MOLECULAR ECOLOGY

Molecular Ecology (2012) doi: 10.1111/j.1365-294X.2012.05540.x

Divergent natural selection with gene flow along major environmental gradients in Amazonia: insights from genome scans, population genetics and phylogeography of the characin fish *Triportheus albus*

GEORGINA M. COOKE,*§ NING L. CHAO† and LUCIANO B. BEHEREGARAY*‡

*Molecular Ecology Laboratory, Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia, †Departamento de Ciências Pesqueiras, Universidade Federal do Amazonas, Manaus, Brazil, ‡Molecular Ecology Lab, School of Biological Sciences, Flinders University, Adelaide, SA 5001, Australia

Abstract

The unparalleled diversity of tropical ecosystems like the Amazon Basin has been traditionally explained using spatial models within the context of climatic and geological history. Yet, it is adaptive genetic diversity that defines how species evolve and interact within an ecosystem. Here, we combine genome scans, population genetics and sequence-based phylogeographic analyses to examine spatial and ecological arrangements of selected and neutrally evolving regions of the genome of an Amazonian fish, *Triportheus albus*. Using a sampling design encompassing five major Amazonian rivers, three hydrochemical settings, 352 nuclear markers and two mitochondrial DNA genes, we assess the influence of environmental gradients as biodiversity drivers in Amazonia. We identify strong divergent natural selection with gene flow and isolation by environment across craton (black and clear colour)- and Andean (white colour)-derived water types. Furthermore, we find that heightened selection and population genetic structure present at the interface of these water types appears more powerful in generating diversity than the spatial arrangement of river systems and vicariant biogeographic history. The results from our study challenge assumptions about the origin and distribution of adaptive and neutral genetic diversity in tropical ecosystems. In addition, they have important implications for measures of biodiversity and evolutionary potential in one of the world's most diverse and iconic ecosystems.

Keywords: Adaptive divergence, ecological genomics, isolation by selection gradient, ecological speciation, AFLP

Received 26 February 2011; revision received 22 January 2012; accepted 1 February 2012

Introduction

Knowledge about the distribution of adaptive genetic diversity is key to understand how organisms evolve and interact within an ecosystem. Adaptive diversity generated by natural selection can occur via two routes. The first, mutation-order speciation, occurs when populations diverge after accumulating a different assortment of mutations under equivalent selection pressures (Mani & Clarke 1990; Nosil & Flaxman 2011). The second, ecological speciation, results from the evolution of reproductive isolation between populations or subsets of a single population as a direct consequence of ecologically based divergent natural selection (Schluter 2000, 2001; Rundle & Nosil 2005; Schluter & Conte 2009). While natural selection drives alleles to fixation in both scenarios, selection favours divergence between environments only in ecological speciation (Schluter & Conte 2009). Thus, ecological speciation results in reproductive isolation

Correspondence: Luciano Beheregaray, Fax: 61 (8) 82013015; E-mail: luciano.beheregaray@flinders.edu.au

§Present address: School of Biological, Earth and Environmental Sciences, University of New South Wales, Kensington, NSW 2052, Australia.

between environments, where 'environments' refer to abiotic or biotic elements of a habitat, including interspecific interactions (Schluter 2001).

In the past 10 years, there has been sizeable progress in the understanding of divergent natural selection and how it results in ecological speciation. Numerous studies have demonstrated how ecological divergence and specific phenotypic traits under selection alter reproductive success (Via 1999; Rundle *et al.* 2000; Rundle & Nosil 2005). However, the genetics of ecologically based divergent natural selection remains comparatively mysterious (reviewed in Schluter & Conte 2009). Currently, it is being examined by employing approaches such as the mapping of quantitative trait loci or the identification of genes or regulatory control regions affecting individual phenotype and reproductive success (e.g. Orr & Turelli 2001; Bradshaw & Schemske 2003; Mackay *et al.* 2009; Bernatchez *et al.* 2010). These methods are powerful but require a priori genomic knowledge that is often unavailable in non-model organisms. Another approach that is growing in popularity and is accessible to studies of non-model organisms involves 'genome scans'. A genome scan screens many individuals from ecologically different populations or species at numerous markers and compares allelic variation at marker loci spread throughout the genome. Here, signatures of selection can be detected in 'outlier loci' that exhibit exceptionally high levels of differentiation (Luikart *et al.* 2003; Beaumont 2005). Although genome scans cannot identify the precise genetic condition that drives phenotypic traits, they can estimate selected genomic regions and compare the evolutionary patterns between selected and neutral loci (e.g. Campbell & Bernatchez 2004; Emelianov *et al.* 2004; Scotti-Saintagne *et al.* 2004; Achere *et al.* 2005; Turner *et al.* 2005; Bonin *et al.* 2006; Murray & Hare 2006; Nosil *et al.* 2008). A potential caveat of a genome scan is that genes that have diverged under mutation-order processes can also leave the same signature of outlier loci (Barton & Bengtsson 1986; Kondrashov 2003). Nonetheless, the key distinction between ecological and mutation-order divergence lies in the association of outlier loci to contrasting environments.

Genome scans are effective in identifying divergent natural selection and ecological speciation primarily because of the heterogeneous nature of genomic divergence through time (reviewed in Nosil *et al.* 2009). Because of this, the temporal context of divergence is critical when examining the outlier loci within the paradigm of ecological speciation. During the early phases of ecologically based divergent natural selection, selection will be highest at genomic regions that affect or are tightly linked via divergence hitchhiking to ecologically important traits. Simultaneously, the homogenizing effects of gene flow may continue throughout the rest of the genome (Wu 2001; Gavrilets & Vose 2005; Via 2009). However, as time passes, if speciation progresses, genetic divergence will occur in other genomic regions that were not originally affected by divergent selection via drift or new selection pressures, resulting in reproductive isolation and eventual 'ecological speciation' between lineages (Mayr 1963; Funk 1998; Schluter 2000; Rundle & Nosil 2005; Via 2009). This process has been described as isolation by adaptation (IBA) and results in a positive correlation between the degree of adaptive phenotypic divergence and neutral differentiation with time (Nosil *et al.* 2008). In this way, ecologically based adaptive divergence also promotes the accumulation of neutral genomic divergence via genetic drift (Slatkin 1987; Nosil *et al.* 2008). Thus, if divergent natural selection is identified, time since divergence and accumulated neutral genetic differentiation should also be considered alongside loci under selection to better understand the historical relationship between organism and ecosystem.

