá clade remain the same as was identified with the complete dataset (benthic species resolve as sister taxa with 88% bootstrap replicates), within the Apoyo flock there is moderate support for a sister relationship between the limnetic *A. zaliosus* and the benthic *A. astorquii* (76% of bootstrap replicates) (Fig. S1b, Supporting information).

As expected, absolute branch lengths vary considerably between the trees based on different components



**Fig. 2** Population-level neighbour-joining trees based on the complete dataset (1030 loci). Nodes show per cent bootstrap support (1000 replicates). Pictures show representative specimens. Ecotypes are given in brackets. Arrows mark lakes of origin for respective species flocks.

of the dataset. The tree based on outlier loci had the longest branches and the 'neutral' tree indicated the least evolutionary change. In all datasets, and in accor-

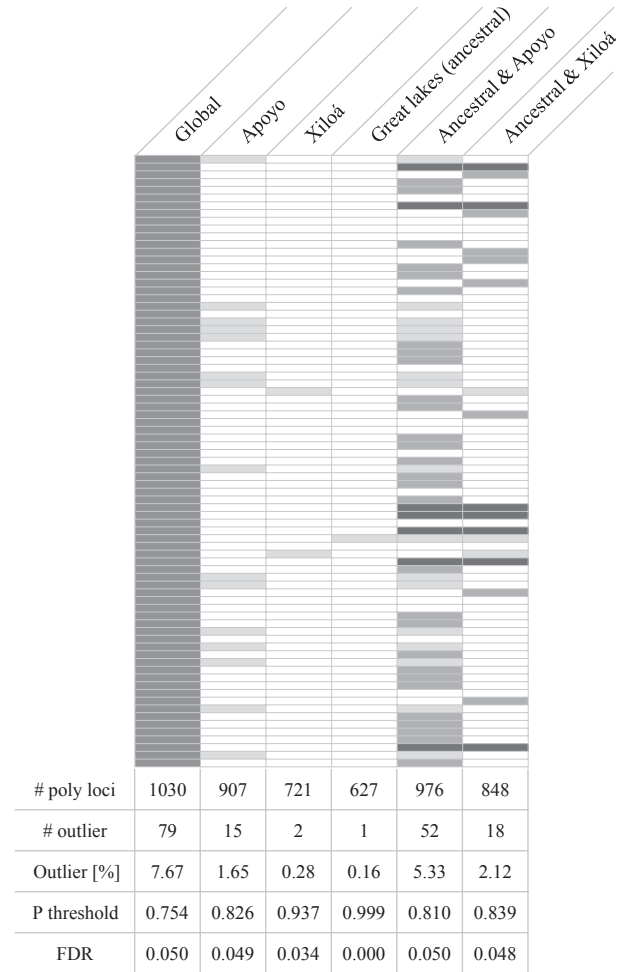
dance with its greater age, species from the Apoyo clade are more genetically differentiated from the source populations than are species from the Xiloá clade. The extraordinary long branch length of the lineage of *A. astorquii* is obvious in all trees.

### Genome scans

We performed outlier analyses including all eight species (global analysis) and at two different hierarchical levels: between sympatric species of lakes (each three species for the crater lakes and two species for the great lakes) and between the great lake species and each of the crater lake flocks (five species each comparison). In the global analysis, 79 of 1030 polymorphic loci (7.7%) are found to be outliers (Fig. 3). All outlier loci exhibit a positive value for the locus-specific parameter  $\alpha$ , indicating positive selection. Not a single case of supposedly balancing selection (outlier with a negative value of  $\alpha$ ) was identified. The fact that no such signs were detected is most likely because of methodological limitations rather than biological reality, as genome scans are known to have limited power in identifying signs of balancing selection (Beaumont & Balding 2004). Outlier loci are distributed among all 13 primer combinations (Table S1, Supporting information).

In the lake-specific analysis of Apoyo we find 15 of 907 polymorphic loci (1.7%) to be outliers. With two outliers out of 721 polymorphic loci (0.3%), significantly fewer outliers were determined in the analysis of species from Xiloá (Fisher's exact test,  $P$ -value = 0.006). The two loci determined as outliers in the Xiloá flock are not among the 15 outlier loci identified in the Apoyo flock. When crosschecked, loci that have been found to be outliers in one radiation show no signs for an elevated differentiation in the other radiation (data not shown). Between *A. citrinellus* and *A. labiatus* from the great lakes, only one of 627 polymorphic loci (0.2%) was found to exhibit outlier status, which in turn was not among the outliers identified in the intralacustrine analyses of the crater lake flocks.

For all 15 outliers in Apoyo, the species-specific allele frequencies are drastically different between *A. zalius* and *A. astorquii* with a high range for the '1' allele from 0.776 to 0.982 as compared with a low range from 0.019 to 0.155, respectively (Table S4, Supporting information). Allele frequencies for *A. chancho* are in five cases close to *A. zalius* (0.824–0.984) and in the other 10 instances intermediate (0.328–0.675). In the species from the great lakes and Xiloá, allele frequencies at these loci are generally high and close to *A. zalius*, except for one case where allele frequencies are close to *A. chancho* and intermediate between *A. zalius* and *A. astorquii*. In Xiloá, allele



**Fig. 3** Appearance of outlier loci. Rows represent loci detected as outliers in a global scan with more stringent criteria (prior odds of 10). Columns show lake-specific analyses and such comparing crater lake flocks with their ancestral species. These analyses were carried out with less-stringent criteria (prior odds of 1). Filled cells indicate outlier status in respective analyses. Differential shading designates outlier role, with light grey indicating lake-specific outliers, intermediate grey indicating outlier potentially related to adaptation to crater lake environment and dark grey indicating the latter which occurred in parallel among both crater lakes. Given at the bottom are the number of screened polymorphic loci, the number and proportion of detected outliers, the applied posterior probability ( $P$ ) threshold to define outliers and the false discovery rate for respective analyses. See also text for details.

frequencies at the two outlier loci are varying among all three species and no pattern is apparent (Table S4, Supporting information).

