

Genomic architecture of ecologically divergent body shape in a pair of sympatric crater lake cichlid fishes

PAOLO FRANCHINI,*§ CARMELO FRUCIANO,*§ MARIA L. SPREITZER,* JULIA C. JONES,*† KATHRYN R. ELMER,*‡ FREDERICO HENNING* and AXEL MEYER*

*Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, Universitätsstraße 10, 78457 Konstanz, Germany, †Zukunftskolleg, University of Konstanz, 78457 Konstanz, Germany

Abstract

Determining the genetic bases of adaptations and their roles in speciation is a prominent issue in evolutionary biology. Cichlid fish species flocks are a prime example of recent rapid radiations, often associated with adaptive phenotypic divergence from a common ancestor within a short period of time. In several radiations of freshwater fishes, divergence in ecomorphological traits — including body shape, colour, lips and jaws — is thought to underlie their ecological differentiation, specialization and, ultimately, speciation. The Midas cichlid species complex (*Amphilophus* spp.) of Nicaragua provides one of the few known examples of sympatric speciation where species have rapidly evolved different but parallel morphologies in young crater lakes. This study identified significant QTL for body shape using SNPs generated via ddRAD sequencing and geometric morphometric analyses of a cross between two ecologically and morphologically divergent, sympatric cichlid species endemic to crater Lake Apoyo: an elongated limnetic species (*Amphilophus zaliosus*) and a high-bodied benthic species (*Amphilophus astorquii*). A total of 453 genome-wide informative SNPs were identified in 240 F_2 hybrids. These markers were used to construct a genetic map in which 25 linkage groups were resolved. Seventy-two segregating SNPs were linked to 11 QTL. By annotating the two most highly supported QTL-linked genomic regions, genes that might contribute to divergence in body shape along the benthic limnetic axis in Midas cichlid sympatric adaptive radiations were identified. These results suggest that few genomic regions of large effect contribute to early stage divergence in Midas cichlids.

Keywords: ddRADseq, geometric morphometrics, limnetic-benthic, quantitative trait loci, sympatric speciation

Introduction

Rapid radiations of species are associated with adaptive phenotypic divergence from a common ancestor in a short period of time (Schluter 2000). Determining the genetic basis of such adaptations is very much a focus of contemporary evolutionary biology research

(reviewed in Elmer & Meyer 2011). One of the major open questions in this field is to determine whether adaptive speciation is typically driven by a few genes of large effect or by many genes with small effects (e.g. Michel *et al.* 2010; Nosil & Feder 2012). Theoretical models predict that phenotypic and fitness effect sizes have an exponential distribution with few alleles of large effect and many of small effects (Orr 1999). However, much of the work underlying these predictions does not take into account divergence with gene flow (Smadja & Butlin 2011; Feder *et al.* 2012) such as would be expected in situations of sympatric speciation by disruptive selection (Gavrilets 2003; Gavrilets *et al.* 2007).

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

‡ Present address: Institute of Biodiversity, Animal Health & Comparative Medicine, College of Medical, Veterinary & Life Sciences, University of Glasgow, Glasgow, G12 8QQ, UK

§These authors contributed equally to this work.

According to theoretical models, the few genes under selection or physically linked to loci experiencing strong disruptive selection can diverge, whereas gene flow will homogenize the remainder of the genome, resulting in genomic islands of speciation (e.g. Turner *et al.* 2005; Noor & Feder 2006; Feder *et al.* 2012; Via 2012). Recently, it has been shown (Michel *et al.* 2010) that in *Rhagoletis pomonella* an insect model species for sympatric ecological speciation divergence was driven by selection on numerous genomic regions as patterns of genetic differentiation in geographical space were present in most of the 33 microsatellite markers analysed and the observed patterns of divergence could not be explained by genetic drift or isolation by distance. In the same species, Via *et al.* (2012) proposed 10 genomic regions spread throughout all linkage groups where divergence outliers cluster around divergently selected QTL. However, in Atlantic cod, a single large genomic region, specific to differentiation between two interbreeding ecotypes, has been identified (Hemmer-Hansen *et al.* 2013). How pervasive these two alternative genomic patterns of divergence with gene flow are is still unknown and will require more work on more species and ecological circumstances. More genome-wide studies are required to better understand the genetic mechanisms driving speciation at different stages, to link genotype and phenotype and to understand how the genetic architecture of the diverging populations is shaped during this process (reviewed in Elmer & Meyer 2011).

In particular, the identification of the number, size and distribution of genomic regions underlying the evolution of divergent adaptive phenotypes represents an important first step in the analysis of the genomic bases of adaptive divergence (Bernatchez *et al.* 2010). Once genomic regions associated with divergent adaptive phenotypes are identified, their levels of genetic differentiation in natural populations can be compared with the levels of differentiation in the rest of the genome to determine the size of the affected genomic regions and how strongly they have been impacted by selection (Rogers & Bernatchez 2005; Renaut *et al.* 2012).

Recent advances in protocols of next-generation sequencing techniques (Baird *et al.* 2008; Peterson *et al.* 2012) and bioinformatic analyses now allow genome-wide scans of genomic variation. High-throughput next-generation sequencing approaches, such as restriction site-associated DNA sequencing (RAD genotyping), enable the detection and genotyping of thousands of single nucleotide polymorphisms (SNPs) distributed throughout the genome (Baird *et al.* 2008). The advantage of these approaches is that now genome-wide information can be screened for a large number of individuals even in nonmodel organisms (Hohenlohe *et al.*

2010). These tools allow the construction of very dense genetic linkage maps, subsequent population genomic assessments and phylogenetic analyses (e.g. Hohenlohe *et al.* 2010; Chutimanitsakun *et al.* 2011; Jones *et al.* 2013). Specifically, the double-digest (dd) RAD method (Peterson *et al.* 2012) allows targeted robust recovery of homologous genomic regions even without a reference genome.

Interestingly, the first axis of differentiation in many freshwater systems appears to be the limnetic benthic one, which is normally reflected in body shape, that is, in a nutshell, how 'stretched-out' a fish is. Adaptive divergence into benthic and limnetic forms has been documented in a range of freshwater fishes, including sticklebacks (Larson 1976; McPhail 1984; Wootton 2009; Willacker *et al.* 2010), whitefish and other salmonids (Fenderson 1964; McCart 1970; Bernatchez *et al.* 1999, 2010; Landry *et al.* 2007; Bertrand *et al.* 2008; Gowell *et al.* 2012; Siwertsson *et al.* 2013), perch (Svanbäck & Eklöv 2002; Kekäläinen *et al.* 2010), but also in Neotropical (Meyer 1990a,b) and African cichlids (Hulsey *et al.* 2013).

In sticklebacks, a linkage map of sympatric benthic and limnetic forms found that different chromosomal regions seemed to affect components of skeletal anatomy, yet some functional traits mapped to similar chromosome regions, suggesting that genetic linkage or pleiotropy might also play an important role (Pechel *et al.* 2001). In benthic limnetic species pairs of stickleback, Jones *et al.* (2012) identified divergence in four genomic regions consistent across all the three lakes they studied, also putatively associated with skin pigmentation and pathogen resistance. However, all such studies would benefit from the application of multiple approaches (genome scans, analyses of gene expression, QTL studies) on the same sets of populations to better understand the link between phenotype and genomic island of divergence (Hendry *et al.* 2013). Studies on whitefish using the same population pairs for both QTL mapping and genome scans (Rogers & Bernatchez 2005, 2007; Renaut *et al.* 2012; Gagnaire *et al.* 2013) reported a colocalization of regions associated with phenotypic variation and regions of reduced gene flow between sympatric benthic limnetic forms. These studies have also suggested that the early stages of speciation in whitefish are characterized by large islands of divergence and that there is a correlation between the overall level of differentiation and the size and number of genomic regions of divergence (Renaut *et al.* 2012). However, taxonomically broader comparisons are needed to further understand the fundamental genetic patterns underlying this ubiquitously used limnetic benthic axis of divergence among freshwater fishes.

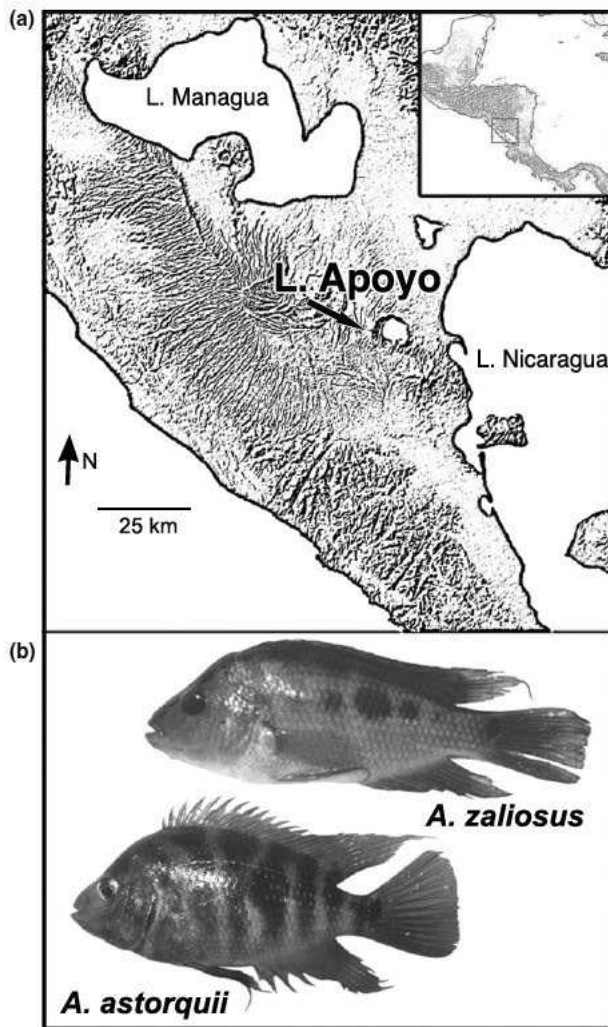


Fig. 1 (a) Map of the Nicaraguan main lakes and Lake Apoyo. (b) Two typical specimens of *Amphilophus astorquii* and *A. zaliosus*.

Cichlid fish have been widely used in studies of biological diversification due to their extremely fast rates of evolution and spectacular adaptive radiations (Meyer *et al.* 1990; Meyer 1993; Kocher 2004; Seehausen 2006; Kuraku & Meyer 2008). The Midas cichlid species complex, a Neotropical lineage that comprises 13 currently described species in the genus *Amphilophus*, is a fitting model system for the study of adaptive radiation and sympatric speciation (e.g. Barluenga *et al.* 2006; Elmer *et al.* 2010a; Recknagel *et al.* 2013a). Midas cichlids occur in Nicaragua chiefly in two large basins, Lake Nicaragua and Lake Managua, and in a number of smaller, younger crater lakes that formed in dormant volcanoes (Fig. 1). In these young crater lakes, at least two adaptive radiations of different ages have formed (Barluenga & Meyer 2004, 2010; Barluenga *et al.* 2006; Fan *et al.*

2012), repeatedly evolving forms polymorphic in colour (Meek 1907; Barlow 1976; Villa 1976; Henning *et al.* 2010, 2013), lip morphology (Barlow & Munsey 1976; Elmer *et al.* 2010c; Manousaki *et al.* 2013), lower pharyngeal jaw size and shape (Meyer 1989, 1990a,b; Klingenberg *et al.* 2002; Barluenga *et al.* 2006), and body shape along the deep-elongated axis of variation typically associated with benthic limnetic divergence (Meyer 1990a,b; Klingenberg *et al.* 2003; Barluenga *et al.* 2006; Elmer *et al.* 2010a).