Tropical ecosystems provide an unparalleled concentration of biodiversity, with tropical forests alone accommodating perhaps two-thirds of the Earth's terrestrial diversity (Gardner *et al.* 2009). The Amazon Basin is no exception. Apart from its huge terrestrial biodiversity, the Amazon fluvial system sustains the world's richest freshwater fish fauna, with over 3000 species and 60 fish families (Reis *et al.* 2003). The evolutionary and ecological forces elemental to the genesis of such diversity are controversial and poorly resolved (Beheregaray 2008). Like other tropical systems (reviewed in Moritz *et al.* 2000), Amazonian diversity has largely been described within the context of climatic and geological history with limited focus on ecologically based divergent natural selection. However, as evidence accumulates supporting Darwin's theory (1859) that reproductive isolation can arise as a consequence of the interaction between individuals and their environment (reviewed in Schluter 2000; Coyne & Orr 2004), it is also becoming recognized that ecologically driven adaptive divergence may actually be a widespread force in the accumulation of tropical diversity (Smith *et al.* 1997, 2001; Schneider *et al.* 1999; Garcia-Paris *et al.* 2000; Ogden & Thorpe 2002; Freedman *et al.* 2010; Lopez-Fernandez *et al.* 2010; Cooke *et al.* 2012).

The Amazonian aquatic environment provides an ideal platform to examine the influence of divergent natural selection in the generation of tropical biodiversity. In particular, it sustains dramatic hydrochemical and ecological gradients that impose physiological constraints upon its aquatic communities (Junk *et al.* 1983; Henderson & Crampton 1997; Rodriguez & Lewis 1997;

Saint-Paul *et al.* 2000; Petry *et al.* 2003). These aquatic conditions have been grouped into three water types or 'colours' and are differentiated largely by sediment composition, geochemistry and optical characteristics (Sioli 1984): (i) white water that has an Andean origin is turbid in nature and characterized by large amounts of dissolved solids; (ii) clear water that is comparatively transparent contains a low content of dissolved solids; and (iii) black water that is transparent yet stained by tannins and humic acids leached from vegetation differs most dramatically from the latter by its low pH. Both clear and black water differ from white water in that they are craton born, draining the Brazilian and Guyana shields, respectively (Hoorn *et al.* 2010a). As a result, their sediment composition and channel formation are different from the fast-turbid white waters. Major ecological gradients and discontinuities have been shown to generate diversity via divergent natural selection elsewhere (Endler 1973; Smith *et al.* 1997, 2001; Thorpe *et al.* 2008; Berner *et al.* 2009). Thus, we predict that major differences in water colour between rivers of the Amazon Basin provide ecological opportunities for divergent natural selection to drive genetic differentiation between populations of aquatic organisms.

This study combines genome scans, population genetics, sequence-based phylogeographic analyses and a powerful sampling design to assess the influence of hydrochemical gradients as biodiversity drivers in the Amazonian fish *Triportheus albus*. *Triportheus albus* (Cope 1872) is a medium-sized benthopelagic characin fish abundant in white, black and clear waters of the Amazon Basin. Our study system consists of five major rivers (Amazon, Madeira, Branco, Negro and Tapajós Rivers), three hydrochemical settings (white, black and clear water), two ecological gradients (i.e. black meets white and white meets clear water) and two controls in which river systems of the same water colour meet. Surveys aimed to assess the influence of divergent natural selection in the landscape usually need to disentangle ecological divergence from allopatric biogeographic history. Thus, our study aims were framed to (i) clarify the phylogeographic history in *T. albus* from white, black and clear water; (ii) quantify divergent natural selection using genome scans and examine associations of selected genomic regions with water colour; and (iii) assess the extent of isolation by environment between populations of *T. albus* spanning different water colours.

Within our study system, we show that the selection gradients generated by water colour may actually be more powerful in generating diversity in the Amazon Basin than geographic isolation *per se*. Our results have important implications for programmes aimed at measuring ecological diversity and evolutionary potential in one of the world's most diverse and iconic ecosystems. It also raises fundamental questions about the origin and distribution of adaptive genetic diversity in tropical rainforests, a topic of global significance because of the threatened nature of these poorly studied ecosystems.

Materials and methods

Sampling design and molecular methods

Our study includes populations from white, black and clear water habitat. It encompasses five major river systems and the three main hydrochemical settings of the Amazon Basin: the Amazon (white), Madeira (white), Branco (seasonally black), Negro (black) and Tapajós (clear) rivers. Within our sampling transect, the black waters of the Negro, the white waters of the Madeira and the clear waters of the Tapajós Rivers coalesce into the Amazon River that drains west to east trans-continentally. Additionally, our transect includes the black water portion of the Branco River that meets the Negro River. In this way, our sampling design consists of two ecological gradients, where black (Negro) and clear (Tapajós) water meets white (Amazon) water and two controls, where rivers of the same water colour meet (Branco to Negro and Madeira to Amazon) (Fig. 1). During January and February of 2005 and 2008, we conducted expeditions along ~2200 km of riverine distance, all in Brazilian territory and sampled 179 *Triportheus albus* from 10 sampling sites (Fig. 1; Table 1). *Triportheus elongatus*, whose distribution is sympatric with *T. albus*, were also sampled for outgroup comparison. Geographic coordinates of sampling sites were recorded using a global positioning system. Fish were caught using beach seine nets, euthanized and muscle tissue dissected and preserved in 95% ethanol. Temperature, pH, turbidity (measured with a Secchi disk), dissolved oxygen and oxygen saturation were measured at each sampling site in both years.

DNA was extracted using a modified salting-out method (Sunnucks & Hales 1996). The mitochondrial adenosine triphosphatase subunits 6 and 8 (ATPase 6 and 8) were amplified via polymerase chain reaction (PCR) and sequenced for 114 samples of *T. albus*. ATPase 6 and 8 were amplified using the primers ATP8.2 and CO3.2 (Bermingham & Martin 1998) with PCR conditions as in Corrigan *et al.* (2008). All PCR products were sequenced on an automated Sequencer AB3730xl.

Amplified fragment length polymorphism (AFLP) profiling was performed on all samples (n = 179) using a modified protocol after Zenger *et al.* (2006). Restriction fragments were amplified in a single selective PCR round using two selective *Eco*RI primer carrying three

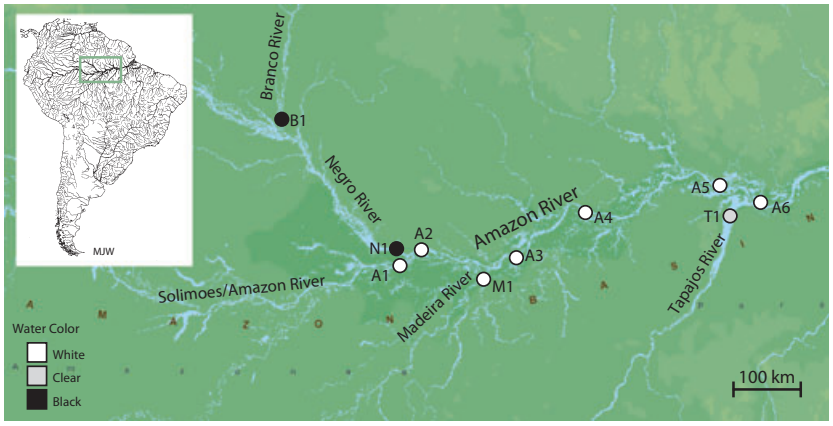

Fig. 1 Sampling localities of *Triportheus albus* in the Amazon Basin. Each site label has been shaded according to water colour. Black water sites include B1 and N1, white water sites include A1–A6 and M1, clear water sites include T1. Inset shows the study area in the northern region of South America.