The comparison between the ancestral species and the Apoyo flock shows a higher proportion of outlier loci (52 out of 976, 5.3%) than the equivalent comparison of the ancestral species with the Xiloá flock (18 out of 848, 2.1%). All loci that were identified as



outliers in one of the three intralacustrine analyses were also identified as outlier in the analyses between lakes. Excluding those intralacustrine outlier loci, significantly more loci were found to be outliers in the comparison between the ancestral species vs. the Apoyo flock (36 out of 976, 3.7%) than for the ancestral species vs. the Xiloá flock (15 out of 848, 1.8%) (Fisher's exact test,  $P$ -value = 0.015). Out of these 51 great lake to crater lake outlier loci, seven (13.7%) are shared among both comparisons (Fig. 3). Allele frequencies at these outlier loci showed no consistent pattern between different populations within lakes.

## Discussion

We applied an AFLP-based population genomics approach to investigate genome-wide signatures of divergence in two young and repeated radiations of Midas cichlid fishes inhabiting Nicaraguan crater lakes and their source populations from the two great lakes of Nicaragua. This represents one of the first efforts to characterize the genetic bases underlying the rapid and often parallel phenotypic diversifications (Elmer *et al.* 2010) in this natural system. Specifically, we (i) detected loci associated with the divergence of sympatric species and (ii) investigated whether the same regions have been involved in this process within the two repeated radiations. At the inter-lacustrine level, i.e. by comparing crater lake flocks with the ancestral species, we (iii) found outlier loci potentially associated with allopatric ecological divergence and (iv) examined whether there are shared outlier loci indicating adaptation to the crater lake environment in general. By comparing the two differently progressed radiations, our (v) results provide evidence that the number of outlier is increasing with time. Finally, we (vi) reconstructed the phylogenetic relationships among sympatric species. As this last analysis serves as baseline information for the interpretation of the genome scan results, we will begin by discussing this result first.

### *Parallel sequence of sympatric speciation events*

Stemming from source populations inhabiting the great lakes in Nicaragua, our phylogeny based on the complete dataset (1030 loci) demonstrates that Midas cichlid fishes colonized crater lakes Apoyo and Xiloá independently and speciated subsequently *in situ*. The ancestral individuals from those lakes will likely have been benthic in their morphology as there are no limnetic species or morphs in the shallow and murky waters of the great lakes. Interestingly, the tree topology suggests that the sequence of diversification followed a parallel pattern, with a primary divergence among

the limnetic-benthic axis in both crater lake radiations (Fig. 2). Hence, the most parsimonious scenario is the following; the first of the extant species to diverge from, and coexist with, the benthic founder population in each of the two crater lakes was the limnetic species. In a second set of speciation events, the benthic species subsequently split into two benthic species and therefore, in both lakes the two new endemic benthic species tend to be younger and more closely related to each other. Levels of genetic differentiation including all markers (1030 loci) reflect the same pattern, with comparisons of sympatric species showing generally less genetic differentiation than pairwise comparisons of allopatric species, and the two benthic species being less differentiated from each other than to the limnetic species. Yet, this pattern is blurred by comparisons with *A. astorquii*, which showed generally high  $F_{ST}$ -values in sympatry and in allopatry. Nonetheless, *A. astorquii* is, as expected, more differentiated from any allopatric population than to its two sympatric species and in turn more differentiated to the sympatric limnetic species *A. zalius* than to the other sympatric benthic species *A. chancho* (Table 1). The patterns of the formation of the parallel adaptive radiations in each of the crater lakes is therefore repeated across lakes, suggesting that the processes were the same in both crater lakes following their initial colonization.

These AFLP results agree with our previous results on the Midas cichlid species complex that demonstrated the monophyly of crater lake flocks (Barluenga *et al.* 2006; Barluenga & Meyer 2010). In addition, the lower genetic differentiation of Xiloá species might be expected given the lake's younger age [max. 6100 years (Kutterolf *et al.* 2007)] as compared to Apoyo [max. 24 000 years (Kutterolf *et al.* 2007)] and therefore more recent colonization by Midas cichlids (Barluenga *et al.* 2006; Elmer *et al.* 2012). Moreover, a sister relationship between the two benthic species in both crater lakes had already been suggested by means of genetic differentiation based on microsatellite markers and mtDNA in a recent study (Barluenga & Meyer 2010). Yet, another study based on different samples and a slightly reduced set of the same microsatellite markers found *A. amarillo* to be the most differentiated species within Xiloá (Elmer *et al.* 2009). Based on different genetic markers and using different metrics of population distance as well as assessing statistical support, our results support a scenario of a parallel primary diversification of limnetic and benthic species in both crater lakes. Therefore, the radiations of Midas cichlids agree with numerous other examples of freshwater fish taxa for which sympatric divergence into a new open-water niche has happened (Schluter & McPhail 1993; Robinson & Wilson 1994) and within a relatively short

amount of time (Schluter 1996; McKinnon & Rundle 2002; Rogers & Bernatchez 2007). Furthermore, inferences from mathematical models (Gavrilets & Losos 2009) as well as empirical data from vertebrates (Streelman & Danley 2003) suggest that the primary divergence in adaptive radiations often occurs in this manner among macrohabitats.

We note, however, that the use of genetic distances and levels of genetic differentiation between populations as a proxy for time since divergence is not without potential problems in scenarios of sympatric speciation. This is because adaptive divergence may decrease gene flow (i.e. isolation by adaptation) and the higher levels of genetic differentiation between the limnetic and benthic species as compared to the two benthic species might reflect a stronger (historical) reduction in gene flow rather than time since divergence (Nosil *et al.* 2008; Nosil 2012). Trying to disentangle these not always mutually exclusive hypotheses, for example by coalescent-based approaches, is an ongoing effort, but estimating the timing of gene flow remains difficult (Niemiller *et al.* 2010; Strasburg & Rieseberg 2010, 2011).

We find a discrepancy between tree topologies based on the complete and outlier datasets vs. the presumably neutral dataset. Whereas using 79 outlier loci (Fig. S1a, Supporting information) provides a congruent tree topology as the complete dataset, applying the 951 presumably neutral loci yields a slightly different topology, with the benthic species *A. chancho* diverging first in Apoyo (Fig. S1b, Supporting information). This inconsistency might be attributable to the fact that during the early stages of adaptive divergence, loci under divergent selection may actually constitute the branch defining loci. Neutrally evolving loci may not informatively reflect the species tree, as they may have been homogenized by recent or ongoing gene flow or reflect stochastic events of lineage sorting (Via 2001, 2009). Especially during the process of ecological speciation and in the face of gene flow, this phylogenetic discordance of different genomic regions may persist through the speciation process and possibly well beyond (Via 2009). Taking into account the young age and sympatric origin of crater lake species in this system, we therefore think that the use of all available markers (neutral and outlier loci) is the most appropriate reflection of species' histories.