The oldest (about 22 000 years; Kutterolf *et al.* 2007) and largest of these crater lakes is Lake Apoyo, in which one open water limnetic species *Amphilophus zaliosus* and five benthic species have arisen rapidly, likely via ecological disruptive selection causing sympatric speciation, from a Lake Nicaragua '*Amphilophus citrinellus*' ancestor (Barluenga *et al.* 2006; Stauffer *et al.* 2008; Barluenga & Meyer 2010; Elmer *et al.* 2010a; Geiger *et al.* 2010). *A. zaliosus* diverged soon after the origin of Lake Apoyo and *A. astorquii* one of the benthic species endemic to Lake Apoyo is an even younger species that diverged more recently during the formation of the radiation of Midas cichlids in crater Lake Apoyo (Kautt *et al.* 2012). Not all aspects of these incipient adaptive radiations via sympatric speciation are understood yet (Barluenga *et al.* 2006; Gavrilets *et al.* 2007; Elmer *et al.* 2009, 2010a), and in particular, the genetic basis of divergence and adaptation between these sympatric, ecologically distinct species, has not been determined.

In a previous transcriptomic analysis, these two species, *A. zaliosus* and *A. astorquii*, were compared, and 14 candidate genes showing signs of positive selection were identified (Elmer *et al.* 2010b). Variation in external body morphology is important in adaptation and has been shown to be associated with ecological divergence in sticklebacks (Walker 1997; Walker & Bell 2000; Aguirre & Bell 2012; Kaeuffer *et al.* 2012) the best-known text-book model in ecological speciation as well as in other fish species (Langerhans *et al.* 2007; Roy *et al.* 2007; Volanthen *et al.* 2009). But so far, the only attempts to explore the genetic basis of body shape differences using dense linkage maps have been made in sticklebacks (Albert *et al.* 2008; Malek *et al.* 2012; Rogers *et al.* 2012) and salmon (Boulding *et al.* 2008).

Here, we aim to extend our understanding of the genetic underpinnings of ecological divergence in sympatry by analysing through QTL mapping the genetic basis of three morphological and putatively adaptive traits: body shape, pectoral fin base size and caudal peduncle depth. Body shape was chosen as in previous studies benthic and limnetic species in Lake Apoyo showed divergence in this particular trait (Barluenga *et al.* 2006; Stauffer *et al.* 2008; Elmer *et al.*

2010a). Pectoral fin base size and caudal peduncle depth are, besides body shape, also important characteristics for swimming performance and manoeuvrability (Pettersson & Brönmark 1999; Blake *et al.* 2005; Rouleau *et al.* 2010). Our study is the first QTL analysis of external body morphology, based on a dense linkage map, in fish species that have speciated sympatrically (as sticklebacks have not speciated sympatrically). We also functionally annotated the QTL regions detected by the mapping analyses in an effort to identify a set of candidate genes that might be responsible for the divergence in external body morphology between benthic and limnetic Midas cichlids from Lake Apoyo.

Materials and methods

The present study is based on wild-caught *Amphilophus astorquii* and *Amphilophus zalius* that were collected in Lake Apoyo, Nicaragua, in 2005 and 2007 as juveniles and subsequently reared under laboratory conditions. Fishes were collected and exported with permission of MARENA (Nicaragua).

Genetic crosses

The genetic cross is based on a wild-caught female *A. astorquii* with a wild-caught male *A. zalius*. Eggs of the P_0 were removed from the parental tank once spawned. Upon maturity (ca. 1 year of age), we randomly chose one pair (F_1) as it formed, isolated them into a different tank and allowed them to breed. This F_1 pair produced a total of 240 F_2 individuals. When the F_2 individuals were 18 months of age, all fish were photographed in a standardized manner for morphometric analyses, and a fin clip was taken and preserved in ethanol for molecular analyses. To characterize the external morphology of wild-caught, laboratory-reared *A. astorquii* ($n = 16$) and *A. zalius* ($n = 41$), we took standardized photographs of adults and conducted both traditional and geometric morphometric analyses.

Phenotypic analysis of laboratory-reared populations

From each photograph, 20 homologous landmarks and semi-landmarks (Fig. S1, Supporting information) were digitized by a single investigator (MLS) using the software tpsDig v2.16 (Rohlf 2006). The landmark/semi-landmark configurations obtained were then subjected to a generalized Procrustes analysis with sliding of semilandmarks (Bookstein 1997) in tpsRelw v1.49 (Rohlf 2007a) using the minimization of the squared Procrustes distance as the sliding criterion (Perez *et al.* 2006). Centroid size estimates for each specimen were also computed. Non biologically-relevant shape variation

due to dorso-ventral arching of the fish body was modelled in this study as a shape change vector (Valentin *et al.* 2008; Fruciano *et al.* 2011, 2012) using a random subset of 10 specimens each measured at five different levels of dorso-ventral arching (10 shape change vectors, one per model specimen, were obtained, and then, a single shape change vector was computed by averaging them). This shape variation was then removed from the data set by projecting the shape variables constituting the data set to the multivariate subspace orthogonal and therefore uncorrelated to such a vector (Valentin *et al.* 2008; Fruciano *et al.* 2011, 2012). To remove the allometric component of shape variation, we performed a multivariate regression of shape on centroid size in MorphoJ v1.05f (Klingenberg 2011) and used the regression residuals in all downstream analyses.

To test the hypothesis that body shape differed between adult individuals of the two species of Midas cichlids used here, despite being raised under the same laboratory conditions, we used the permutational procedure based on Procrustes distances implemented in MorphoJ (1000 permutations). Differences in mean shape between species were visualized with a wire-frame graph.

To analyse specific external body morphological traits that have a well-understood effect on swimming performance (Pettersson & Brönmark 1999; Blake *et al.* 2005; Rouleau *et al.* 2010) and hence are expected to be of ecological relevance, linear measurements were also calculated from landmarks for the pectoral fin base size and the depth of the caudal peduncle. We used ANCOVAs on each log-transformed linear measurement using the logarithm of centroid size as a covariate (thus controlling for allometry) to test for the difference in these traits among the two wild-caught laboratory-reared groups.

Phenotypic analyses for QTL

QTL studies commonly use scores of individuals along individual principal component axes as a measure of shape variation (Carrier *et al.* 2005; Mezey *et al.* 2005). However, scores along individual principal components are not necessarily biologically relevant. A plot of the scores of the parental groups and F_2 individuals along the first two principal component axes (Fig. 2) clearly reveals that the biologically relevant variation in shape between *A. astorquii* and *A. zalius* is not parallel to the first or the second principal component axis, so scores along one of these axes would not have biological meaning. For this reason, we used between-group principal component analysis (Boulesteix 2005; Mitteroecker & Bookstein 2011) to compute a between-species principal component and obtained the projection of each F_2 individual along this axis with NTSYSpc v2.200

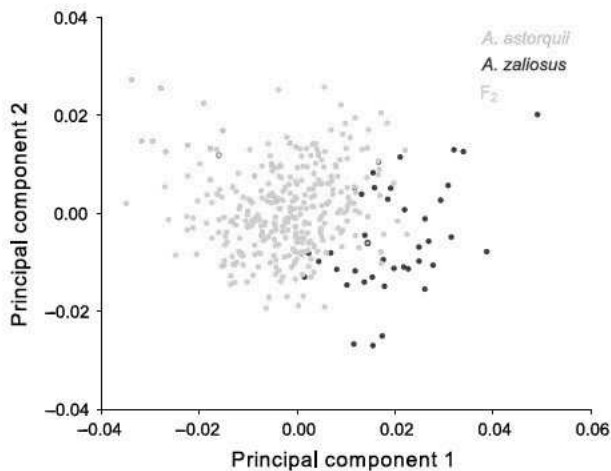


Fig. 2 Scatterplot of scores along the first two principal components of a PCA based on all individuals used in the present study. Empty circles represent the individuals of each parental species used for the cross.

(Rohlf 2007b) (individual scores can thus be interpreted as a measure of how much a given F_2 individual is 'zaliosus-like' or 'astorquii-like'). We used the scores along this axis as a phenotypic measure for QTL mapping. For QTL mapping, we also used a measure of body shape partially redundant with scores along the between-group principal component: the ratio of body depth (distance between landmarks 6 and 9) and standard length (distance between landmarks 1 and 14).

For QTL mapping of linear measures of pectoral fin base width, caudal peduncle height and the ratio body depth/standard length, we used residuals from the regression of each log-transformed linear measurement on the logarithm of centroid size, to control for allometry.

Molecular methods for genotyping

Genomic DNA was extracted from fin tissue for the two P_0 , two F_1 and 240 F_2 individuals using the Qiagen DNeasy Blood & Tissue Kit (Qiagen, Valencia, USA) following the manufacturer's protocol including the RNase treatment to remove residual RNA. The DNA integrity of each sample was assessed by agarose gel electrophoresis and quantified using a Qubit v2.0 fluorometer (Life Technologies, Darmstadt, Germany). Approximately 300 ng of DNA template of each sample was double-digested using the restriction enzymes *Pst*I-HF and *Msp*I (New England BioLabs, Beverly, USA) in one combined reaction. The double-digest restriction site-associated DNA (ddRAD) library preparation protocol followed the methodology originally described in Peterson *et al.* (2012) with the modifications introduced by Recknagel *et al.* (2013b). Briefly, after digestion, each

sample was ligated to barcoded Illumina primers and then pooled at equimolar concentrations. A total of five RAD libraries were prepared, each including 50 individuals. Each F_2 individual was included in one library ($5 \times 48 = 240 F_2$ s), the P_0 was included in three libraries and the F_1 samples in two libraries to maximize the depth of sequencing of these catalogue loci. The libraries were size-selected for a range of 335–405 bp using a Pippin Prep[®] electrophoresis system (Sage Science, Beverly, USA). After cluster amplification in a cBOT, the ddRAD libraries were run on an Illumina Genome Analyzer IIx (GAIIx) at the Genomics Center University of Konstanz (GeCKo) using a single-end protocol with 120 cycles following the manufacturer's protocol. Because of the low base composition diversity in the first 10 bp (barcode and restriction enzyme sequences) inherent to RAD samples, a separate PhiX genome control lane was used to improve cross-talk and phasing calculations.