Table 1 Water 'colour', sampling locations, sample size and average hydrochemical variables collected in 2005 and 2008 [temperature, °C; pH; turbidity, cm; dissolved oxygen (mg/L); oxygen saturation O_2 (%)]

River	Site	Latitude	Longitude	*N*	°C	pH	cm	OD	O_2%
Black									
Branco	B1	1°15′59.00′ S	61°50′55.00″ W	24	29.8	6.91	83.4	6.92	90.4
Negro	N1	3°4′44.00″ S	60°14′44.00″ W	19	29.7	5.2	76.0	6.4	82.3
White									
Madeira	M1	3°28′14.00″ S	58°52′5.00″ W	10	29.7	7.1	5.5	5.7	82.4
Amazon	A1	3°20′40.00″ S	60°7′10.00″ W	13	28.8	7.2	12.3	6.7	86.3
Amazon	A2	3° 6′56.00″ S	59°32′19.00″ W	41	29.6	7.1	18.8	6.3	85.2
Amazon	A3	3° 4′39.00″ S	58°13′13.00″ W	34	28.7	7.1	18.3	4.8	84.0
Amazon	A4	2°33′7.00″ S	57°1′59.00″ W	4	29.2	7.2	10.5	6.4	85.6
Amazon	A5	2°10′21.00″ S	54°58′21.00″ W	3	29.0	7.2	12.5	6.3	82.0
Amazon	A6	2°28′10.00″ S	54°30′5.00″ W	5	29.7	7.2	15	6.6	87.9
Clear									
Tapajós	T1	2°52′17.00″ S	55° 9′38.00″ W	25	29.5	6.7	118	7.1	97.5

All measurements (except turbidity) are averaged from riverbed and surface measurements.

selective nucleotides (5′-GAC TGC GTA CCA ATT C + AGT or AAC-3′) and four selective *Mse*I primers carrying four selective nucleotides (5′-GAT GAG TCC TGA GTA A + CAAC, CTGC, CAGC or CTTC-3′) (Applied Biosystems). The fragments were then denatured and separated on an AB3130xl automated sequencer.

The presence and absence of AFLP fragments were determined using GENEMAPPER 4.0 (Applied Biosystems), within a range of 50–500 bp and a 1-bp call accuracy. Peaks with heights less than 50 relative fluorescent units were not scored, and every bin fragment position identified using GENEMAPPER was checked manually. AFLPScore1.4 (Whitlock *et al.* 2008) was employed to normalize and score our data within the predefined fragment locations and to calculate mismatch error rates for each primer combination using approximately 10% of samples per primer combination that was rerun as replicates.

Phylogeography and demographic history

Mitochondrial sequences were aligned using SEQUENCHER 4.1 (Gene Codes Corporation, Ann Arbor, MI, USA). ATPase 6 and 8 sequence diversity was estimated as haplotypic and nucleotide diversity (Nei 1987) in ARLEQUIN 3.11 (Excoffier *et al.* 2005). Genealogical relationships were investigated by constructing a haplotype network in TCS (Clement *et al.* 2000) using the ATPase 6 and 8 mtDNA sequences.

We further investigated phylogenetic relationships using character- and model-based methods: maximum

parsimony (MP) and maximum likelihood (ML) in PAUP*4.0b10 (Swofford 2003). For MP, a heuristic search strategy was used, and all characters were treated as unordered and unweighted. Bootstrap resampling was based on 1000 replicates for a majority-rule consensus tree. For ML, we used MODELTEST v3.07 (Posada & Crandall 1998) to estimate the most appropriate model of sequence evolution. According to the Akaike Information Criterion, the Tamura & Nei (1993) plus invariant sites (+I) was the most likely model of DNA substitution for our data set. The ML tree was constructed using a heuristic search strategy and appropriate model parameters as specified by MODELTEST v.3.0.7.

To gain insight about the history of migration rate across water colours, we used the isolation with migration model IMa to analyze the ATPase 6 and 8 sequence data (Hey & Nielsen 2004, 2007). After separating from an ancestral population, the IMa model includes gene flow with rates m_1 and m_2, respectively. We used IMa to estimate posterior probabilities of migration rates between black and white water and between white and clear water populations. Using this estimation, we calculated the population migration rate as $2N_1m_1$. Each comparison was run twice until stationary distribution was achieved, and the minimum effective sample size exceeded 50 (Hey 2005). Each run included four Metropolis-coupled Markov chains (MCMC), a linear heating scheme with the first-heating parameter set to 0.05, and a maximum of 10 multiple chain swapping attempts.

Major demographic events potentially associated with the evolutionary history of *T. albus* populations were assessed by the mismatch analysis as implemented in ARLEQUIN 3.01 (Excoffier *et al.* 2005) by testing the data against a model of demographic expansion (Rodgers & Harpending 1992). When applicable, we calculated the estimator of time to the expansion (τ) and mutation parameter (θ) according to Schneider & Excoffier (1999). From this, we used the formula $t = \tau/(2\mu)$ to estimate the timing of putative population expansion where t is given in generations, and μ is the mutation rate for ATPase 6 and 8. We adopted the ATPase 6 and 8 mutation rate of 1.4% per million years after Bermingham *et al.* (1997). Additionally, Fu's (1997) test of selective neutrality was employed to assess the signal of demographic expansion in the data set. In the event of demographic expansion, large negative F_S values are generally observed (Fu 1997).

Analysis of adaptation and population structure

Outlier loci were detected in our nuclear data set using two methods. We first used the program DFDIST distributed by M. A. Beaumont, which implements the approach of Beaumont & Balding (2004), with modifications for dominant markers. The methodology is based on the assumption that loci under divergent selection (positive selection) will display significantly higher F_{ST} values than the majority of neutral loci in a sample. In contrast, loci under stabilizing selection (negative selection) are expected to display significantly low F_{ST} values. From the empirical data set, null allele frequencies for each locus are estimated using the Bayesian method of Zhivotovsky (1999). DFDIST then calculates a 'trimmed' mean F_{ST} by removing 30% of the highest and lowest values (Beaumont & Balding 2004). This 'trimmed' mean represents the 'neutral' F_{ST} values, uninfluenced by loci under positive or negative selection (outlier loci). The model then implements a hierarchical-Bayesian approach to compute F_{ST} values conditional on heterozygosity in a subdivided population under the infinite island model (Wright 1943). Empirical and simulated distributions are then compared, identifying upper and lower confidence intervals and the resulting outlier loci. We chose not to specifically test for loci with unusually low F_{ST} because divergence-based methods are known to have little power to detect stabilizing selection (Beaumont & Nichols 1996; Beaumont & Balding 2004).