Conspicuous in all trees is the extraordinary long branch of *A. astorquii* in Apoyo. This is also reflected in the high gene diversity and levels of genetic differentiation of *A. astorquii* and hints at an elevated evolutionary rate. Yet, if this were true we would expect to see the same pattern in other neutral markers such as microsatellites; but this is not the case (Barluenga & Meyer 2010). Furthermore, because of the recent demographic

bottleneck of the crater lake populations we would expect reduced levels of gene diversity within the crater lake flocks as compared to within the great lake populations. Interestingly, whereas previous analyses based on microsatellite markers found such a pattern (Barluenga & Meyer 2010), estimates of gene diversity based on AFLPs in this study are generally in a similar range for all populations, with the exception of the high gene diversity of *A. astorquii*. Although the absolute values of gene diversity based on microsatellite and AFLP markers are expected to differ based on differences in the number of alleles per locus (multi-allelic vs. bi-allelic) and marker type (co-dominant vs. dominant), we are unaware of an explanation for why our results show a pattern of relative gene diversity between populations that differs from estimates based on microsatellites. Nevertheless, because of our automated and objective data generation procedure and the biologically meaningful phylogenetic results, we think that this difference and the distinctiveness of *A. astorquii* are biologically real rather than because of a methodological bias. It should be noted that our general conclusions would still hold even if the elevated rate of evolution in *A. astorquii* were an artifact.

#### *Signatures of divergent selection between sympatric species*

We identified genomic regions under divergent selection within both radiations, which fits with expectations given that the crater lake species originated in sympatry and occupy different ecological niches (Vivas & McKaye 2001; Barluenga *et al.* 2006). We found 15 (1.7%) and two (0.3%) outlier loci in the intralacustrine analyses of each three species inhabiting crater lakes Apoyo and Xiloá, respectively. We interpret the detected outlier markers as being located in regions that have been, or still are, resisting gene flow. The species-specific allele frequencies for the 15 outlier loci in Apoyo revealed an intriguing pattern: *A. astorquii* consistently has lower allele frequencies for the '1' allele than all other species at these loci (Table S4, Supporting information). This suggests that the signatures of divergent selection in Apoyo are strongly driven by the divergence of *A. astorquii*. As a result of the anonymity of the AFLP markers, we are unable to infer whether or not the identified outlier loci reside on different genomic regions or if they are physically linked and occur only on one or a few regions under selection. The pattern of consistently deviant allele frequencies might, for example, result from several of the outliers being captured in a chromosomal inversion that occurs in high frequency in *A. astorquii*. In this case, genetic differentiation of loci within the inver-

sion, because of a reduction in recombination, would be expected to be much higher than in co-linear genomic regions (e.g. Cheng *et al.* 2012) and the spread of one adaptive allele located on the inversion could bring about a signature of selection on multiple loci simultaneously. Alternatively, several adaptive alleles could reside on the inversion, thereby being protected from recombination and facilitating the speciation process (Kirkpatrick & Barton 2006). A role for chromosomal inversions involved in the divergence-with-gene-flow of Midas cichlid would not be surprising, given the theoretical and empirical support for inversions in this process in general (reviewed in Hoffmann & Rieseberg 2008). Alternatively, several outliers could be located on distinct regions that nonetheless harbour genes involved in a co-adapted gene complex. Testing these architectural alternatives with markers of known position is important for understanding the speciation process and currently in progress in this laboratory.

In a biologically similar scenario, on average 3.2% of screened AFLP loci have been found to be outliers in the replicated evolution of dwarf and normal ecotypes of postglacial lake whitefish (*Coregonus clupeaformis*) (Campbell & Bernatchez 2004). This might suggest that fewer loci are responsible for species differences in Midas cichlids compared to lake whitefish. However, the number of outlier loci is not easily comparable across studies as lower sample sizes and a lower relative marker density decrease the chances of detecting outliers (Foll & Gaggiotti 2008) and different outlier detection methods vary in their power (Perez-Figueroa *et al.* 2010). In any case, the low number of outlier loci detected in our study suggests that our approach to determine outlier loci is rather conservative. For example, the only two outlier loci detected in Xiloá are unlikely to represent the only genomic regions diverged under selection among the three species.

The pairwise comparisons between the two ancestral species, *A. citrinellus* and *A. labiatus* from the great lakes, revealed the least genetic differentiation (Table 1) and it is not significant after removal of the 79 outlier loci (Table S2, Supporting information). Further, only a single outlier locus (0.2%) was detected between *A. citrinellus* and *A. labiatus*. The low level of genetic differentiation found in this study is consistent with previous results (Barluenga & Meyer 2010) and suggests that there is substantial ongoing gene flow. Hence, the species status of these two forms may be questionable. However, the two species are phenotypically quite different: in body shape (Barlow 1976b; Klingenberg *et al.* 2003; Elmer *et al.* 2010), lip size (Barlow & Munsey 1976; Klingenberg *et al.* 2003) as well as diet and gene expression profiles (Manousaki *et al.*, accepted). This may suggest

that a single genomic region harbouring several strongly linked genes (or few genes with pleiotropic effects) is responsible for their phenotypic differences, and that selection acting on this one region is not sufficient to facilitate further divergence (Nosil 2012). Alternatively, more genomic regions might be under selection, but remain undetected by our population genomics approach (Butlin 2008; Michel *et al.* 2010). Divergence of *A. citrinellus* and *A. labiatus* in comparison to the younger, but more differentiated, sympatric species in the crater lakes could make an interesting case about whether strong selection on a few loci or multifarious selection on many loci is more inductive to further divergence and speciation (Nosil & Feder 2012). Efforts to further characterize the genomic regions of divergence in this system are currently underway in our laboratory.