Marker detection

Raw Illumina reads were processed into candidate RAD loci (i.e. RAD stacks) using the open source Stacks pipeline v0.99997 (Catchen *et al.* 2011). Reads were separated by barcode, cleaned from erroneous and low-quality reads and truncated to a length of 100 bp using the Stacks component 'process radtags'. The 'denovo map.pl' Stacks Perl script (which sequentially runs the Stacks component 'ustacks', 'cstacks' and 'sstacks') was used to process the 244 individual's sets of filtered reads in order to (i) build loci in each sample and calling SNPs in each 'ustacks'; (ii) create a catalogue of loci for the parents and the progeny 'cstacks'; (iii) match each catalogue of loci of the sample against the parent catalogues 'sstacks'. The denovo map.pl parameters were set as the default except for the following parameters: minimum number of identical raw reads required to create a stack in each parent (-m 3) and in each progeny (-P 3); number of mismatches allowed between loci from the parents (-n 3); removal or disjoining of highly repetitive RAD tags (-t). Multiple SNPs in one RAD stack were treated as a single polymorphic marker. Finally, the progeny genotypes present in at least 100 individuals were exported using the Stacks 'export sql.pl' Perl script.

We used a marker selection procedure to obtain a set of reliable markers to be used in downstream linkage map construction and QTL mapping. Of the polymorphic loci shared among parents, F_1 and F_2 individuals, we first removed loci for which both parents were heterozygous or one of the two parent's genotype was missing. Then, based on frequencies of genotypes in the F_2 progeny (χ^2 test threshold $P = 0.001$), we removed: a)

noninformative markers that are not heterozygous in any of the F_1 s and that generate an aa:ab:bb ratio of 1:0:0, 0:0:1 or 0:1:0; b) unmappable markers where the parents share an allele (i.e. they are aa \times ab or ab \times aa), and one F_1 inherited the P_1 - (or P_2 -) exclusive allele, but the other did not, thus producing a test-cross (ab \times aa) that generates a 1:1 ratio; c) under severe segregation distortion, which most likely reflects bad marker quality and technical artefacts. To assess the robustness of our results with different levels of stringency in marker selection, we also used a more stringent threshold for either segregation distortion (step c above, χ^2 test $P < 0.01$) or proportion of missing data (markers present in at least 80% of individuals).

Linkage map construction

A biparental F_2 linkage map was constructed using the F_2 design of JoinMap v4 software (Van Ooijen 2006). The RAD-based approach generates large amounts of missing data and incorrect genotypes (i.e. miscalled homozygotes) (Ward *et al.* 2013). This can lead to incorrect ordering at small map distances and inflated map size. An algorithm to correct these mistakes has been recently developed, but is not yet applicable for biparental maps (Ward *et al.* 2013). To address these issues and maintain quality control, markers were used for linkage map construction were selected with the procedure described above to minimize the impact of missing data on linkage map estimation and QTL detection (Fig. S2, Supporting information). The parameters used for linkage map construction were as follows: a linkage LOD threshold above 14.0 was used to group loci on linkage groups (LGs). Loci were ordered using the Kosambi function with LOD >3 and recombination frequencies <0.40 . This generated 30 LGs, which were reduced to 25 by sequentially assigning markers to the groups of highest and strongest cross-link (SCL) using a SCL-LOD threshold of 10.0.

QTL mapping

QTL analyses were performed for the three traits described above (body shape, fin base width and caudal peduncle depth) using the RAD markers from which the linkage map was built. Interval mapping (IM) analysis was conducted using the MapQtl v5 program (Van Ooijen 2004) with a 1 cM increment. LOD thresholds for testing the significance of QTL peaks were calculated using 1000 permutations for each of the trait data sets and a significance level of $P < 0.05$. To increase the precision of the QTL mapping, markers with LOD scores exceeding the calculated threshold were selected as cofactors, then the automatic cofactor

selection routine (ACS) was applied to retain a cofactor among the previously selected ones for each linkage group (P -value for automatic cofactor selection = 0.02). The Multiple QTL mapping analysis (MQM) was finally applied using the selected cofactors. QTL were designated for those peaks that reached the genome-wide and chromosome-wide significance level of 0.05 determined by permutations.

Gene prediction and functional annotation

To annotate the detected QTL regions, the Nile tilapia (*Oreochromis niloticus*), an African cichlid fish, genome (Broad Institute, MIT, USA) was used as a reference (Orenil1.0, Ensembl v72). To retrieve the genome sequences of the highly supported QTL regions (exceeding the genome-wide significance threshold), the following pipelines were used: (i) the RAD loci significant at the genome-wide level and the most proximal loci significant at the chromosome level (CL) (one upstream and one downstream to the genome-wide significant loci considering the linkage map distance) were aligned to the draft of the Midas genome (unpublished data) using the BLASTn algorithm; (ii) a custom Perl script was used to extract, for each locus, a sequence of 1000 bp from the Midas genome (respectively, 450 bp downstream and upstream relative to the BLAST-hit coordinates); (iii) the 1000 bp sequences thus obtained were aligned to the tilapia genome using BLASTn and the region spanning the BLAST-hits was retained; (iv) gene sequences were retrieved from these regions using the available gene prediction of the tilapia genome. Blast2GO (Conesa *et al.* 2005) was used to perform the functional annotation of the genes included in the QTL regions. To test for the presence of significant GO term frequency differences in the genes occurring in the QTL regions, each QTL-linked gene set was compared with a baseline including all tilapia genes using the enrichment analysis implemented in Blast2Go using the Fisher's exact test (Fisher 1934) and setting the false discovery rate to 0.05 (Benjamini & Yekutieli 2001). The enrichment analysis was also performed comparing the two QTL-linked gene sets.

Results

Phenotypic variation between species and among F_2 individuals

We found a significant difference in mean body shape between the two parental (laboratory-reared) species *Amphilophus astorquii* and *Amphilophus zaliosus* (Procrustes distance 0.041; $P < 0.001$). The difference in mean shape between the two groups reflected the differences between *A. astorquii* and *A. zaliosus* previously

documented from wild adult specimens (Elmer *et al.* 2010a), with *A. zaliosus* having a more elongate and laterally compressed body shape (Fig. 3a). The between-group principal component scores used for QTL mapping reflect mainly variation in body depth (Fig. 4). The laboratory-reared *A. astorquii* and *A. zaliosus* also differed significantly in their width of the base of the pectoral fin (ANCOVA, $F_{1,54} = 14.56$ $P = 0.0003$) and depth of caudal peduncle (ANCOVA, $F_{1,54} = 20.79$ $P = 0.0003$), with *A. astorquii* having a larger pectoral fin base and a deeper caudal peduncle that is typical of a more benthic morphology (Fig. 3b,c).

Marker detection

Illumina sequencing generated 178 325 350 raw reads, averaging 35 729 707 reads per library across five libraries (standard deviation: 2 318 057). This number was reduced to 150 278 093 sequences (averaging

30 055 619 reads per library; standard deviation 2 358 130) after the cleaning pipeline was implemented. For the female and male parental samples, 1 422 347 and 1 740 156 reads were obtained, respectively. For the F_1 hybrid female and the F_1 hybrid male, the number of reads was 961 724 and 1 042 071, respectively. On average, the F_2 progeny libraries provided 604 632 reads per individual. Twenty F_2 individuals were removed from the final progeny data set because of low coverage, leaving a final F_2 progeny data set of 220 individuals (Table S1, Supporting information). We identified 165 915 loci in the sire and 165 517 loci in the dam. Of these, respectively, 9299 and 8165 were polymorphic and 2056 of the polymorphic loci were shared among parents, F_1 s and F_2 progeny. Of these, 545 markers were excluded because they were not informative for downstream QTL analyses (both parents were heterozygous or one of the two parent's genotype was missing). The remaining potentially informative 1 511

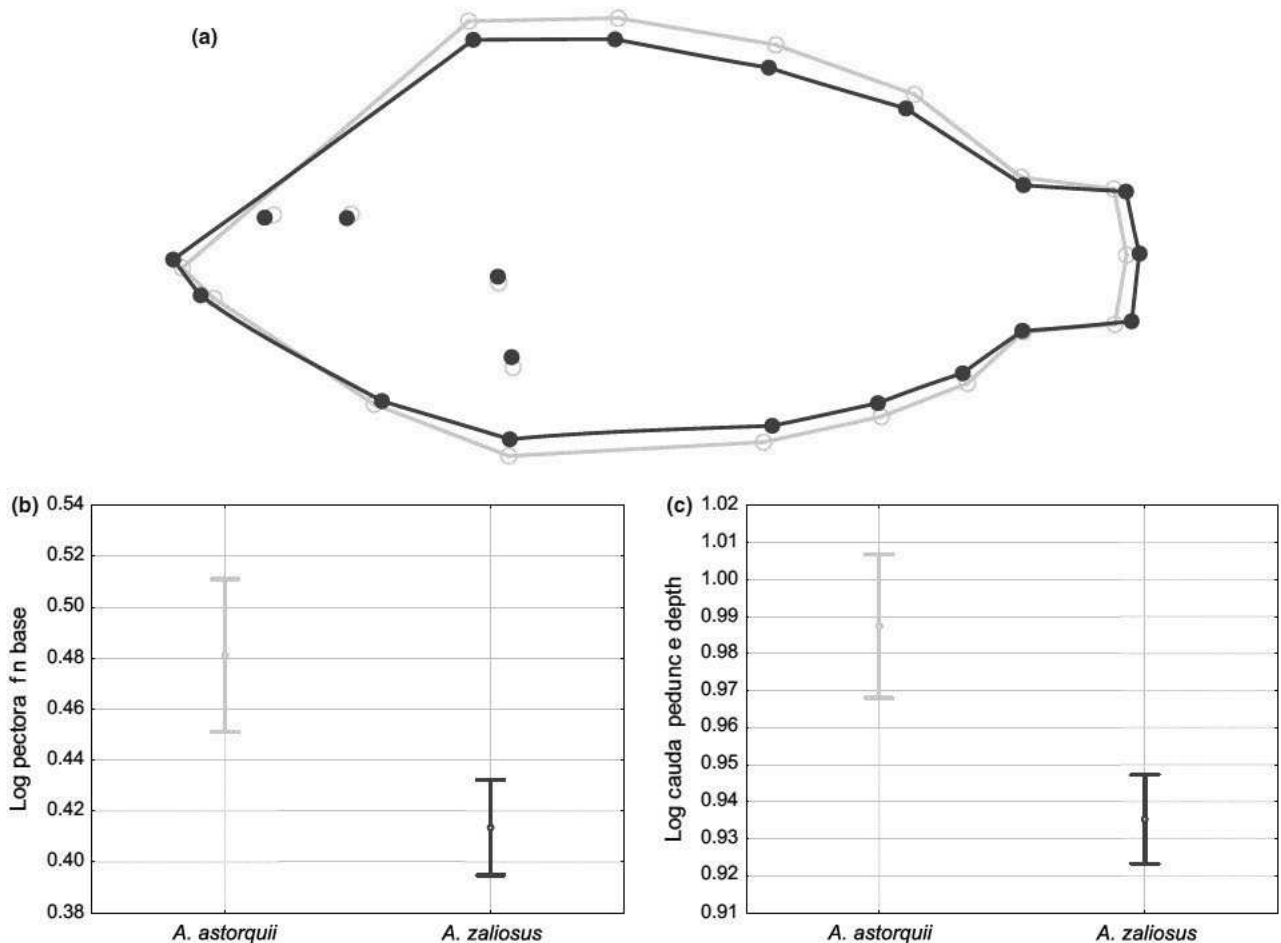


Fig. 3 (a) Body shape variation between wild caught, laboratory reared *A. astorquii* and *Amphilophus zaliosus*. Green dots represent *A. astorquii* average shape ($n = 16$), the blue dots represent *A. zaliosus* ($n = 41$). Least squares means estimated from the ANCOVAs of pectoral fin base size (b) and caudal peduncle depth (c) at the average covariate value (logarithm of centroid size). Error bars represent 0.95 confidence intervals.