We searched for (positive) directional selection within 12 pairwise population comparisons and three additional comparisons between water colours (Fig. 4). For each analysis, loci with allele frequencies higher than 0.98 were excluded. The trimmed mean F_{ST} was calculated, and a null distribution of F_{ST} was generated using 50 000 realizations. This was compared to the empirical distribution, and outlier loci were identified at the upper 95th quantile using a smoothing parameter of 0.04.

Loci were classified into three groups after Nosil *et al.* (2008): (i) putatively neutral loci, those that were not outliers in any comparison, (ii) nonrepeat outlier loci, outliers detected in only one pairwise comparison and (iii) repeat outlier loci, outliers detected in multiple comparisons. Additionally, we noted the geographic and ecological setting of these outliers to make inferences about the possible forces driving outlier behaviour. Nonrepeat outlier loci were considered potential false positives, irrespective of their *P*-values. While methods implemented in DFDIST explicitly address the issue by employing a prior distribution to correct for the potential problem of multiple testing (Beaumont & Balding 2004), we chose this conservative approach to further minimize type I errors rather than minimizing type II errors by including all outliers.

The second program we used to detect outlier loci was BAYESCAN (http://www-leca.ujf-grenoble.fr/logiciels.htm) that implements the method of Foll & Gaggiotti (2008). BAYESCAN employs a Bayesian method based on

the multinomial Dirichlet model and directly estimates the probability of each locus being subject to selection. BayeScan is considered robust when dealing with complex demographic scenarios because it estimates population-specific and locus-specific F_{ST} coefficients (Foll & Gaggiotti 2008). To identify outlier loci, we focused on the posterior distribution of α_i where a positive value is indicative of positive selection and a negative value is indicative of stabilizing selection. The significance of α_i for each locus is estimated in BAYESCAN using a reversible jump Metropolis-coupled Markov chains (MCMC) algorithm, and loci were classified according to Jeffreys (1961) scale of evidence. We therefore used the Bayesian approach of BAYESCAN on the entire data set in addition to the pairwise comparisons performed using DFDIST. Outliers detected by both methods are more likely to represent truly adaptive regions within the genome because the two separate approaches have completely different assumptions and employ separate algorithms. All outlier loci detected using both methods, together with repeat outliers detected using DFDIST were removed from the total data set for the subsequent data analyses.

Allele frequencies were estimated in AFLP-SURV (Vekemans *et al*. 2002) using the Bayesian approach of Zhivotovsky (1999) for dominant markers with a non-uniform prior and assuming Hardy–Weinberg. The number and proportion of polymorphic loci, expected heterozygosity and average gene diversity was calculated for all populations, with the unbiased estimator of Lynch & Milligan (1994). All data analyses were performed using only polymorphic loci. Population structure (F_{ST}) based on the estimated allele frequencies were calculated according to Lynch & Milligan (1994). The significance of F_{ST} values was determined with 1000 bootstrap replicates.

Hierarchical patterns of population structure were further explored with AMOVA using ARLEQUIN for mtDNA (Φ_{ST}) and GENALEX 6.3 (Peakall & Smouse 2006) for AFLPs (based on the F_{ST} analogue Φ_{ST} that accounts for binary data). Two separate AMOVAS were conducted. AMOVA 1 partitioned the data set into regions defined by hydrochemistry: (i) seasonally black (B1), (ii) black (N1), (iii) white (A1-A6, M1) and (iv) clear water (T1) populations. AMOVA 2 in contrast was our control that tested for population subdivision between two true white water rivers: (i) The Madeira River (M1) and (ii) the Amazon River (A1–A5).

Genetic subdivision was also investigated using Structure v.2.3.1 (Falush *et al*. 2003). Using the admixture model, we determined the number of populations, *K*, by comparing the log-likelihood ratios across multiple independent runs by varying the assumed number of *K* from 1 to 10. Each run was replicated three times and consisted of a burn-in phase of 100 000 iterations, followed by 1 000 000 iterations. Structure was run using both our data set of putatively neutral AFLP loci as well as the total data set.

The distribution of AFLP genotypes in space was explored using a spatial autocorrelation analysis (Smouse & Peakall 1999). Spatial autocorrelation identifies the scale at which spatial patterns occur by assessing the genetic relationship between pairs of individuals at defined distance classes. Distance classes were set at 100 km intervals, and our data were tested against the null hypothesis of randomly distributed genotypes using GENALEX with 9999 permutations and 10 000 bootstrap replicates.

Results

We obtained and confidently aligned 792 bp of ATPase 6 and 8 genes for 114 *Triportheus albus* samples and the out-group. Of these, 28 unique haplotypes were identified. These were composed of 30 variable characters, of which nine were parsimony informative. AFLP profiles were resolved for 177 individuals, and 360 AFLP loci were scored. Mismatch error rates calculated from AFLPScore were on average 7.5% per primer combination. This is within the acceptable AFLP error rate range of 5–10% (Bonin *et al*. 2007). Summary statistics for both mtDNA and the putatively neutral AFLP loci are shown in Table 2. In each case, moderate levels of variability were detected at each population.

Population history

The mitochondrial data show a very strong association between haplotype and water colour (Fig. 2). Both the ML and MP phylogenetic trees and the haplotype network identify two major clades. The basal and presumably ancestral clade consists of samples from white water populations (i.e. Amazon and Madeira River haplotypes), with the exception of one haplotype also found in the Negro River. In the network, TCS identified the most frequent white water haplotype as ancestral. In contrast, the most recently derived clade consists of all samples from either black or clear water populations, with the exception of one haplotype also sampled in white water. Strikingly, samples from black and clear water populations separated by over 1000 km appear more related to each other than to neighbouring white water populations. Despite the genealogical subdivision observed in the data set, all maternal lineages sampled appear as closely related, as evidenced by very low nucleotide diversity at the population level (π ranged from 0.0004 to 0.0033; Table 2) and for the whole sample (π = 0.00860).