#### *Genetic non-parallelism between replicated radiations?*

The pattern of parallel phenotypic evolution, as exhibited by initial speciation of the two limnetic species, led us to expect that at least some of the outlier loci in our genome scans might be shared between the two intralacustrine analyses (Campbell & Bernatchez 2004; Bonin *et al.* 2006; Nosil *et al.* 2008; Hohenlohe *et al.* 2010a; Deagle *et al.* 2011). However, we found no evidence for parallel genetic divergence between both radiations so far. The most conspicuous divergent trait in Midas cichlid fishes is body shape, which is significantly different among all species (Elmer *et al.* 2010). Body shape is likely to be under polygenic control, with a few loci of major and multiple loci of minor effect, as has been found in threespine stickleback (Albert *et al.* 2007). It is hence possible that the parallel phenotypic evolution across both radiations (Elmer *et al.* 2010) can be explained (at least partly) by a lake-specific genetic basis in the gene regions linked to the two and 15 different outlier loci that we detected. Indeed, this conclusion is not unreasonable, as a role for different loci involved in parallel phenotypic evolution has been indicated in several of the few studies that exist to date where the genetic basis for (parallel) adaptation has been discovered (Colosimo *et al.* 2005; Borowsky 2008; Steiner *et al.* 2009; Chan *et al.* 2010).

Alternatively, there could be several reasons why we did not find the same outlier loci in the repeatedly evolved eco-types across adaptive radiations. Firstly, population genomic approaches tend only to detect loci with a high effect size and most likely not all the genes that contribute to a trait will be tracked (Rogers & Bernatchez 2005). Hence, it is likely there are alleles with minor contribution to adaptive divergence that have indeed been selected in parallel in both crater lakes, but

remained undetected with this approach. The fact that we only detected two outlier loci in the Xiloá flock exemplifies this. Secondly, the ability of genome scans to identify regions under selection depends strongly on the history of a selected allele, since with time the reduction of variability around a selected locus will decay because of new mutations and recombination (Nadeau & Jiggins 2010). Under a soft-sweep-scenario, where selection operates on standing genetic variation or on different mutations at the same locus, footprints of selection will be much weaker than under a hard sweep, where advantageous alleles arise *de novo* and spread in a population rapidly (Barrett & Schluter 2008). Thus, detecting the genetic bases of quantitative traits selected from standing genetic variation is especially difficult via a population genomics approach (Hohenlohe *et al.* 2010b). The young age and parallel diversification of our studied repeated radiations suggests selection on standing genetic variation is likely. Thirdly, the 1030 polymorphic markers screened in this study are estimated to yield an average coverage of approximately one marker per Mb (Megabase), assuming that Midas cichlids have a 1 Gb (Gigabase) genome size (Henning *et al.* unpublished data) similar to African cichlids (Majumdar & McAndrew 1986; Kocher *et al.* 1998; Lang *et al.* 2006). As the AFLP population genomics approach relies heavily on the effects of genetic hitchhiking, this coverage might be too low to detect signatures of selection resulting from standing genetic variation, where blocks of linkage disequilibrium (LD) are short. Moreover, the variance of allelic associations is large when selection is acting on standing genetic variation (Przeworski *et al.* 2005). Taking this into account, it is possible that the same loci of large effect have been selected in both crater lakes, but association between these loci and our AFLP markers has decayed differently in both radiations, leading to a pattern of non-overlapping outlier loci.

Hence, although the complex genetic architecture of the trait of interest and limitations of our methodology might simply preclude us from detecting an existing underlying genetic parallelism, it is possible that different outlier loci are indeed mainly responsible for the parallel phenotypic evolution in both radiations. Further studies investigating the hypothesis of genetic non-parallelism in Midas cichlids are needed and currently underway in our laboratory using other population genomics approaches that promise to yield more complete genome coverage.

#### *Signatures of divergence between allopatric lake flocks*

After excluding outlier loci involved in the intralacustrine differences, we found 36 (3.7%) and 16 (1.8%)

outlier loci in the comparisons of the ancestral species vs. the Apoyo and Xiloá flock, respectively. These loci might be linked to regions that are involved in the significant phenotypic divergence of crater lakes populations as compared to the ancestral species (Elmer *et al.* 2010). Although these loci might very well be under divergent selection and involved in adaptation to the crater lake environment, we cannot rule out that at least some of them reflect stochastic differences of allele frequencies driven by strong founder effects (Foll & Gaggiotti 2008).

#### *Genetic parallelisms involved in allopatric divergence of crater lake populations*

Seven loci appear as outliers in both comparisons of a crater lake flock with the source population; a pattern that is unlikely to result from random effects (Nosil *et al.* 2008). The substantial number of shared outlier loci furthermore suggests at least some role for standing genetic variation, since the occurrence of seven *de novo* mutations in the same regions in both crater lake populations in such a short time (*ca.*  $5400 \pm 750$  years for Xiloá, Elmer *et al.* 2012) seems unlikely. An alternative explanation that an independent *de novo* mutation occurring in the same region would affect seven outlier loci simultaneously (a region probably spanning several megabases) seems equally unlikely, as islands of divergence in an allopatric model are predicted to be small and distributed among the genome (Nosil 2012).

Although it would be more parsimonious to interpret the replicated outliers to indicate selective sweeps that occurred in the ancestral populations, rather than having occurred independently in both the crater lake flocks, we think the demographic history of Midas cichlids renders that unlikely. The great lakes are relatively old and provide a presumably fairly stable environment. Hence, there is no indication that strong selective pressures have been working in the populations of the great lakes after the split of the founder population of Xiloá. In contrast, the crater lake populations have experienced a presumably strong ecological change with their recent founding of a new environment very different from the great lakes in water clarity, depth, water chemistry and ecological community (Barlow *et al.* 1976; Elmer *et al.*, 2010; Elmer *et al.*, unpublished data). Thus, we think it is more likely that the outliers identified in the ancestral species to crater lake flock comparisons represent signatures of selection that acted on the crater lake populations. Consequently, we believe that the outliers suggest a role for ecological divergence in allopatry, and the fact that some of them are shared indicates that at least some of the same genomic regions are involved in the adaptation to the crater lake environ-

ment. This does not seem unlikely, given that they may have started with a similar repertoire of standing genetic variation because of their recent common ancestry (Barlow 1976a; Barluenga & Meyer 2010). Nonetheless, we are currently generating data to test the origin of loci under selection using linkage information (e.g. the integrated haplotype score (iHS), Voight *et al.* 2006) from high-resolution genomic data with positional information.