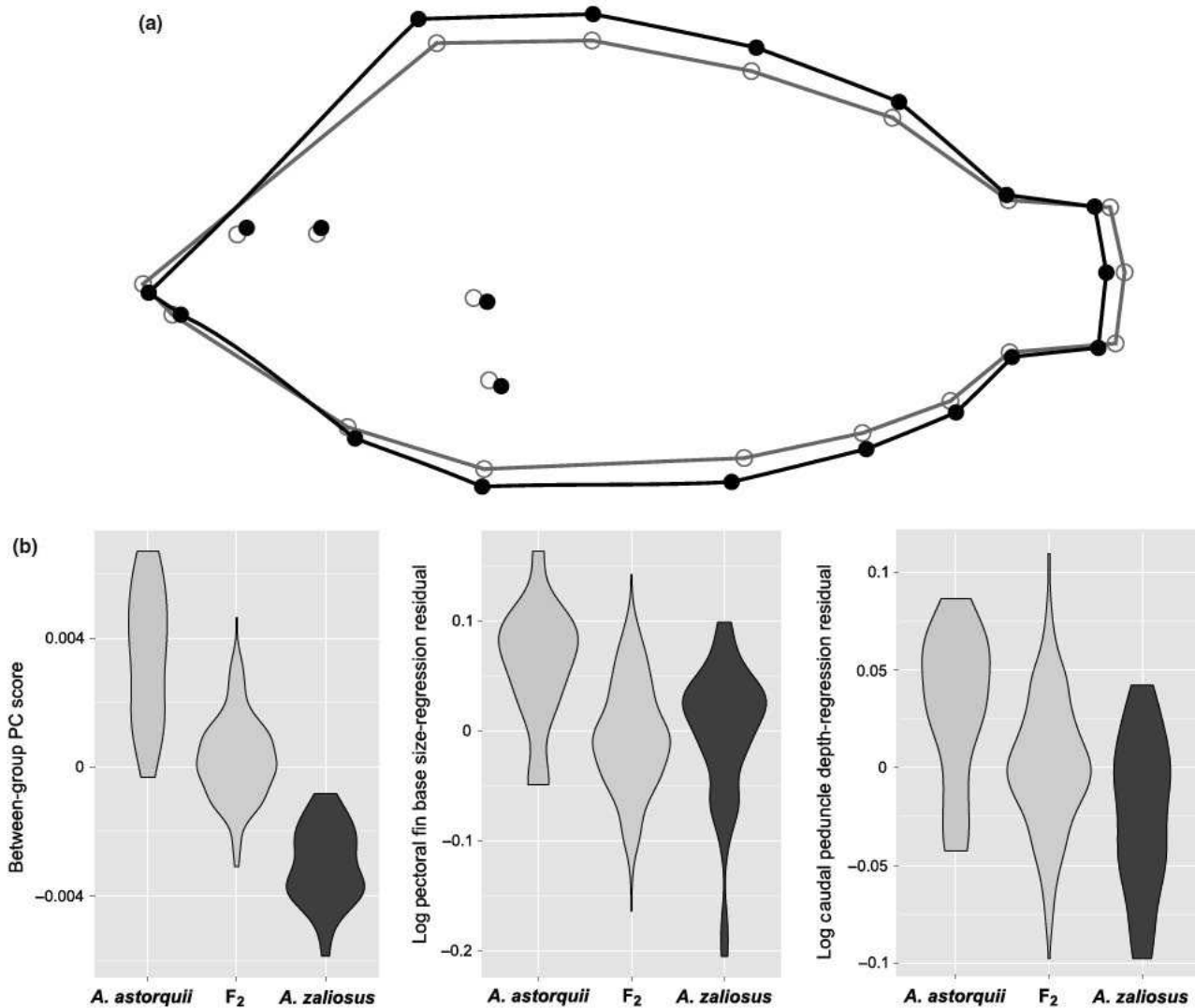


Fig. 4 (a) Shape variation in the F_2 individuals as predicted by the between species principal component used for QTL mapping. The grey dots represent the average shape of F_2 individuals, the black dots show the shape associated with a positive variation along the between species principal component (i.e. an *astorquii* like body shape). (b) Violin plots (Hintze & Nelson 1998) – a modified version of box plots (Tukey 1977) showing the distribution of the dependent variable plotted for the phenotypic traits used in QTL mapping in the parental species and in the F_2 generation.

markers were included in four categories according to their genotype frequency in the F_2 progeny (χ^2 test, $P < 0.001$): a) *noninformative*: markers that are not heterozygous in any of the F_1 s that generate an aa:ab:bb ratio of 1:0:0, 0:0:1 or 0:1:0 (300 of 1511, 19.6%); b) *unmappable*: markers where the parents share an allele (i.e. they are aa x ab or ab x aa), and one F_1 inherited the P_1 - (or P_2 -) exclusive allele, but the other did not, thus producing a test-cross (ab x aa) that generates a 1:1 ratio. (575 of 1511, 38.1%); c) *under severe segregation distortion*: which most likely reflects bad marker quality and technical artefacts (179 of 1511, 11.8%; 28.1% considering only the 636 markers left after the filtering described at points a and b above); d) *informative*

markers: remaining data set of loci after filtering out the a, b and c categories (457 markers). Among the informative markers, four remained as singletons or distorted the map and were therefore excluded. The final linkage map constructed with the remaining 453 markers genotyped in 220 F_2 individuals had a 14.4% average number of missing values. On average, 387 markers per F_2 individual were scored. Each marker was scored on average in 188 F_2 s.

Using more stringent thresholds for missing data and segregation distortion, we obtained genotype data sets of 301 (6.5% missing values) and 415 (14.5% missing values) markers, respectively, that were used for downstream linkage map and QTL analyses.

Linkage map

The three linkage maps resolved 25 LGs (Fig. S3, Supporting information). The total size of the map with 453 markers is 1 368.198 cM, representing an average marker spacing of approximately 3 cM. The size and number of LGs is comparable with current knowledge of the chromosome number of Neotropical cichlids ($n = 24$) (Feldberg *et al.* 2003) and previous studies on Midas cichlids that also resolved 25 LGs (Recknagel *et al.* 2013b).

QTL mapping

Across the three mapped major phenotypic traits – body shape, pectoral fin width and caudal peduncle width – a total of eleven QTL regions were found (Table 1). The two most prominent QTL peaks, significant at the genome-wide threshold ($P < 0.05$), were for body shape (LG3 LOD score: 3.90) and for pectoral fin width (LG6 LOD score: 4.06) (Table 1 and Fig. 5). The remaining nine QTL regions were statistically significant ($P < 0.05$) at the chromosome-wide level. Of these, three QTL associated with body shape were located on three different linkage groups (LG7 LOD score: 2.75; LG10 LOD score: 2.38; LG14 LOD score: 2.61), two QTL associated with pectoral fin width were located on two linkage groups (LG5 LOD score: 2.34; LG18 LOD score: 2.77), and four QTL associated with caudal peduncle depth (LG2 LOD score: 2.44; LG8 LOD score: 2.48; LG9 LOD score: 2.38; LG10 LOD score: 2.68). A total of 72 markers were linked to the QTL regions considering the 1.5 CI (confidence interval in which the LOD score is within 1.5 units of its maximum). Two QTL regions significant at the chromosome-wide level (one for body shape and one for caudal peduncle depth) showed a wide overlap on LG10. The

QTL found using the more stringent map in terms of missing values (301 markers) were located in the same regions as the ones found with the less stringent map (Fig. S3, Supporting information). Implementing the map obtained using a more stringent segregation distortion threshold, we found 10 of the 11 QTL found in the map with 453 loci. The QTL obtained with the less stringent map but absent in the more stringent one was located in a region of LG9 where there was only a marker that was excluded in the latter map. Analysing the ratio of body depth and standard length as a redundant measure of body shape recovered the same QTL regions found when analysing scores along the between-group principal component.

Gene prediction and functional annotation

The two narrow QTL regions significant at the genome-wide level were annotated based on the tilapia genome: body shape and pectoral fin size. Tilapia is the only cichlid species with an annotated genome, and, even though African and Neotropical cichlids shared a common ancestor 60–90 MYA (Farias *et al.* 1999), the overall levels of synteny between these two lineages are highly conserved (Recknagel *et al.* 2013b). For the body shape QTL, we annotated the genomic region spanning the locus significant at the genome-wide level and the upstream locus significant at the chromosome-wide level (loci included in the 1.5-LOD confidence interval: 4.156–5.056 cM). As no loci that were significant at the chromosome-wide level were linked to this QTL region downstream from the locus significant at the genome-wide level, we extended the annotation to a downstream region of the same length of the selected

Table 1 Summary of the QTL detected for the three traits analysed – external body shape morphology, pectoral fin base size and caudal pedunculum depth. For each QTL, nearest locus, position, LOD score, 1.5 CI (confidence interval in which the LOD score is within 1.5 units of its maximum) and explained variance are shown. For each trait, LOD significance threshold was calculated by 10 000 permutations. QTL exceeding the genome wide (GW) significance threshold are in bold. QTL and confidence interval position are given in centimorgan (cM)

Trait	Nearest locus	Linkage group	Position (cM)	LOD	1.5 CI (cM)	Variance explained (%)
Body shape	L 105438	3	4.609	3.90	4.156–5.056	7.9
Body shape	L 110835	7	34.528	2.75	17.452–38.528	5.2
Body shape	L 158912	10	52.893	2.38	30.751–54.017	4.5
Body shape	L 7781	14	0.000	2.61	0.000–11.191	4.9
Pectoral fin	L 31360	5	1.584	2.34	0.000–18.947	4.3
Pectoral fin	L 37242	6	55.594	4.06	53.098–56.594	8.5
Pectoral fin	L 5930	18	35.135	2.77	10.000–49.094	5.1
Caudal peduncle	L 47275	2	3.000	2.44	0.000–9.815	5.1
Caudal peduncle	L 132266	8	36.310	2.48	31.972–62.416	5.2
Caudal peduncle	L 48952	9	37.466	3.38	24.012–57.546	5.7
Caudal peduncle	L 53297	10	40.790	2.68	23.751–54.017	5.5