Results from mismatch analysis and Fu's test of neutrality suggest that the history of *T. albus* has been

Table 2 Genetic diversity at mtDNA and AFLP markers

	mtDNA diversity				AFLP diversity		
Site	*n*	*N*	*h*	π	*n*	PLP	He
B1	20	7	0.8158 (0.0575)	0.00201 (0.00139)	24	81.7	0.32530
N1	8	5	0.8571 (0.1083)	0.00204 (0.00153)	19	89.5	0.37084
M1	8	4	0.6429 (0.1841)	0.00149 (0.00149)	10	90.1	0.33492
A1	8	6	0.9286 (0.0844)	0.00299 (0.00299)	13	91	0.32529
A2	20	4	0.2842 (0.1284)	0.00038 (0.00046)	41	89.5	0.36037
A3	22	6	0.6710 (0.0768)	0.00150 (0.00111)	34	87.3	0.34117
A4	3	1	0.00	0.00	4	81.4	0.30731
A5	3	2	0.6667 (0.3143)	0.00339 (0.00304)	3	74.3	0.33681
A6	5	3	0.7000 (0.2184)	0.00279 (0.00215)	5	79.3	0.32838
T1	21	5	0.5810 (0.1025)	0.00084 (0.00075)	24	86.1	0.32619

Mitochondrial diversity is summarized by (*n*) sample size, number of haplotypes (*N*), haplotypic diversity (*h*) and nucleotide diversity (π). AFLP diversity is summarized by sample size (*n*), estimated proportion of polymorphic loci at the 5% level per population (PLP) and expected heterozygosity (He).
AFLP, amplified fragment length polymorphism.

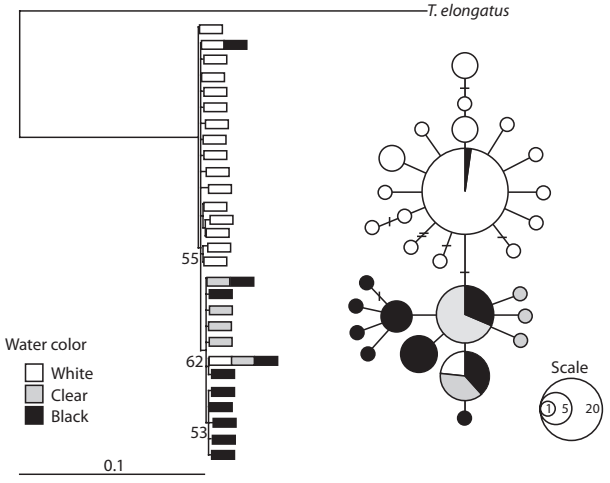

Fig. 2 Maximum likelihood (ML) phylogeny with bootstrap support for the major clades and statistical parsimony network of *Triportheus albus* based on 28 mtDNA haplotypes identified with ATPase 6 and 8 sequences. The shade/s of the circle or rectangle represents the water colour. In the network, each circle denotes a unique haplotype and the area of the circle is proportional to its frequency in the sample. *Triportheus elongatus* is the out-group used for the ML analysis.

influenced by strong demographic expansion(s), a finding corroborated by the star-like genealogy seen in the haplotype network. First, our data did not significantly deviate from a model expected under demographic expansion (sum of squared deviations = 0.001, $P = 0.45$; raggedness index = 0.026, $P = 0.736$). Also, consistent with a scenario of demographic expansion, Fu's test of neutrality gave a negative F_S value (−6.158). Assuming an ATPase 6 and 8 mutation rate proposed for groups of geminate fishes across the Panama Isthmus (Bermingham *et al*. 1997), the estimated τ values suggest a major population expansion around 1 Ma (±1).

Given the association between genealogy and water colour, we considered the directionality of migration between white and black or clear water populations using IMa. While single-locus posterior distributions must be interpreted with caution, in the comparison between black and white water populations, we were able to reject the simple isolation model ($m_1 = m_2 = 0$). The posterior probability of no migration was practically zero. Rather, we identified a bidirectional signature of migration, albeit very small, between black and white water populations (Fig. 3, Table S1, Supporting information). Here, the population migration rate from white to black water was higher than from black to white water (white → black = 2.14; black → white = 3.46). In contrast, the comparison between white and clear water identified a largely unidirectional signature of migration (Fig. 3, Table S1, Supporting information). In this instance, the population migration rate from white to clear water was lower than from clear to white water (white → clear = 1.78; clear → white = 22.68). It should be noted that while the population migration rate from clear to white water is vastly higher than all other pairwise estimates, it is likely due to the slightly bimodal posterior probability curve. Nonetheless, in both pairwise comparisons, the directionality of gene flow is towards white water rather than away from white water.

Spatial distribution of adaptive and neutral population structure

Using DFDIST to search for positive selection between populations and across water colours, we identified 65

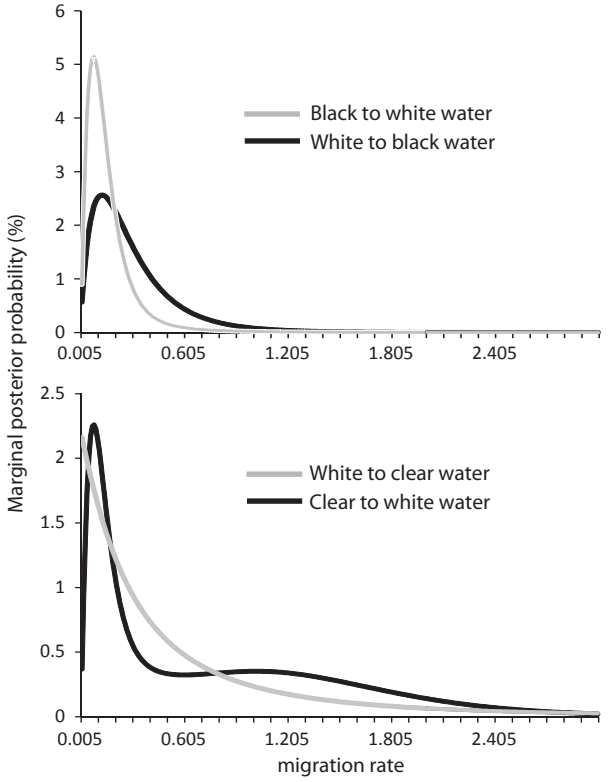

Fig. 3 The posterior distributions of migration rate (m_1 and m_2) estimates for pairwise comparisons between black and white and white and clear water populations using the IMa program.

outlier loci (18.5% of the total) at the 99% quantile. Of the total outlier loci identified, 29 (8% of the total) were classified as 'repeat outliers' that were unlikely to represent false positives (Cooper 2000; Campbell & Bernatchez 2004). In every possible case, more outlier loci were identified in comparisons between water colours than within (Fig. 4; Table 3). For example, 19 repeat outlier loci were identified when comparing the black water site N1 and the white water site A2. In contrast, fewer than five repeat outlier loci were identified in all comparisons of the same water colour ($n = 7$). All repeat outlier loci were removed to create a putatively neutral data set with 323 polymorphic AFLP loci.