#### *Number of outlier loci increases with time*

Levels of genetic differentiation between the ancestral species and Apoyo species ( $F_{ST}$ -values: 0.134–0.229) are generally higher than the equivalent comparisons with species from Xiloá ( $F_{ST}$ -values: 0.067–0.089) (Table 1). The same pattern is also reflected in the branch lengths of our phylogeny (Fig. 2). This is in keeping with expectations from geological ages of these lakes: 24 000 years ago for Apoyo and 6100 years ago for Xiloá (Kutterolf *et al.* 2007), and in congruence with previous demographic analyses (Barluenga & Meyer 2010). In addition, phenotypic differences are higher among species of Apoyo than among species in Xiloá (Elmer *et al.* 2010). These results indicate that the radiations in the two crater lakes represent two different stages of divergence, but might be along a similar trajectory of phenotypic diversification.

Our AFLP genome scans reveal that there are significantly more outlier loci between sympatric species in Apoyo (15 out of 907 polymorphic loci, 1.7%) than in Xiloá (two out of 721 polymorphic loci, 0.3%) (Fisher's exact test,  $P$ -value = 0.006), and that the number of outliers is also higher in comparison between the ancestral species and the Apoyo flock (36 out of 976 polymorphic loci, 3.7%) than in the equivalent comparison with species from Xiloá (15 out of 848 polymorphic loci, 1.8%) (Fisher's exact test,  $P$ -value = 0.015). This result implies that more loci are affected by selection over time, whereas species continue to diverge and adapt to their environments.

Yet, as discussed above, a linkage of outlier markers in Apoyo might bias the comparison of the number of outliers between sympatric species within the two crater lake flocks towards an overestimation in Apoyo. However, note that this would not hold true for the comparison in allopatry. Moreover, our results could be influenced by the demographic history of crater lake populations, which might increase the chance of obtaining false positives in the allopatric analyses (Foll & Gaggiotti 2008); albeit all results are presented at a fixed FDR of 5%. However, such an influence would most likely be present in both respective analyses. Also, the higher levels of neutral

genetic differentiation between Apoyo and the ancestral species, as compared to Xiloá, are expected to rather decrease the power of detecting directional selection, hence resulting in a relative underestimation of the number of outliers in this analysis (Foll & Gaggiotti 2008). Hence, we think our data provides evidence that more loci are affected by selection over time in Midas cichlids.

Research on the lake whitefish system also supports the hypothesis that large genomic islands are affected in the early stages of divergence, and that their number and size tend to increase in more differentiated species pairs (Renaut *et al.* 2012) and a similar pattern has also been found in other taxa, like threespine stickleback (Roesti *et al.* 2012) and *Heliconius* butterflies (Nadeau *et al.* 2012). Owing to the anonymity of AFLP markers, we are not able to infer whether indeed more genomic regions are selected or if the regions of differentiation simply grow in size through divergence hitchhiking (Via & West 2008), thereby spanning multiple AFLP markers.

#### **Conclusion and outlook**

The replicated and extremely young radiations of Midas cichlid fishes provide a framework that allows us to investigate the footprints of divergent selection in a meaningful comparative approach. We found an increased number of outlier loci in the older, phenotypically more differentiated radiation and provided evidence for a role of divergent selection, acting partly on the same genomic regions, in the allopatric divergence of crater lake populations. We found no evidence for a parallel genetic basis across both crater lake radiations, albeit this could be because of limitations of the AFLP-methodology and needs further testing with other approaches. This comprehensive dataset strongly supports the independent colonization of both crater lakes from the great lakes. In addition, we provide evidence for a parallel primary divergence in both crater lakes between the limnetic and benthic habitats. This suggests that the same ecological forces drove the initial divergence and that the adaptive radiations of Midas cichlids evolve by the same rules, as has been suggested before for the repeated radiations of Anoles lizards (Losos *et al.* 1998; Losos 2009, 2010). Although the conclusions we derive are somewhat limited by the anonymity and coverage of our markers, our AFLP genome scans provided valuable results towards a general understanding of ecological speciation and parallel genetic evolution of phenotypes during repeated radiations (Elmer & Meyer 2011). Furthermore, ongoing studies will make use of reference genomes for cichlids and dense, mapped markers to further examine genome-wide patterns of

divergence-with-gene-flow in the Midas cichlid species complex.

## Acknowledgements

We thank some former members of the Meyer lab who participated in previous field expeditions from whom some samples were included in this study—in particular M. Barluenga, K. Stölting, W. Salzburger and T. Lehtonen. M. Foll kindly provided help with BayeScan. We thank F. Henning for valuable comments and discussions. We are grateful to three anonymous reviewers and the Editor for comments on the manuscript. Fishes were collected with the permission of MARENA, Nicaragua. A.F.K. is partially supported by the International Max Planck Research School (IMPRS) for Organismal Biology. This research was funded by support of the University of Konstanz to K.R.E. and A.M. and by grants of the Deutsche Forschungsgemeinschaft (DFG). Support by the European Research Council (ERC advanced grant GenAdap 293700) to A. M. is gratefully acknowledged.