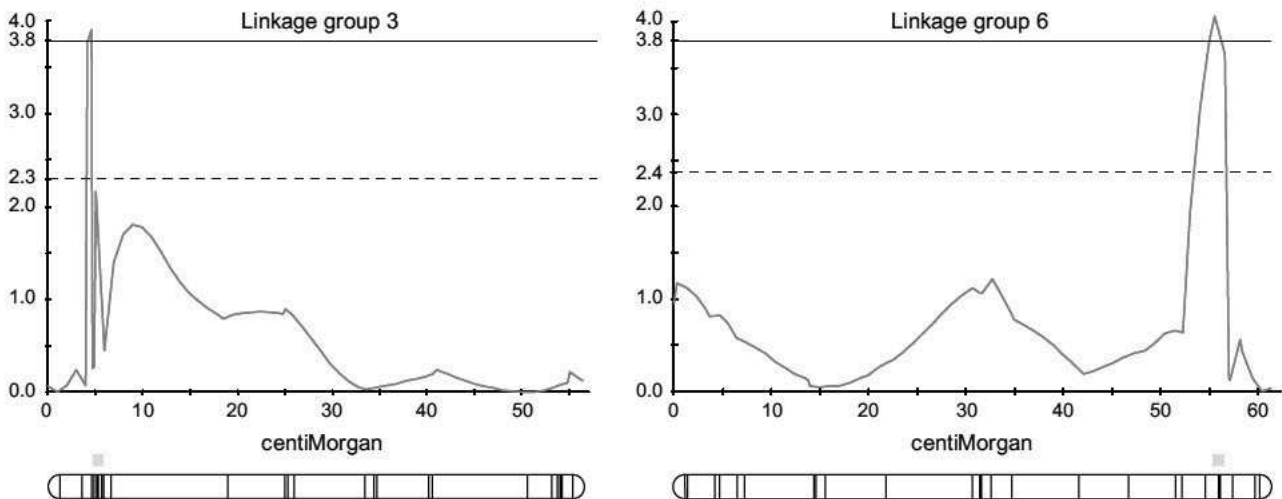


Fig. 5 Linkage groups (LG) showing a QTL region significant at the genome wide level for external body shape (LG3) and for pectoral fin base size (LG6). Red lines represent the LOD score, black lines indicate the genome wide LOD significant value, and dashed black lines indicate the chromosome wide significant values LOD. Green dots show the location of the QTL linked markers with a LOD value exceeding the 0.05 genome wide threshold.

upstream region. The sequence obtained was 3.249 Mbp in length and contains 20 genes (Table S2, Supporting information). For the pectoral fin, only one locus was linked to the QTL and it aligned in the middle of a short scaffold of the tilapia genome (scaffold length: 2.386 Mbp locus mapping position: 1.243 Mbp). Given the wide 1.5-LOD confidence interval (53.098–56.594 cM), the entire scaffold was annotated, a genomic region that includes 54 genes (Table S2, Supporting information). The genes included in the two highly supported QTL regions showed no Gene Ontology (GO) term overrepresentation neither when compared against the tilapia genome nor when compared with each other (Fig. S4, Supporting information), as revealed by the nonsignificance of all the enrichment analyses.

Discussion

The repeated parallel radiations of Midas cichlids in the Nicaraguan crater lakes are one of the few strongly substantiated cases of sympatric speciation (Barluenga *et al.* 2006; Coyne 2007). We aimed to identify genomic regions associated with adaptive differences in morphology along the limnetic benthic axis in the adaptive radiation of Nicaraguan Midas cichlids. In the Midas cichlid species pair studied here, *Amphilophus zaliosus* is an open water species (Barluenga *et al.* 2006), while *Amphilophus astorquii* is specialized on a benthic diet (Stauffer *et al.* 2008; Elmer *et al.* *in prep.*). By investigating F_2 individuals of this hybrid cross, we obtained 453 informative SNP markers and identified a single strongly supported QTL region for each of two ecologically relevant traits: body shape and pectoral fin base

size. Variation in external body morphology in general, and elongation in particular, is likely to play a key role in the early divergence in Midas cichlid species pairs, in this and other crater lakes (Kautt *et al.* 2012).

This type of limnetic benthic split seems to be characteristic of many recently evolved freshwater species, for example stickleback, whitefish and perch in postglacial lakes (see Rogers *et al.* 2013 and references therein). Divergence along the benthic limnetic axis has also occurred multiple times in species-rich Lake Malawi cichlid flocks, thus suggesting that it is of general importance for fish diversification in both temperate and tropical regions (Hulseley *et al.* 2013). By mapping the genomic regions associated with divergence in functionally important morphological traits, we extend the current knowledge of the genomic basis of divergence along the benthic limnetic axis to a Neotropical, sympatrically speciated species pair.

Marker selection and choice of morphological traits

The species targeted in this study are very young (Wilson *et al.* 2000; Barluenga & Meyer 2004, 2010; Barluenga *et al.* 2006; Elmer *et al.* 2010a). We found a high proportion (>2/3) of SNP alleles still shared between both species (even when comparing only two individuals). This is likely a consequence of the very recent (<23 890 years, based on crater lake age, Kutterolf *et al.* 2007) sympatric divergence of these Midas cichlid species (Barluenga & Meyer 2010; Kautt *et al.* 2012). It should be noted that the radiation of Lake Apoyo is, by far, the oldest crater lake radiation, and the species are the oldest and most divergent genetically. SNP sharing

and incomplete lineage sorting is even more prevalent among the younger species such as those forming the crater Lake Xiloá radiation (Kautt *et al.* unpublished data).

Our RADseq data set resulted in 453 informative and high-quality SNPs upon which our QTL analyses were based. To maximize the number of informative markers, we included markers fixed in both parents (*i.e.* aa × bb) and also markers where one of the parents was heterozygous and the parental origin of the alleles could be known (for example, if parent one was aa, parent two was ab, and both of the F_1 parents were ab). We limited the number of missing data in our analyses (the final genotype data set has <15% missing values), because large amounts of missing data might distort results in QTL analyses (Zhang *et al.* 2010). It has been shown that increasing levels of missing markers decreases the power of QTL detection and increases the false discovery rate. Missing loci have greater effects on smaller effect QTL and smaller size populations (Zhang *et al.* 2010) and, not knowing in advance the distribution of the QTL effect size, removing the loci with high proportion of missing data represents a conservative choice. After the rigorous selection of loci, the number of informative markers we detected (453) was slightly smaller than those used in our previous genetic mapping study of Midas cichlids (755) (Recknagel *et al.* 2013b). This discrepancy is mainly due to the fact that the previous study used both a higher number of F_2 progeny (347 individuals) and a less stringent locus selection strategy (about 25% missing values).

Methods for QTL analyses of multidimensional phenotypic data with a high number of loci and in the presence of missing data as is inherent to genotyping-by-sequencing remain to be developed. To map phenotypic differences in body shape, we therefore used a univariate measure that incorporates the multidimensional nature of fish body shape. We did this through a projection on a vector in multivariate shape space (*i.e.* the between-groups principal component). This choice has two advantages relative to other possible measures of body shape variation: 1) it has a clear biological interpretation (*i.e.* how much *zalius*-like or *astorquii*-like each individual is), and 2) it allows for a rigorous and easily interpretable visualization of the morphological variation because scores along the between-groups principal component can be visualized in terms of shape change. We performed separate analyses on two other putatively important functional traits, pectoral fin base size and caudal peduncle depth, because the between-groups principal component captured the main differences between parental species (*i.e.* body depth), but it was not able to capture the more limited variation in the other two traits. All the morphological traits

studied here differ between the two parental species and are therefore statistically correlated (for instance, both pectoral fin base size and caudal peduncle depth are higher in *A. astorquii*). In our QTL analyses, different but statistically correlated traits mapped to different genomic regions and therefore provide an internal validation of our results. In fact, the same genomic region may have otherwise been recovered if the correlation among traits was caused by pleiotropy or close physical linkage, or as a statistical artefact due to their correlation. Because the different traits map to different regions of the genome, this may indicate that divergent selection is acting on different regions of the genome.

Genetic basis of ecologically relevant traits

Previous studies have exclusively investigated morphological variation between wild-caught adult specimens (Elmer *et al.* 2010a) and were therefore unable to rule out the possibility that some of the observed differences were due to phenotypically plastic responses of the examined traits. The parental fish investigated in this study were collected in Lake Apoyo as small fry, were reared under identical laboratory conditions and still retained the expected morphological differences between parental species. This suggests that a strong genetic component underlies the putatively adaptive morphological traits targeted here despite the recent shared ancestry of these species (Barluenga & Meyer 2010; Kautt *et al.* 2012) and despite the fact that it is known that some aspects of the morphology of these fish (*e.g.* the pharyngeal jaws) are phenotypically plastic (Meyer 1990a,b; Muschick *et al.* 2011) as for other cichlids as well (*e.g.* Huyssseune 1995; Gunter *et al.* 2013). The presence of such a strong genetic component in external body morphology allows, in turn, to investigate which genomic regions are responsible for the observed phenotypic divergence between *A. astorquii* and *A. zalius*.

Here, we identified two distinct genomic regions associated with two divergent external morphological traits: body shape and pectoral fin base size. These traits, among others, are associated with hydrodynamic drag and thrust forces and have been suggested to be integral in swimming performance in fish (Drucker & Lauder 2003; Rouleau *et al.* 2010). The elongate body shape found in *A. zalius* is likely adaptive for movement in open water environments, as a more streamlined shape allows for a higher swimming performance, or a faster critical swimming speed and higher endurance level. Feeding in open water would most probably require good prolonged swimming ability in order to travel long distances to search for patchily distributed prey. By comparison, prey searching in the shallower more structurally complex benthic zone

probably requires short fast swimming bouts and more manoeuvring, and benthic individuals with a higher body are likely to exhibit a steeper power curve than open water fish (e.g. Pettersson & Brönmark 1999; Blake *et al.* 2005; Rouleau *et al.* 2010). A broader pectoral fin base is associated with a more powerful fin movement and is typically found in more manoeuvrable benthic fish that are associated with the rocky habitat of the littoral zone (Pettersson & Brönmark 1999; Blake *et al.* 2005; Rouleau *et al.* 2010). In salmonid fishes, for example, the pectoral fins are adducted against the body during constant-speed swimming – the type of swimming behaviour expected in open water environments while these fins are actively recruited for a variety of movement behaviours such as hovering, low-speed turning and rapid deceleration of the body during breaking (Drucker & Lauder 2003). A similar relationship has been found between pectoral fin size and habitat-specific foraging behaviour in a variety of other freshwater fish suggesting a strong functional relationship between pectoral fin size and mode of feeding and high potential for selection acting on this trait in the two different lacustrine habitats (reviewed in Drucker & Lauder 2003). Such traits are therefore likely to be under divergent or disruptive selection in different trophic niches along the benthic limnetic axis.

Candidate genes for body shape differences

In support of this role in adaptive divergence, the most highly supported QTL-linked region was associated with body shape. It spanned 3.249 Mb on a scaffold that included 20 annotated genes. This set of genes does not show any particular enrichment in specific GO terms. However, although at present we cannot estimate their potential role in the phenotypic divergence of the studied species pair, two genes may be of particular relevance to body shape. The first codes for the α -3 subunit of Collagen VI, a structural protein that plays a critical role in bone development and remodelling and muscle growth (Christensen *et al.* 2012 and references therein). The second gene codes for a glycerol kinase that plays a key role at the interface of glucose lipid metabolism and has been proposed as an important regulator in the metabolic pathways of food assimilation (Panserat *et al.* 2009).