The hierarchical-Bayesian approach of BAYESCAN performed on all data identified 133 loci with positive α_i values. Of the 65 outlier loci detected using Dfdist, 59 of these were also identified using BAYESCAN. Of the 29 'repeat' outliers detected using DFDIST, BAYESCAN identified 100% of these. However, using the Jeffreys (1961) scale of evidence, only 12 loci were supported with $\log_{10}$ (Bayes factor) >2, which is considered 'decisive' evidence for selection, one was supported with $\log_{10}$ (Bayes factor) >1.5, which is considered 'very strong', two were supported with $\log_{10}$ (Bayes factor) >1, which is considered 'strong' and two were supported with $\log_{10}$ (Bayes factor) >0.5, which is considered 'substantial'. The remaining loci with positive α_i values had $\log_{10}$ (Bayes factor) <0.5, which is poor evidence for selection. However, of the 17 loci (5% of the total) with $\log_{10}$ (Bayes factor) >0.05, 15 were also detected using DFDIST. Two of these were detected in only white vs. black water comparisons, one was identified only in white vs. clear water comparisons, while 12 were identified in both black vs. white and white vs. clear water comparisons. Only one of the

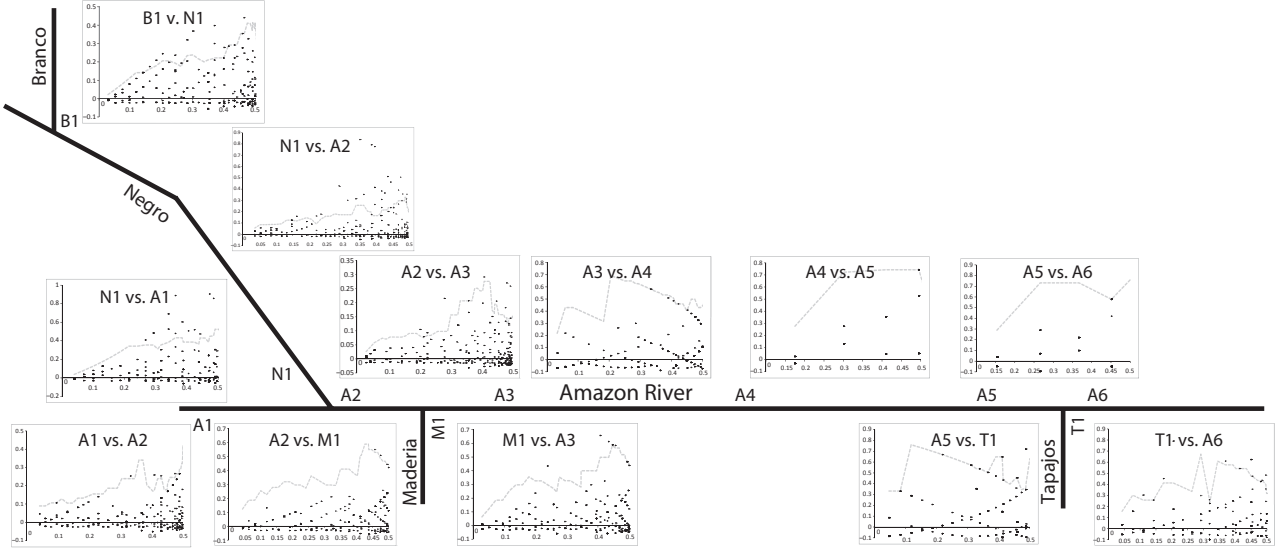

Fig. 4 The empirical distributions of FST values from DFDIST analysis for amplified fragment length polymorphism loci for each population pairwise comparison. The solid line represents the 99th quantile, and dots exceeding the 99th quantile are 'outlier loci'. Results are overlaid over a schematic representation of the study region and sampled locations.

Table 3 Number of outlier and repeat outlier AFLP loci found in each pairwise comparison, total F_{ST}, the putative 'neutral' F_{ST} value, Nei's unbiased genetic distance (after Lynch & Milligan 1994) and the mtDNA Φ_{ST} value

Comparison	No. 99% outliers	No. repeat outliers	Total F_{ST}	Neutral F_{ST}	Nei's genetic distance	mtDNA Φ_{ST}
Black vs. white	22	21	–	0.0273*	0.0152	–
Black vs. clear	5	3	–	0.0237*	0.0123	–
White vs. clear	17	16	–	0.0340*	0.0180	–
B1 vs. N1	11	2	0.0456*	0.0403*	0.0221	0.09544
N1 vs. A2	25	19	0.0588*	0.0193*	0.0114	0.7844*
N1 vs. A1	13	12	0.1060*	0.0617*	0.0351	0.54739*
A1 vs. A2	6	2	0.0303*	0.0302*	0.0160	0.1503*
A2 vs. A3	9	3	0.0215*	0.0195*	0.0107	0.09902*
A2 vs. M1	2	2	0.0465*	0.0427*	0.0238	0.07714*
M1 vs. A3	5	4	0.0704*	0.0648*	0.0360	0.07231
A3 vs. A4	0	–	0.0828*	0.0796*	0.0421	−0.0879
A4 vs. A5	0	–	0.1014*	0.1030*	0.0558	0.5000
A5 vs. A6	0	–	0.0539*	0.0506*	0.0269	0.49197
A5 vs. T1	2	1	0.0844*	0.0720*	0.0392	0.81327*
T1 vs. A6	9	5	0.0702*	0.0437*	0.0225	0.29399*

AFLP, amplified fragment length polymorphism.
*$P \leq 0.05$.

outlier loci detected was also represented in white vs. white water population comparisons.

Fixation indices showed a strong signal of population structure associated with water colour (Table 3). For mtDNA Φ_{ST}, all pairwise population comparisons between water colour samples were statistically significant, while Φ_{ST} for comparisons of the same water colour were either nonsignificant or much lower. For both the total and putatively neutral AFLP data sets, all pairwise population comparisons were significant. However, after the removal of outlier loci, F_{ST} estimates were generally much lower for pairwise comparisons that bridged different water colours whereas comparisons of the same water type showed little difference (Table 3). For pairwise comparisons involving black and white water populations, F_{ST} values for our putatively neutral data were in some cases less than halved from those using the total data set (N1 vs. A1: total $F_{ST} = 0.106$, neutral $F_{ST} = 0.062$; N1 vs. A2: total $F_{ST} = 0.059$, neutral $F_{ST} = 0.019$). Similarly, comparisons involving white and clear water were also markedly reduced in our neutral compared to total data sets (A5 vs. T1: total $F_{ST} = 0.084$, neutral $F_{ST} = 0.072$; T1 vs. A6: total $F_{ST} = 0.070$, neutral $F_{ST} = 0.044$). Pairwise comparisons between white water or between black water sites showed little change in F_{ST} after the removal of outlier loci.