## References

- Albert AYK, Sawaya S, Vines TH, *et al.* (2007) The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. *Evolution*, **62**, 76–85.
- Barlow GW (1976a) The Midas Cichlid in Nicaragua. In: Investigations of the ichthyology of Nicaraguan lakes (ed. Thorson TB), pp. 333–358. University of Nebraska Press, Lincoln.
- Barlow GW (1976b) The Midas Cichlid in Nicaragua. In: Investigations of the Ichthyofauna of Nicaraguan lakes (ed. Thorson TB), pp. 333–358. University of Nebraska Press, Lincoln, Nebraska.
- Barlow GW, Munsey JW (1976) The red devil-Midas-arrow cichlid species complex in Nicaragua. In: Investigations of the Ichthyofauna of Nicaraguan lakes (ed. Thorson TB), pp. 359–370. University of Nebraska Press, Lincoln, Nebraska.
- Barlow GW, Baylis JR, Roberts D (1976) Chemical analyses of some crater lakes in relation to adjacent Lake Nicaragua. In: Investigations of the Ichthyofauna of Nicaraguan lakes (ed. Thorson TB), pp. 17–20. University of Nebraska Press, Lincoln, Nebraska.
- Barluenga M, Meyer A (2004) The Midas cichlid species complex: incipient sympatric speciation in Nicaraguan cichlid fishes? *Molecular Ecology*, **13**, 2061–2076.
- Barluenga M, Meyer A (2010) Phylogeography, colonization and population history of the Midas cichlid species complex (*Amphilophus spp.*) in the Nicaraguan crater lakes. *BMC Evolutionary Biology*, **10**, 326.
- Barluenga M, Stoltz KN, Salzburger W, Muschick M, Meyer A (2006) Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature*, **439**, 719–723.
- Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, **23**, 38–44.
- Barton NH (2001) The role of hybridization in evolution. *Molecular Ecology*, **10**, 551–568.
- Barton NH, Charlesworth B (1984) Genetic revolutions, founder effects, and speciation. *Annual Review of Ecology and Systematics*, **15**, 133–164.
- Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.
- Bensch S, Akesson M (2005) Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology*, **14**, 2899–2914.
- Bolnick DI (2006) Multi-species outcomes in a common model of sympatric speciation. *Journal of Theoretical Biology*, **241**, 734–744.
- Bolnick DI, Fitzpatrick BM (2007) Sympatric speciation: models and empirical evidence. *Annual Review of Ecology, Evolution and Systematics*, **38**, 459–487.
- Bonin A, Bellemain E, Eidesen PB, *et al.* (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, **13**, 3261–3273.
- Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution*, **23**, 773–783.
- Borowsky R (2008) Restoring sight in blind cavefish. *Current Biology*, **18**, R23–R24.
- Butlin RK (2008) Population genomics and speciation. *Genetica*, **138**, 409–418.
- Butlin R, Debelle A, Kerth C, *et al.* (2012) What do we need to know about speciation? *Trends in Ecology & Evolution*, **27**, 27–39.
- Campbell D, Bernatchez L (2004) Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology and Evolution*, **21**, 945–956.
- Carson HL (1975) The genetics of speciation at the diploid level. *American Naturalist*, **109**, 83–92.
- Carson HL, Templeton AR (1984) Genetic revolutions in relation to speciation phenomena – the founding of new populations. *Annual Review of Ecology and Systematics*, **15**, 97–131.
- Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. *Evolution*, **21**, 550–570.
- Chan YF, Marks ME, Jones FC, *et al.* (2010) Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science*, **327**, 302–305.
- Cheng CD, White BJ, Kamdem C, *et al.* (2012) Ecological genomics of *Anopheles gambiae* along a latitudinal cline: a population-resequencing approach. *Genetics*, **190**, 1417–1432.
- Colbeck GJ, Turgeon J, Sirois P, Dodson JJ (2011) Historical introgression and the role of selective vs. neutral processes in structuring nuclear genetic variation (AFLP) in a circum-polar marine fish, the capelin (*Mallotus villosus*). *Molecular Ecology*, **20**, 1976–1987.
- Colosimo PF, Hosemann KE, Balabhadra S, *et al.* (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, **307**, 1928–1933.
- Coyne JA, Orr HA (2004) Speciation. Sinauer Associates, Sunderland.
- Deagle BE, Jones FC, Chan YF, *et al.* (2012) Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. *Proceedings of the Royal Society B*, **279**, 1277–1286.
- Dieckmann U, Doebeli M (1999) On the origin of species by sympatric speciation. *Nature*, **400**, 354–357.