The genomic region associated with pectoral fin base size was located on a different linkage group and spanned 2.386 Mb that included 54 annotated genes (Table S2, Supporting information). These genes have a wide set of functions, as shown by an absence of over-representation of specific GO categories. Of particular interest for further work might be two genes that code for proteins known to play a critical role in food intake

and processing, hydroxypyruvate isomerase and a cocaine- and amphetamine-regulated transcript (CART). Hydroxypyruvate isomerase has been suggested to be associated with the modulation of negative feedback of glucose on food intake (Mennigen 2011). Additionally, transcripts of the CART group have been shown to be involved in the response to fasting in goldfish (Volkoff & Peter 2001), cod (Kehoe & Volkoff 2007), catfish (Kobayashi *et al.* 2008) and Atlantic salmon (Murashita *et al.* 2009) and in an increase in motor activity in rat (Kimmel *et al.* 2000). In Cypriniformes, orexin A (the response to which is modulated by CART) does not only stimulate food consumption, but also enhances locomotor activity (Matsuda *et al.* 2011), and amphetamine has been shown to reduce response to food and increased swimming activity in the cichlid *Pelvicachromis pulcher* (Munro 1986).

In a third putatively adaptive morphological trait, the depth of the caudal peduncle, we found no association that was significant at the genome-wide level, although we did identify an association at the less stringent chromosome level of significance.

The size, location and gene composition of the genomic regions identified should at this stage be treated with some caution as they were annotated using the Nile tilapia genome, which being an African cichlid is up to *ca.* 90 MYA divergent from the Neotropical cichlids (Farias *et al.* 1999, 2001; Azuma *et al.* 2008). We are currently sequencing the complete genome of the Midas cichlid (Fan *et al.*, unpublished data), which will soon allow us to confirm the location of the QTL regions we identified here. However, in support of our inferences, previous synteny analyses found a high degree of conservation between tilapia and Midas cichlid (see Recknagel *et al.* 2013b).

How do RNAseq and QTL data compare?

Previous studies seeking genomic differences between benthic and limnetic Midas cichlids have generally also found few regions of divergence between sister taxa (Elmer *et al.* 2010b; Fan *et al.* 2012; Kautt *et al.* 2012). So far, however, we did not find common candidate genes between this study and genes previously found to show signs of positive selection (high K_a/K_s) in our earlier transcriptomic comparison of the same pair of species (Elmer *et al.* 2010b). There are at least five possible reasons why the different analyses do not detect the same genes: 1) the number of markers detected throughout the genome in this study is not large enough to capture all the genes correlated with the traits (we sequenced 118 636 loci shared between the parents while an *in silico* exploration of the draft Midas genome recovered 599 773 potential RAD tags); 2) in the RNAseq study 1

721 ESTs that is, a small portion of the coding regions of the Midas cichlid genome were tested for signatures of diversifying selection; 3) the phenotypic effect size of the genes detected by Elmer *et al.* (2010b) is small and difficult to detect using a QTL approach; 4) the tissues collected for the transcriptomic analyses (Elmer *et al.* 2010b) were collected at ontogenetic stages where genes relevant for the divergence in swimming ability between the two species might not be expressed; 5) rare transcripts might have been missed or miRNAs could be responsible for phenotypic differentiation at early stages during speciation (a possibility we are following up on through small RNA libraries and comparative genomics for miRNA differences among incipient species of cichlid fishes).

Genomics of limnetic and benthic body shapes in cichlids and other fish

In contrast to our findings in Midas cichlids, many different genes were found to be associated with body shape in recently diverged populations of sticklebacks (Albert *et al.* 2008). For example, QTL studies on body shape divergence between marine and freshwater populations of sticklebacks reported 26 chromosome regions (QTL) on 17 linkage groups, 36 QTL on 12 linkage groups and 41 QTL on 13 linkage groups (Albert *et al.* 2008; Rogers *et al.* 2012). A genomic scan of natural populations (Hohenlohe *et al.* 2010) has shown nine regions consistently differentiated between marine and freshwater sticklebacks and that one of these regions is adjacent to one of the QTL identified by Albert *et al.* (2008). A more recent genomic scan on multiple marine freshwater stickleback populations (Jones *et al.* 2012) identified F_{ST} outliers in 16 regions of 10 chromosomes. Body shape in salmon was also found to be associated with many small effect size QTL and few QTL of major effect, spread across 15 different linkage groups (Boulding *et al.* 2008). Like Midas cichlids, whitefish have benthic and limnetic ecomorphs, but in those cases, QTL identified were strongly associated with growth rate, behavioural differences and life history; QTL associated with body shape in a morphometric sense have not been identified so far. Alternatively, it may be the case that with the current QTL approach we were only detecting highly differentiated regions, while other small effect size regions in the genome could not be detected.

The field of evolutionary genomics is beginning to be able to investigate genome-wide patterns and mechanisms underlying the genetics of speciation. Increasingly, it is being shown that different organisms show different patterns of divergence across the genome at similar stages of speciation (Mayr 1942; Wu &

Ting 2004; Orr 2005; Elmer & Meyer 2011; Nosil & Schluter 2011; Nadeau *et al.* 2012; Nosil & Feder 2012; Andrew & Rieseberg 2013). Processes fundamentally related to those differences are gene flow, selection, recombination, divergence hitchhiking and genome hitchhiking (Feder *et al.* 2012; Nosil & Feder 2012). Here, we uncovered patterns of genome-wide divergence in a study system where sympatric speciation is in the relatively early stages (Meyer 1990a,b; Wilson *et al.* 2000; Barluenga *et al.* 2006; Elmer *et al.* 2010a).

Outlook: genomics of parallel evolution in Midas cichlids

The current study opens the door for future integrated studies on determining how selection is acting on natural populations of divergent cichlid morphs that will provide insights into the patterns and processes of genomic divergence at different points along the speciation continuum. In particular, this study presents one of the few accounts so far of the genomic basis of divergent external body morphology in a fish species. As the adaptive value of body shape difference is well understood, any further knowledge about the genomic architecture of these traits is of great interest for evolutionary biologists. By identifying specific genomic regions associated with variation in an ecologically relevant set of traits in this pair of species, we enable future population-level studies.

In future studies, we plan to investigate whether these regions play a major role in the adaptive sympatric radiation of cichlid fish in Lake Apoyo, and if these regions are associated with reproductive isolation between species pairs. A study on the genetic basis of divergence in trophic morphology (such as pharyngeal jaws) in this species pair is currently underway. This follow-up study will allow us to investigate the number of genomic regions responsible for divergence in this other important adaptive trait and if these are colocalized with the regions we identified here.

Lastly, given that morphologically parallel benthic and limnetic species are also found in crater Lake Xiloá (Elmer *et al.* 2010a), a QTL mapping project with the same design as the present is currently underway. Identifying the genomic regions associated with divergence in morphology along the benthic limnetic continuum in a similar species pair from Lake Xiloá using the same design will allow us to determine whether parallel variation in these adaptive traits arises through genetic variation at the same or different genomic regions. Because the Midas cichlid species flock in crater Lake Xiloá is younger (Elmer *et al.* 2010a, 2013), comparing QTL regions identified in fish inhabiting Lake Xiloá with those identified here will allow us to determine whether

the number of QTL regions are similar or not, whether they are homologous or whether there are changes at different time points along the speciation continuum. Eventually, we hope to identify the genes responsible for these parallel morphological solutions to similar ecological challenges and to determine whether the same, homologous, genes and even similar mutations might be responsible or not. At this point, it is hard to predict the answer, because evolution may not always use the same genetic or genomic route to reach the same phenotypic end.

Acknowledgements

We thank T. Lehtonen for collecting fry and help setting up initial breeding crosses, J. Sieling and her team for caring for the fish, L. Beck for laboratory assistance, S. Fan, T. Frickey and H. Recknagel for bionformatic support. We thank three anonymous referees for their helpful suggestions on the manuscript. MARENA granted permits for our field work in Nicaragua, and we thank them kindly for their continued support over the years. CF was funded by DAAD (B1 scholarship A/11/78461) and Marie Curie (IEF fellowship PlasticitySpeciation). KRE was funded by Alexander von Humboldt and NSERC postdoctoral fellowships. MLS was funded by the International Max Planck Research School (IMPRS). JCJ was funded by the Zukunftscolleg at the University of Konstanz and the DFG. The University of Konstanz is thanked for its support of the Meyer laboratory and the GeCKo (Genomic Center Konstanz). Funding for this project came from the European Research Council through ERC advanced grant 293700 GenAdap (AM).

References

- Aguirre WE, Bell MA (2012) Twenty years of body shape evolution in a threespine stickleback population adapting to a lake environment. *Biological Journal of the Linnean Society*, **105**, 817–831.
- Albert AYL, Sawaya S, Vines TH *et al.* (2008) The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. *Evolution*, **62**, 76–85.
- Andrew RL, Rieseberg LH (2013) Divergence is focused on few genomic regions early in speciation: incipient speciation of sunflower ecotypes. *Evolution*, **67**, 2468–2482.
- Azuma Y, Kumazawa Y, Miya M, Mabuchi K, Nishida M (2008) Mitogenomic evaluation of the historical biogeography of cichlids toward reliable dating of teleostean divergences. *BMC Evolutionary Biology*, **8**, 215.
- Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.
- Barlow GW (1976) The Midas cichlid in Nicaragua. In: *Investigations of the Ichthyofauna of Nicaraguan Lakes* (ed. Thorson TB), pp. 333–358. University of Nebraska, 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.
- 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, Stolting KN, Salzburger W, Muschick M, Meyer A (2006) Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature*, **439**, 719–723.
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate under dependency. *Annals of Statistics*, **29**, 1165–1188.
- Bernatchez L, Chouinard A, Lu G (1999) Integrating molecular genetics and ecology in studies of adaptive radiation: white fish, *Coregonus*, as a case study. *Biological Journal of the Linnean Society*, **68**, 173–194.
- Bernatchez L, Renaut S, Whiteley AR *et al.* (2010) On the origin of species: insights from the ecological genomics of lake whitefish. *Philosophical Transactions of the Royal Society B*, **365**, 1783–1800.
- Bertrand M, Marcogliese DJ, Magnan P (2008) Trophic polymorphism in brook charr revealed by diet, parasites and morphometrics. *Journal of Fish Biology*, **72**, 555–572.
- Blake RW, Law TC, Chan KHS, Li JFZ (2005) Comparison of the prolonged swimming performances of closely related, morphologically distinct three spined sticklebacks. *Journal of Fish Biology*, **67**, 834–848.
- Bookstein FL (1997) Landmark methods for forms without landmarks: localizing group differences in outline shape. *Medical Image Analysis*, **1**, 225–243.
- Boulding EG, Culling M, Glebe B, Berg PR, Lien S, Moen T (2008) Conservation genomics of Atlantic salmon: SNPs associated with QTLs for adaptive traits in parr from four trans Atlantic backcrosses. *Heredity*, **101**, 381–391.
- Boulesteix AL (2005) A note on between group PCA. *International Journal of Pure and Applied Mathematics*, **19**, 359–366.
- Carrier D, Chase K, Lark KG (2005) Genetics of canid skeletal variation: size and shape of the pelvis. *Genome Research*, **15**, 1825–1830.
- Catchen JM, Amore A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short read sequences. *G3*, **3**(1), 171–182.
- Christensen SE, Coles JM, Zelenski NA *et al.* (2012) Altered trabecular bone structure and delayed cartilage degeneration in the knees of Collagen VI null mice. *PLoS ONE*, **7**, e33397.
- Chutimanitsakun Y, Nipper RW, Cuesta Marcos A *et al.* (2011) Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley. *BMC Genomics*, **12**, 4.
- Conesa A, Gotz S, Garcia Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674–3676.
- Coyne JA (2007) Sympatric speciation. *Current Biology*, **17**, 787–788.
- Drucker EG, Lauder GV (2003) Function of pectoral fins in rainbow trout: behavioral repertoire and hydrodynamic forces. *Journal of Experimental Biology*, **206**, 813–826.