Regional patterns of population structure associated with water colour were also evident in AMOVAS for both mtDNA and AFLPs (Table 4). The first AMOVA,

Table 4 Analysis of molecular variance (AMOVA) for mtDNA and AFLP data

	AMOVA 1				AMOVA 2			
Data type	Source of variation	% Variation	FI	*P*	Source of variation	% variation	FI	*P*
mtDNA	Among regions	46.29	Φ_{CT}: 0.463	0.008*	Among regions	−27.16	Φ_{CT}: −0.272	0.665
	Among populations	13.50	Φ_{SC}: 0.251	0.000*	Among populations	45.09	Φ_{SC}: 0.355	0.000*
	Among individuals	40.21	Φ_{ST}: 0.598	0.000*	Among individuals	82.08	Φ_{ST}: 0.179	0.000*
AFLP	Among regions	3	Φ_{RT}: 0.027	0.000*	Among regions	1	Φ_{RT}: 0.014	0.057
	Among populations	5	Φ_{PR}: 0.053	0.000*	Among populations	4	Φ_{PR}: 0.045	0.000*
	Among individuals	92	Φ_{PT}: 0.079	0.000*	Among individuals	94	Φ_{PT}: 0.058	0.000*

In AMOVA 1, regions include (i) seasonally black/white (B1), (ii) black (N1), (iii) white (M1, A1, A2, A3, A4, A5, A6) and (iv) clear water populations (T1). In AMOVA 2, regions include (i) the Madeira River and (ii) white water sites A2-A6.
FI, fixation index; AFLP, amplified fragment length polymorphism.
*Significant results.

which tested for genetic structure across black, white and clear water populations was significant for both data sets (mtDNA: $\Phi_{CT} = 0.463$, $P = 0.008$; AFLP: $\Phi_{RT} = 0.027$, $P < 0.001$). In contrast, AMOVA 2, our control that tested for structure associated with the confluence of the Madeira (white water) and the Amazon (white water) Rivers was nonsignificant in both cases (mtDNA: $\Phi_{CT} = -0.272$, $P = 0.665$; AFLP: $\Phi_{RT} = 0.014$, $P = 0.057$).

The STRUCTURE analyses for both the total and putatively neutral AFLP data sets were consistent with the mtDNA haplotype network analysis that grouped the geographically distant black and clear water populations separately from white water populations (Fig. 5a, Table S2, Supporting information). The Structure analysis based on all polymorphic loci ($n = 352$) identified four populations [ln Pr ($X \mid K = 4$) = −24575.4]. Consistent with mtDNA, there is a pronounced population subdivision at the interface of the Negro and Amazon Rivers and at the Tapajós River. Similarly, samples from the Tapajós River were identified as belonging to the same population as those from the Branco and Negro Rivers. There were no obvious breaks between the Madeira River and other white water populations of the Amazon River.

Results from STRUCTURE analysis based on our putatively neutral loci ($n = 323$) differed distinctly from the total AFLP data set for the Negro River population (Fig. 5b). Here, the most likely number of populations was 6 [ln Pr ($X \mid K = 6$) = −21860.2], yet this did not resolve any further clusters from $K = 3$ [ln Pr ($X \mid K = 3$) = −22560.6]. Thus, we chose $K = 3$ as the most likely number of populations for our putatively neutral loci, because increasing K did not inform further on population structuring. In this analysis, the Negro River samples were no longer similar to those from the Branco and Tapajós River. Rather, the Negro River samples appeared more similar to those from white water populations. The genotypic composition in the Negro River however was mildly distinct from the white water populations, with a greater incidence of genotypes with high membership to a third green population than to other white water populations. In contrast, samples from the Branco and Tapajós Rivers appeared largely unaffected by the removal of outlier loci, although some admixture can be seen between the Tapajós River and nearby white water populations. Apart from the marked population breaks observed, the Structure analysis also points out the existence of a few migrant and intermixed individuals between different water colour populations (e.g. from the Amazon to Negro River and from the Amazon to Tapajós River), a result consistent with the expectations based on a model of divergence with gene flow.

The results from spatial autocorrelation analysis reiterated the results from STRUCTURE analyses that the overall *T. albus* population structure is not driven solely by geographic isolation (Fig. 6, Table S3, Supporting information). Significant positive spatial structure is seen between pairs of individuals from the same sampling site. After the intercept at approximately 130 km, genetic similarity fell significantly lower than random expectation and decreased until 600 km. After 600 km, there is a sharp increase in autocorrelation signal that exceeds random expectation and becomes positive at

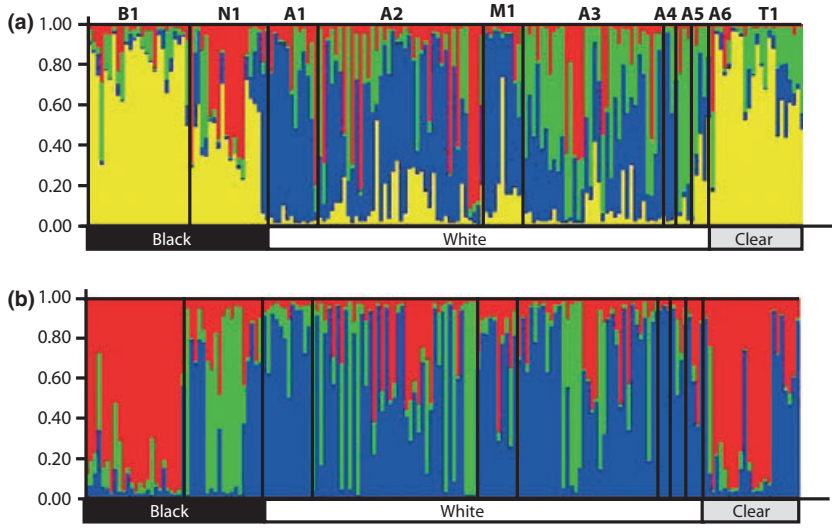

Fig. 5 STRUCTURE results of total amplified fragment length polymorphism (AFLP) data set (a: $n = 352$ loci) and the putatively neutral AFLP data set (b: $n = 323$ loci). Individuals are grouped by sampling location, and each individual is represented by one vertical column. Sampling sites are shown at the top of the figure and separated by black bars. The water colour of each site is shown at the bottom.