- Dittmann MT, Roesti M, Indermaur A, *et al.* (2012) Depth-dependent abundance of Midas Cichlid fish (*Amphilophus spp.*) in two Nicaraguan crater lakes. *Hydrobiologia*, **686**, 277–285.
- Elmer KR, Meyer A (2011) Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends in Ecology & Evolution*, **26**, 298–306.
- Elmer KR, Lehtonen TK, Meyer A (2009) Color assortative mating contributes to sympatric divergence of neotropical cichlid fish. *Evolution*, **63**, 2750–2757.
- Elmer KR, Kusche H, Lehtonen TK, Meyer A (2010) Local variation and parallel evolution: morphological and genetic diversity across a species complex of neotropical crater lake cichlid fishes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **365**, 1763–1782.
- Elmer KR, Lehtonen TK, Fan S, Meyer A (2012) Crater lake colonization by Neotropical cichlid fishes. *Evolution*, doi:10.1111/j.1558-5646.2012.01755.x.
- Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics*, **28**, 342–350.
- Fischer MC, Foll M, Excoffier L, Heckel G (2011) Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). *Molecular Ecology*, **20**, 1450–1462.
- Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.
- Freedman AH, Thomassen HA, Buermann W, Smith TB (2010) Genomic signals of diversification along ecological gradients in a tropical lizard. *Molecular Ecology*, **19**, 3773–3788.
- Gavrilets S (2003) Perspective: models of speciation: what have we learned in 40 years? *Evolution*, **57**, 2197–2215.
- Gavrilets S, Losos JB (2009) Adaptive radiation: contrasting theory with data. *Science*, **323**, 732–737.
- Gavrilets S, Vose A (2005) Dynamic patterns of adaptive radiation. *Proceedings of the National Academy of Sciences USA*, **102**, 18040–18045.
- Gavrilets S, Vose A (2007) Case studies and mathematical models of ecological speciation. 2. Palms on an oceanic island. *Molecular Ecology*, **16**, 2910–2921.
- Gavrilets S, Vose A, Barluenga M, Salzburger W, Meyer A (2007) Case studies and mathematical models of ecological speciation. 1. Cichlids in a crater lake. *Molecular Ecology*, **16**, 2893–2909.
- Geiger MF, McCrary JK, Stauffer JR (2010) Description of two new species of the Midas cichlid complex (Teleostei: Cichlidae) from Lake Apoyo, Nicaragua. *Proceedings of the Biological Society of Washington*, **123**, 159–173.
- Günther ACLG (1864) On some new species of Central-American fishes. *Proceedings of the Zoological Society of London*, **1**, 23–27.
- Harmon LJ, Schulte JA, Larson A, Losos JB (2003) Tempo and mode of evolutionary radiation in iguanian lizards. *Science*, **301**, 961–964.
- Hoffman JL, Peck LS, Linse K, Clarke A (2011) Strong population genetic structure in a broadcast-spawning Antarctic Marine Invertebrate. *Journal of Heredity*, **102**, 55–66.
- Hoffmann AA, Rieseberg LH (2008) Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology, Evolution and Systematics*, **39**, 21–42.
- Hohenlohe PA, Bassham S, Etter PD, *et al.* (2010a) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.
- Hohenlohe PA, Phillips PC, Cresko WA (2010b) Using population genomics to detect selection in natural populations: key concepts and methodological considerations. *International Journal of Plant Sciences*, **171**, 1059–1071.
- Kirkpatrick M, Barton N (2006) Chromosome inversions, local adaptation and speciation. *Genetics*, **173**, 419–434.
- Klingenberg CP, Barluenga M, Meyer A (2003) Body shape variation in cichlid fishes of the *Amphilophus citrinellus* species complex. *Biological Journal of the Linnean Society*, **80**, 397–408.
- Kocher TD, Lee WJ, Sobolewska H, Penman D, McAndrew B (1998) A genetic linkage map of a cichlid fish, the tilapia (*Oreochromis niloticus*). *Genetics*, **148**, 1225–1232.
- Kolbe JJ, Leal M, Schoener TW, Spiller DA, Losos JB (2012) Founder effects persist despite adaptive differentiation: a field experiment with lizards. *Science*, **335**, 1086–1089.
- Kondrashov AS, Kondrashov FA (1999) Interactions among quantitative traits in the course of sympatric speciation. *Nature*, **400**, 351–354.
- Kutterolf S, Freundt A, Perez W, Wehrmann H, Schmincke HU (2007) Late Pleistocene to Holocene temporal succession and magnitudes of highly-explosive volcanic eruptions in west-central Nicaragua. *Journal of Volcanology and Geothermal Research*, **163**, 55–82.
- Lang M, Miyake T, Braasch I, *et al.* (2006) A BAC library of the East African haplochromin cichlid fish *Astatotilapia burtoni*. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, **306B**, 35–44.
- Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of theory of selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.
- Losos JB (2009) Lizards in an evolutionary tree: the ecology of adaptive radiation in anoles. University of California Press, Berkeley.
- Losos JB (2010) Adaptive radiation, ecological opportunity, and evolutionary determinism. *American Naturalist*, **175**, 623–639.
- Losos JB (2011) Convergence, adaptation, and constraint. *Evolution*, **65**, 1827–1840.
- Losos JB, Jackman TR, Larson A, de Queiroz K, Rodriguez-Schettino L (1998) Contingency and determinism in replicated adaptive radiations of island lizards. *Science*, **279**, 2115–2118.
- Majumdar KC, McAndrew BJ (1986) Relative DNA content of somatic nuclei and chromosomal studies in 3 Genera, Tilapia, Sarotherodon, and Oreochromis of the Tribe Tilapiini (Pisces, Cichlidae). *Genetica*, **68**, 175–188.
- Manel S, Conord C, Despres L (2009) Genome scan to assess the respective role of host-plant and environmental constraints on the adaptation of a widespread insect. *BMC Evolutionary Biology*, **9**, 288.
- Manousaki T, Hull PM, Kusche H, Machado-Schiaffino G, Franchini P, Harrod C, Elmer KR, Meyer A (2012) Parsing parallel evolution: trophic ecology, morphometrics and comparative transcriptomics in the repeated evolution of hypertrophied-lips in the Nicaraguan Midas cichlid fish adaptive radiations. *Molecular Ecology*, in press.
- Mayr E (1954) Change of genetic environment and evolution. In: *Evolution as a process* (eds. Huxley J, Hardy AC, Ford EB), pp. 157–180. Allen & Unwin, London.

- McKinnon JS, Rundle HD (2002) Speciation in nature: the threespine stickleback model systems. *Trends in Ecology & Evolution*, **17**, 480–488.
- McNiven VT, Levasseur-Viens H, Kanippayoor RL, Laturney M, Moehring AJ (2011) The genetic basis of evolution, adaptation and speciation. *Molecular Ecology*, **20**, 5119–5122.
- Meudt HM, Clarke AC (2007) Almost forgotten or latest practice? AFLP applications, analyses and advances *Trends in Plant Science*, **12**, 106–117.
- Meyer A (1990a) Ecological and evolutionary consequences of the trophic polymorphism in *Cichlasoma citrinellum* (Pisces, Cichlidae). *Biological Journal of the Linnean Society*, **39**, 279–299.
- Meyer A (1990b) Morphometrics and allometry in the trophically polymorphic cichlid fish, *Cichlasoma citrinellum* – alternative adaptations and ontogenetic changes in shape. *Journal of Zoology*, **221**, 237–260.
- Meyer A (1993) Phylogenetic-relationships and evolutionary processes in East-African cichlid fishes. *Trends in Ecology & Evolution*, **8**, 279–284.
- Michel AP, Sim S, Powell THQ, *et al.* (2010) Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences USA*, **107**, 9724–9729.
- Mueller UG, Wolfenbarger LL (1999) AFLP genotyping and fingerprinting. *Trends in Ecology & Evolution*, **14**, 389–394.
- Nadeau NJ, Jiggins CD (2010) A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends in Genetics*, **26**, 484–492.
- Nadeau NJ, Whibley A, Jones RT, *et al.* (2012) Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **367**, 343–353.
- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences USA*, **70**, 3321–3323.
- Niemiller ML, Nosil P, Fitzpatrick BM (2010) Corrigendum: recent divergence-with-gene-flow in Tennessee cave salamanders (Plethodontidae; *Gyrinophilus*) inferred from gene genealogies. *Molecular Ecology*, **19**, 1513–1514.
- Nosil P (2008) Speciation with gene flow could be common. *Molecular Ecology*, **17**, 2103–2106.
- Nosil P (2012) Ecological Speciation. Oxford University Press, New York, New York.
- Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **367**, 332–342.
- Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: “Isolation by adaptation” and multiple roles for divergent selection. *Evolution*, **62**, 316–336.
- Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.
- Paris M, Boyer S, Bonin A, *et al.* (2010) Genome scan in the mosquito *Aedes rusticus*: population structure and detection of positive selection after insecticide treatment. *Molecular Ecology*, **19**, 325–337.
- Parisod C, Joost S (2010) Divergent selection in trailing- versus leading-edge populations of *Biscutella laevigata*. *Annals of Botany*, **105**, 655–660.
- Paun O, Bateman RM, Fay MF, *et al.* (2010) Stable epigenetic effects impact adaptation in allopolyploid orchids (*Dactylorehiza*: Orchidaceae). *Molecular Biology and Evolution*, **27**, 2465–2473.
- Perez-Figueroa A, Garcia-Pereira MJ, Saura M, Rolan-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, **23**, 2267–2276.
- Pinho C, Hey J (2010) Divergence with gene flow: models and data. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 215–230.
- Przeworski M, Coop G, Wall JD (2005) The signature of positive selection on standing genetic variation. *Evolution*, **59**, 2312–2323.
- R Development Core Team (2010) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Renaut S, Maillet N, Normandeau E, *et al.* (2012) Genome-wide patterns of divergence during speciation: the lake whitefish case study. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **367**, 354–363.
- Robinson BW, Wilson DS (1994) Character release and displacement in fishes – a neglected literature. *American Naturalist*, **144**, 596–627.
- Roesti M, Hendry AP, Salzburger W, Berner D (2012) Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology*, **21**, 2852–2862.
- Rogers SM, Bernatchez L (2005) Integrating QTL mapping and genome scans towards the characterization of candidate loci under parallel selection in the lake whitefish (*Coregonus clupeaformis*). *Molecular Ecology*, **14**, 351–361.
- Rogers SM, Bernatchez L (2007) The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonus* sp Salmonidae) species pairs. *Molecular Biology and Evolution*, **24**, 1423–1438.
- Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters*, **8**, 336–352.
- Saitou N, Nei M (1987) The neighbor-joining method – a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–425.
- Salzburger W, Meyer A (2004) The species flocks of East African cichlid fishes: recent advances in molecular phylogenetics and population genetics. *Naturwissenschaften*, **91**, 277–290.
- Schluter D (1996) Ecological speciation in postglacial fishes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **351**, 807–814.
- Schluter D (2000) *The Ecology of Adaptive Radiation*, 1. publ. edn. Oxford University Press, Oxford.
- Schlüter PM, Harris SA (2006) Analysis of multilocus fingerprinting data sets containing missing data. *Molecular Ecology Notes*, **6**, 569–572.
- Schluter D, McPhail JD (1993) Character displacement and replicate adaptive radiation. *Trends in Ecology & Evolution*, **8**, 197–200.
- Smadja CM, Butlin RK (2011) A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology*, **20**, 5123–5140.
- Stapley J, Reger J, Feulner PG, *et al.* (2010) Adaptation genomics: the next generation. *Trends in Ecology & Evolution*, **25**, 705–712.