- 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 (2010a) 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 B*, **365**, 1763–1782.
- Elmer KR, Fan S, Gunter HM *et al.* (2010b) Rapid evolution and selection inferred from the transcriptomes of sympatric crater lake cichlid fishes. *Molecular Ecology*, **19**, 197–211.
- Elmer KR, Lehtonen TK, Kautt AF, Harrod C, Meyer A (2010c) Rapid sympatric ecological differentiation of crater lake cichlid fishes within historic times. *BMC Biology*, **8**, 60.
- Elmer KR, Lehtonen TK, Fan S, Meyer A (2013) Crater lake colonization by Neotropical cichlid fishes. *Evolution*, **67**, 281–288.
- Fan S, Elmer KR, Meyer A (2012) Genomics of adaptation and speciation in cichlid fishes: recent advances and analyses in African and Neotropical lineages. *Philosophical Transactions of the Royal Society B*, **367**, 385–394.
- Farias IP, Ortí G, Sampaio I, Schneider H, Meyer A (1999) Mitochondrial DNA phylogeny of the family Cichlidae: monophyly and fast molecular evolution of the Neotropical assemblage. *Journal of Molecular Evolution*, **48**, 703–711.
- Farias IP, Ortí G, Sampaio I, Schneider H, Meyer A (2001) The cytochrome b gene as a phylogenetic marker: the limits of resolution for analyzing relationships among cichlid fishes. *Journal of Molecular Evolution*, **53**, 89–103.
- Feder JL, Egan SP, Nosil P (2012) The genomics of speciation with gene flow. *Trends in Genetics*, **28**, 342–350.
- Feldberg E, Porto JIR, Bertollo LAC (2003) Chromosomal changes and adaptation of cichlid fishes during evolution. In: *Fish Adaptations* (eds Val AL & Kapoor BG), pp. 285–308. Science Publisher, New Delhi and New York.
- Fenderson OC (1964) Evidence of subpopulations of lake whitefish, *Coregonus clupeaformis*, involving a dwarfed form. *Transactions of the American Fisheries Society*, **93**, 77–94.
- Fisher RA (1934) *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.
- Fruciano C, Tigano C, Ferrito V (2011) Geographic morphological variation within and between colour phases in *Coris julis* (L. 1758), a protogynous marine fish. *Biological Journal of the Linnean Society*, **104**, 148–162.
- Fruciano C, Tigano C, Ferrito V (2012) Body shape variation and colour change during growth in a protogynous fish. *Environmental Biology of Fishes*, **94**, 615–622.
- Gagnaire PA, Pavey SA, Normandeau E, Bernatchez L (2013) The genomic architecture of reproductive isolation across the speciation continuum in lake whitefish species pairs assessed by RAD sequencing. *Evolution*, **67**, 2483–2497.
- Gavrilets S (2003) Perspective: models of speciation: what have we learned in 40 years? *Evolution*, **57**, 2197–2215.
- 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.
- Gowell C, Quinn T, Taylor EB (2012) Coexistence and origin of trophic ecotypes of pygmy whitefish, *Prosopium coulterii*, in southwestern Alaskan lake. *Journal of Evolutionary Biology*, **25**, 2432–2448.
- Gunter H, Xiong F, Fan S, Franchini P, Fruciano C, Meyer A (2013) Shaping development through mechanical strain: the transcriptional basis of diet dependent phenotypic plasticity in *Astatoreochromis alluaudi*. *Molecular Ecology*, **22**, 4516–4531.
- Hemmer Hansen J, Nielsen EE, Therkildsen NO *et al.* (2013) A genomic island linked to ecotype divergence in Atlantic cod. *Molecular Ecology*, **22**, 2653–2667.
- Hendry AP, Peichel CL, Matthews B, Boughman JW, Nosil P (2013) Stickleback research: the now and the next. *Evolutionary Ecology Research*, **15**, 111–141.
- Henning F, Renz AJ, Fukamachi S, Meyer A (2010) Genetic, comparative genomic, and expression analyses of the Mc1r locus in the polychromatic Midas cichlid fish (Teleostei, Cichlidae *Amphilophus* sp.) species group. *Journal of Molecular Evolution*, **70**, 405–412.
- Henning F, Jones JC, Franchini P, Meyer A (2013) Transcriptomics of morphological color change in polychromatic Midas cichlids. *BMC Genomics*, **14**, 171.
- Hintze JL, Nelson RD (1998) Violin plots: a box plot density trace synergism. *The American Statistician*, **52**, 181–184.
- Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD Tags. *PLOS Genetics*, **6**, e1000862.
- Hulseley CD, Roberts RJ, Loh YHE, Rupp MF, Streebman JT (2013) Lake Malawi cichlid evolution along a benthic/limnetic axis. *Ecology and Evolution*, **3**, 2262–2272.
- Huysseune A (1995) Phenotypic plasticity in the lower pharyngeal jaw dentition of *Astatoreochromis alluaudi* (Teleostei, Cichlidae). *Archives of Oral Biology*, **40**, 1005–1014.
- Jones FC, Chan YF, Schmutz J *et al.* (2012) A genome wide SNP genotyping array reveals patterns of global and repeated species pair divergence in sticklebacks. *Current Biology*, **22**, 83–90.
- Jones JC, Fan S, Franchini P, Scharl M, Meyer A (2013) The evolutionary history of *Xiphophorus* fish and their sexually selected sword: a genome wide approach using restriction site associated DNA sequencing. *Molecular Ecology*, **22**, 2986–3001.
- Kaeuffer R, Peichel CL, Bolnick DI, Hendry AP (2012) Parallel and nonparallel aspects of ecological, phenotypic and genetic divergence across replicate population pairs of lake and stream stickleback. *Evolution*, **66**, 402–418.
- Kautt AF, Elmer KR, Meyer A (2012) Genomic signatures of divergent selection and speciation patterns in a 'natural experiment', the young parallel radiations of Nicaraguan crater lake cichlid fishes. *Molecular Ecology*, **21**, 4770–4786.
- Kehoe AS, Volkoff H (2007) Cloning and characterization of neuropeptide Y (NPY) and cocaine and amphetamine regulated transcript (CART) in Atlantic cod (*Gadus morhua*). *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, **146**, 451–461.

- Kekalainen J, Kahkonen J, Kiviniemi V, Huuskonen H (2010) Morphological variation of perch *Perca fluviatilis* in humic lakes: the effect of predator density, competition and prey abundance. *Journal of Fish Biology*, **76**, 787–799.
- Kimmel HL, Gong W, Dall'Vechia S, Hunter RG, Kuhar MJ (2000) Intra ventral tegmental area injection of rat Cocaine and Amphetamine Regulated Transcript Peptide 55 102 induces locomotor activity and promotes conditioned place preference. *Journal of Pharmacology and Experimental Therapeutics*, **294**, 784–792.
- Klingenberg CP (2011) MorphoJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources*, **11**, 353–357.
- Klingenberg CP, Barluenga M, Meyer A (2002) Shape analysis of symmetric structures: quantifying variation among individuals and asymmetry. *Evolution*, **56**, 1909–1920.
- Klingenberg C, 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.
- Kobayashi Y, Peterson BC, Waldbieser GC (2008) Association of cocaine and amphetamine regulated transcript (CART) messenger RNA level, food intake, and growth in channel catfish. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, **151**, 219–225.
- Kocher TD (2004) Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews Genetics*, **5**, 288–298.
- Kuraku S, Meyer A (2008) Genomic analysis of cichlid fish 'natural mutants'. *Current Opinion in Genetics & Development*, **18**, 551–558.
- Kutterolf K, Freundt A, Perez W, Wehrmann H, Schmincke U (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.
- Landry L, Vincent WF, Bernatchez L (2007) Parallel evolution of lake white fish dwarf ecotypes in association with limnological features of their adaptive landscape. *Journal of Evolutionary Biology*, **20**, 971–984.
- Langerhans RB, Gifford ME, Joseph EO (2007) Ecological speciation in *Gambusia* fishes. *Evolution*, **61**, 2056–2074.
- Larson GA (1976) Social behaviour and feeding ability of two phenotypes of *Gasterosteus aculeatus* in relation to their spatial and trophic segregation in a temperate lake. *Canadian Journal of Zoology*, **54**, 107–121.
- Malek T, Boughman JW, Dworkin ID, Peichel CL (2012) Admixture mapping of male nuptial color in a recently formed hybrid population of threespine sticklebacks. *Molecular Ecology*, **21**, 5265–5279.
- Manousaki T, Hull PM, Kusche H *et al.* (2013) Parsing parallel evolution: ecological divergence and differential gene expression in the adaptive radiations of thick lipped Midas cichlid fishes from Nicaragua. *Molecular Ecology*, **22**, 650–669.
- Matsuda K, Kang KS, Sakashita A, Yahashi S, Vaudry H (2011) Behavioral effect of neuropeptides related to feeding regulation in fish. *Annals of the New York Academy of Sciences*, **1220**, 117–126.
- Mayr E (1942) *Systematics and the Origin of Species*. Columbia University Press, New York.
- McCart P (1970) Evidence for the existence of sibling species of pygmy whitefish (*Prosopium coulteri*) in three Alaskan lakes. In: *Biology of Coregonid Fishes* (eds Lindsey CC & Woods CS), pp. 81–98. University of Manitoba Press, Winnipeg.
- McPhail JD (1984) Ecology and evolution of sympatric sticklebacks (*Gasterosteus*): morphological and genetic evidence for a species pair in Enos Lake, British Columbia. *Canadian Journal of Zoology*, **62**, 1402–1408.
- Meek SE (1907) Synopsis of the fishes of the Great Lakes of Nicaragua. *Field Columbian Museum Publication 121*, Zoological Series, **7**, 97–132.
- Mennigen JA (2011) *The serotonergic system as a target of neuroendocrine disruption for the pharmaceutical fluoxetine in the brain of goldfish (Carassius auratus)*. PhD thesis. University of Ottawa, Ottawa.
- Meyer A (1989) Costs and benefits of morphological specialization: feeding performance in the trophically polymorphic Neotropical cichlid fish, *Cichlasoma citrinellum*. *Oecologia*, **80**, 431–436.
- Meyer A (1990a) Ecological and evolutionary aspects of the trophic polymorphism in *Cichlasoma citrinellum* (Pisces: Cichlidae). *Biological Journal of the Linnean Society*, **39**, 279–299.
- Meyer A (1990b) Morphometrics and allometry of 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 cichlids. *Trends in Ecology & Evolution*, **8**, 279–284.
- Meyer A, Kocher TD, Basasibwaki P, Wilson AC (1990) Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature*, **347**, 550–553.
- Mezey JG, Houle D, Nuzhdin AV (2005) Naturally segregating quantitative trait loci affecting wing shape of *Drosophila melanogaster*. *Genetics*, **169**, 2101–2113.
- Michel AP, Sim S, Powell THQ, Taylor MS, Nosil P, Feder JL (2010) Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences*, **107**, 9724–9729.
- Mitteroecker P, Bookstein FL (2011) Linear discrimination, ordination, and the visualization of selection gradients in modern morphometrics. *Evolutionary Biology*, **38**, 100–114.
- Munro AD (1986) The effects of apomorphine, d amphetamine and chlorpromazine on the aggressiveness of isolated *Aequidens pulcher* (Teleostei, Cichlidae). *Psychopharmacology*, **88**, 124–128.
- Murashita K, Kurokawa T, Ebbesson LOE, Stefansson SO, Rönnestad I (2009) Characterization, tissue distribution, and regulation of agouti related protein (AgRP), cocaine and amphetamine regulated transcript (CART) and neuropeptide Y (NPY) in Atlantic salmon (*Salmo salar*). *General and Comparative Endocrinology*, **162**, 160–171.
- Muschick M, Barluenga M, Salzburger W, Meyer A (2011) Adaptive phenotypic plasticity in the Midas cichlid fish pharyngeal jaw and its relevance in adaptive radiation. *BMC Evolutionary Biology*, **11**, 116.
- Nadeau NJ, Whibley A, Jones RT *et al.* (2012) Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by late scale targeted sequencing. *Philosophical Transactions of the Royal Society B*, **367**, 354–363.
- Noor MAF, Feder JL (2006) Speciation genetics: evolving approaches. *Nature Reviews Genetics*, **7**, 851–861.

- Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B*, **367**, 332–342.
- Nosil P, Schluter D (2011) The genes underlying the process of speciation. *Trends in Ecology & Evolution*, **26**, 160–167.
- Orr HA (1999) The evolutionary genetics of adaptation: a simulation study. *Genetics Research*, **74**, 207–214.
- Orr HA (2005) The Genetic basis of reproductive isolation: insights from *Drosophila*. *Proceedings of the National Academy of Sciences*, **102**, 6522–6526.
- Panserat S, Hortopan GA, Plagnes Juan E *et al.* (2009) Differential gene expression after total replacement of dietary fish meal and fish oil by plant products in rainbow trout (*Oncorhynchus mykiss*) liver. *Aquaculture*, **294**, 123–131.
- Peichel CL, Nereng KS, Ohgi KA *et al.* (2001) The genetic architecture of divergence between threespine stickleback species. *Nature*, **414**, 901–905.
- Perez IS, Bernal V, Gonzalez PN (2006) Differences between sliding semilandmark methods in geometric morphometrics, with an application to human craniofacial and dental variation. *Journal of Anatomy*, **208**, 769–784.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non model species. *PLoS ONE*, **7**, e37135.
- Pettersson LB, Bronmark C (1999) Energetic consequences of an inducible morphological defence in crucian carp. *Oecologia*, **121**, 12–18.
- Recknagel H, Kusche H, Elmer KR, Meyer A (2013a) Two new endemic species in the Midas cichlid species complex from Nicaraguan crater lakes: *Amphilophus tolteca* and *Amphilophus viridis* (Perciformes, Cichlidae). *Aqua, International Journal of Ichthyology*, **19**, 207–224.
- Recknagel H, Elmer KR, Meyer A (2013b) A hybrid genetic linkage map of two ecologically and morphologically divergent Midas cichlid fishes (*Amphilophus* spp.) obtained by massively parallel DNA sequencing (ddRADSeq). *G3 Genes Genomes Genetics*, **3**, 65–74.
- 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 B*, **367**, 354–363.
- Rogers SM, Bernatchez L (2005) Integrating QTL mapping and genomic scans towards the characterization of candidate loci under parallel directional 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.
- Rogers SM, Tamkee P, Summers B *et al.* (2012) Genetic signature of adaptive peak shift in threespine stickleback. *Evolution*, **66**, 2439–2450.
- Rogers SM, Mee JA, Bowles E (2013) The consequences of genomic architecture on ecological speciation in postglacial fishes. *Current Zoology*, **59**, 53–71.
- Rohlf FJ (2006) *TpsDig, Digitize Landmarks and Outlines, Version 2.10*. Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, New York.
- Rohlf FJ (2007a) *TpsRelw, Relative Warps Analysis, Version 1.45*. Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, New York.
- Rohlf FJ (2007b) *NTSYSpc: Numerical Taxonomy System, Version 2.20*. Exeter Publishing Ltd, Setauket, New York.
- Rouleau S, Glémet H, Magnan P (2010) Effects of morphology on swimming performance in wild and laboratory crosses of brook trout ecotypes. *Functional Ecology*, **24**, 310–321.
- Roy D, Docker MF, Haffner GD, Heath DD (2007) Body shape vs. colour associated initial divergence in the Telmatherina radiation in Lake Matano, Sulawesi, Indonesia. *Journal of Evolutionary Biology*, **20**, 1126–1137.
- Schluter D (2000) *The Ecology of Adaptive Radiation*. OUP, Oxford.
- Seehausen O (2006) African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Biological Society B*, **273**, 1987–1998.
- Siwertsson A, Knudsen R, Adams CE, Præbel K, Amundsen PA (2013) Parallel and non parallel morphological divergence among foraging specialists in European whitefish (*Coregonus lavaretus*). *Ecology and Evolution*, **3**, 1590–1602.
- Smadja CM, Butlin RK (2011) A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology*, **20**, 5123–5140.
- Stauffer JR Jr, 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.
- Svanback R, Eklov P (2002) Effects of habitat and food resources on morphology and ontogenetic growth trajectories in perch. *Oecologia*, **131**, 61–70.
- Tukey JW (1977) *Exploratory data analysis*. Addison Wesley, Reading, Massachusetts.
- Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic Islands of Speciation in *Anopheles gambiae*. *PLoS Biology*, **3**, e285.
- Valentin AE, Penin X, Chanut JP, Sévigny JM, Rohlf FJ (2008) Arching effect on fish body shape in geometric morphometric studies. *Journal of Fish Biology*, **73**, 623–638.
- Van Ooijen JW (2004) *MapQTL 5. Software for the Mapping of Quantitative Trait Loci in Experimental Populations*. Kyazma BV, Wageningen, The Netherlands.
- Van Ooijen JW (2006) *JoinMap 4. Software for the Calculation of Genetic Linkage Maps in Experimental Populations*. Kyazma BV, Wageningen, The Netherlands.
- Via S (2012) Divergence hitchhiking and the dynamics of ecological speciation with gene flow. *Philosophical Transactions of the Royal Society B*, **367**, 451–460.
- Via S, Conte G, Mason Foley C, Mills K (2012) Localizing FST outliers on a QTL map reveals evidence for large genomic regions of reduced gene exchange during speciation with gene flow. *Molecular Ecology*, **21**, 5546–5560.
- Villa J (1976) Systematic status of the cichlid fishes *Cichlasoma dorsatum*, *C. granadense* and *C. nigritum* Meek. In: *Investigations of the Ichthyofauna of Nicaraguan Lakes* (ed. Thorson TB), pp. 375–383. University of Nebraska, Lincoln, Nebraska.
- Volanthen P, Roy D, Hudson AG, Largiadèr CR, Bittners D, Seehausen O (2009) Divergence along a steep ecological gradient in lake whitefish (*Coregonus* sp.). *Journal of Evolutionary Biology*, **22**, 498–514.
- Volkoff H, Peter RE (2001) Characterization of two forms of cocaine and amphetamine regulated transcript (CART) peptide precursors in goldfish: molecular cloning and distribution, modulation of expression by nutritional status, and interactions with leptin. *Endocrinology*, **142**, 5076–5088.

- Walker JA (1997) Ecological morphology of lacustrine three spine stickleback *Gasterosteus aculeatus* L. (Gasterosteidae) body shape. *Biological Journal of the Linnean Society*, **61**, 3–50.
- Walker JA, Bell MA (2000) Net evolutionary trajectories of body shape evolution within a microgeographic radiation of threespine sticklebacks (*Gasterosteus aculeatus*). *Journal of Zoology*, **252**, 293–302.
- Ward JA, Bhangoo J, Fernandez Fernandez F *et al.* (2013) Saturated linkage map construction in *Rubusidaeus* using genotyping by sequencing and genome independent imputation. *BMC Genomics*, **14**, 2.
- Willacker JJ, von Hippel FA, Wilton PR, Walton KM (2010) Classification of threespine stickleback along the benthic limnetic axis. *Biological Journal of the Linnean Society*, **101**, 595–608.
- Wilson AB, Noack Kunmann K, Meyer A (2000) Incipient speciation in sympatric Nicaraguan crater lake cichlid fishes: sexual selection versus ecological diversification. *Proceedings of the Biological Society B*, **267**, 2133–2141.
- Wootton RJ (2009) The Darwinian stickleback *Gasterosteus aculeatus*: a history of evolutionary studies. *Journal of Fish Biology*, **75**, 1919–1942.
- Wu CI, Ting CT (2004) Genes and speciation. *Nature Reviews Genetics*, **5**, 114–122.
- Zhang L, Wang S, Li H *et al.* (2010) Effects of missing marker and segregation distortion on QTL mapping in F2 populations. *Theoretical and Applied Genetics*, **121**, 1071–1082.

P.F. performed the laboratory work, conducted the molecular data analyses and QTL mapping. C.F. carried out fish caring, performed the morphometric analyses and QTL mapping. M.L.S. carried out fish breeding and caring, and collected morphometric data. J.C.J. conducted part of the laboratory work. F.H. built the linkage map. A.M. and K.R.E. conceived the study and carried out fish breeding. All authors contributed to writing the manuscript.

Data accessibility

Raw Illumina sequences were deposited into the NCBI's Sequence Read Archive (SRA) database with Accession number: SRA110085. Raw landmark coordinates and the final genotypes of the data set used for QTL analysis are archived in DRYAD under doi: 10.5061/dryad.2h1c4.

Fig. S1 Configuration of points and linear distances used in the present study. Circles represent landmarks, triangles semi landmarks.

Fig. S2 Histogram showing the distribution of missing data in the markers used for the construction of the less stringent linkage map (453 markers).

Fig. S3 Hybrid linkage maps constructed with 453 (a.), 301 (using a more stringent missing values threshold) (b.) and 415 (using a more stringent segregation distortion threshold) (c.) informative SNPs.

Fig. S4 Frequency of Gene Ontology (GO) categories for genes in the QTL region associated with body shape (left) and pectoral fin base size (right).

Table S1 Sequencing statistics of each individual.

Table S2 Sequence description, number of Gene Ontology (GO) terms and GO terms description for each of the QTL regions significant at the genome wide threshold.