- Stauffer JJR, McKaye KR (2002) Descriptions of three new species of cichlid fishes (Teleostei: Cichlidae) from Lake Xiloá, Nicaragua. *Cuadernos de Investigación de la Universidad Centroamericana*, **12**, 1–18.
- Stauffer JJR, McCrary JK, Black KE (2008) Three new species of cichlid fishes (Teleostei: Cichlidae) from Lake Apoyo, Nicaragua. *Proceedings of the Biological Society of Washington*, **121**, 117–129.
- Steiner CC, Rompler H, Boettger LM, Schoneberg T, Hoekstra HE (2009) The genetic basis of phenotypic convergence in Beach Mice: similar pigment patterns but different genes. *Molecular Biology and Evolution*, **26**, 35–45.
- Stiassny MLJ, Meyer A (1999) Cichlids of the Rift lakes. *Scientific American*, **280**, 64–69.
- Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, **100**, 158–170.
- Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671–688.
- Strasburg JL, Rieseberg LH (2010) How robust are “Isolation with Migration” analyses to violations of the IM model? A simulation study. *Molecular Biology and Evolution*, **27**, 297–310.
- Strasburg JL, Rieseberg LH (2011) Interpreting the estimated timing of migration events between hybridizing species. *Molecular Ecology*, **20**, 2353–2366.
- Streelman JT, Danley PD (2003) The stages of vertebrate evolutionary radiation. *Trends in Ecology & Evolution*, **18**, 126–131.
- Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, **144**, 389–399.
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.
- Templeton AR (1980) The theory of speciation via the founder principle. *Genetics*, **94**, 1011–1038.
- Vekemans X, Beauwens T, Lemaire M, Roldan-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology Notes*, **11**, 139–151.
- Via S (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution*, **16**, 381–390.
- Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences*, **106**, 9939–9946.
- Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, **17**, 4334–4345.
- Vivas R, McKaye KR (2001) Habitat selection, feeding ecology, and fry survivorship in the *Amphilophus citrinellus* species complex in Lake Xiloá, Nicaragua. *Journal of Aquaculture and Aquatic Sciences*, **IX**, 32–48.
- Voight BF, Kudaravalli S, Wen XQ, Pritchard JK (2006) A map of recent positive selection in the human genome. *PLoS Biology*, **4**, e72.
- Vos P, Hogers R, Bleeker M, *et al.* (1995) AFLP – a new technique for DNA-fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.
- Wake DB, Wake MH, Specht CD (2011) Homoplasy: from detecting pattern to determining process and mechanism of evolution. *Science*, **331**, 1032–1035.
- Whitlock R, Hipperson H, Mannarelli M, Butlin RK, Burke T (2008) An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. *Molecular Ecology Resources*, **8**, 725–735.
- Wright S (1965) The interpretation of population-structure by F-statistics with special regard to systems of mating. *Evolution*, **19**, 395–420.
- Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.
- Zhivotovsky LA (1999) Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology*, **8**, 907–913.

---

A.F.K. generated and analysed the AFLP data. A.M. and K.R.E. collected specimens. K.R.E. and A.M. designed the experiment. All authors contributed to manuscript preparation.

---

## Data accessibility

Binary AFLP data matrix: DRYAD entry: doi: 10.5061/dryad.vh1qk.

## Supporting information

Additional Supporting Information may be found in the online version of this article.

**Fig. S1** Population-level neighbor-joining trees based (a) on only the outlier loci and the (b) presumably neutral data set. Nodes show per cent bootstrap support (1000 replicates).

**Table S1** Information about the 13 used selective primer combinations.

**Table S2** Pairwise  $F_{ST}$ -values based on the presumably neutral data set (951 AFLP loci).

**Table S3** Pairwise  $F_{ST}$ -values based on the outlier data set (79 AFLP loci).

**Table S4** Species-specific allele frequencies at outlier loci as estimated with BayeScan.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.