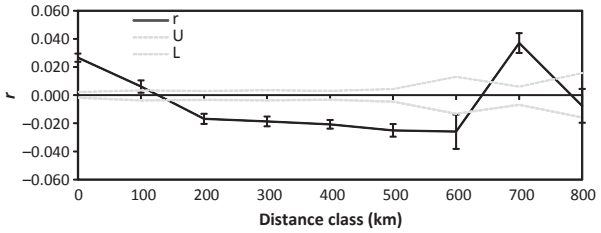

Fig. 6 Correlogram plot of the spatial autocorrelation coefficient, r, as a function of geographic distance. Upper U and lower L bounds for the 95% confidence interval for the null hypothesis of no spatial structure ($r = 0$) based on 9999 random permutations of the data are depicted as dotted lines.

the 700-km distance bin. The 700 km has a relatively low sample size and wider confidence bounds, and the result could also be related to the randomness of autocorrelation patterns and greater distances (Smouse & Peakall 1999). At moderate spatial scales, a pattern of isolation by distance (IBD) is evident. Nonetheless, the pattern observed is consistent with other analyses that suggest samples separated by over 600 km are more genetically similar to one another than samples occurring in closer geographic proximity.

Discussion

Here, we show evidence for strong divergent natural selection with gene flow in *Triportheus albus* at the interface of large-scale hydrochemical gradients in Amazonia. Our analysis of population history and contemporary genetic connectivity consistently demarcates two diverging lineages in *T. albus*: one typical of white water rivers and the other distributed in both black and clear water rivers. Using outlier loci approaches based on genome scans, we quantified divergent selection between populations of *T. albus* and show that heightened selection is present at the interface of different water colours, irrespective of river system and tributary arrangement. Additionally, we assess the extent of outlier loci identified between populations of *T. albus* spanning different water colours and find that a spectrum of adaptation exists because of factors such as geographic isolation and gene flow. Thus, consistent with a pattern of ecological speciation with gene flow, we show that selection generated by water colour may be more powerful in generating diversity in a widespread Amazonian aquatic organism than geographic isolation or vicariant history.

Phylogeography of diverging water colour lineages

Triportheus albus is a primary division fish (i.e. 'true' freshwater fish that does not tolerate saltwater) of the order Characiformes, an ancient group that originated in Gondwana over 100 million years ago (Ma) (Orti & Meyer 1997). Thus, major biogeographic events associated with the evolution of this species should be assessed in the context of the formation of the Amazon and its main tributaries, largely influenced by the history of Andean uplift (Rull 2008; Hoorn *et al.* 2010b). Briefly, increased sedimentation and sea level changes during the late Miocene resulted in the overfilling of the Andean foreland basin. As a result, the proto-Amazon River eventually breached the Purus Arch, flooding a vast area in eastern Amazonia and establishing the modern west to east transcontinental flow of the turbid Amazon River (Campbell *et al.* 2006; Figueiredo *et al.* 2009; Lundberg *et al.* 2010). While the Amazon drainage system was largely formed by the late Miocene (Hoorn *et al.* 1995; Lundberg *et al.* 1998), the full establishment of the modern Amazon River probably occurred only during the late Pliocene around 2.5 Ma (Campbell *et al.* 2006). In our study, both the phylogenetic and phylogeographic evidence (Fig. 2) and summary statistics (e.g. nucleotide diversity, Table 2) strongly indicate that divergence between the two identified clades of *T. albus* is recent. In addition, the results from mismatch analysis and Fu's test of neutrality against population growth suggest at least one major event of demographic expansion in *T. albus*, tentatively dated to the Pleistocene. The emerging scenario supported by geomorphology, paleoecology, river course and genetics is that the final formation of the Amazon River encouraged the colonization and expansion of *T. albus* eastward (Lundberg *et al.* 1998, 2010; Campbell *et al.* 2006). A coalescent-based phylogeographic study using multilocus sequence data (Knowles & Maddison 2002) would probably provide a more statistically rigorous analysis of such a scenario than the one we have presented here. Nonetheless, the eastward downstream colonization hypothesis is also corroborated by comparative phylogeographic studies of three fishes (Amazonian puffer, *Colomesus asellus*, silver croaker, *Plagioscion squamosissimus* and the barred knife fish, *Steatogenys elegans*) sampled from the very same sites as *T. albus* (Cooke *et al.* 2012).

Until relatively recently, craton-derived fluvial systems (e.g. the Negro and Tapajós) actually dominated the Amazon Basin. Thus, the Neogene shift in drainage and sediment source to the Amazon Basin also signified a massive rearrangement of composition and distribution of aquatic species (Hoorn *et al.* 2010a). If ancestral populations of *T. albus* inhabited the paleo-Amazonian white water (as proposed above), the colonization into the black and clear water habitats of the Negro, Branco and Tapajós Rivers must have occurred after the colonization of the modern Amazon River. Thus, the ancestral condition for the species would be a white water ecotype. Indeed, individuals sampled in white water

generally had ancestral haplotypes (i.e. 16 of 17 cases), whereas those from black or clear waters were in a recently derived clade (Fig. 2). Although the mtDNA weakly supports the white water ecotype as ancestral, a valid alternative may be that the black/clear water ecotype is ancestral and the formation of the Amazon River actually promoted the recent evolution of the white water ecotype. This hypothesis is actually more consistent with the unusual relationship between genotype and distance.

Tripotheus albus sampled in black waters are more genetically related to those sampled in clear waters (~1000 km away) than they are to any neighbouring white water population. This remarkable pattern was not only supported by the mtDNA, but also by our nuclear data set as evident in results from Structure, AMOVA and spatial autocorrelation analysis based on AFLPs. For instance, the STRUCTURE analysis based on the total AFLP data (Fig. 5a) grouped the Branco, Negro and Tapajós Rivers as a single population, with some admixture occurring in the Negro River. Structure also grouped the white water Amazon and Madeira Rivers as a single population, an outcome statistically supported by AMOVA (Table 4). The latter is particularly important as it controls for population genetic structure associated with the confluence of a tributary alone.

We have identified a rather novel phylogeographic situation in which *T. albus* genotypes are associated with water colour, or sediment content, irrespective of river system or geographic distance. Considering this pattern, a simple model of allopatric divergence between ecotypes is not satisfactory to account for the phylogeography of *T. albus*. Rather, divergence patterns consistent with environmental conditions are indicative of a scenario of divergence with gene flow or divergent natural selection (Schluter 2001; Coyne & Orr 2004). While the precise migration rates have been interpreted with caution, the posterior distributions of migration between black and white water and white and clear water nonetheless rejected a pure isolation model ($m_1 = m_2 = 0$) because the posterior probability of no migration between ecotypes was practically zero (Fig. 3). Furthermore, in each comparison, the probability of gene flow towards white water (i.e. the expected downstream migration in this nonmigratory fish) was higher than the reverse. Alternatively, this result may indicate that selection against immigrants may be stronger in the black and clear waters than the white water. Migrants and intermixed individuals between different water colour populations were also identified by the Structure analysis based on the more resolving AFLP data set (Fig. 5a). Given this result of nonzero migration between divergent ecotypes, it is important to examine the likelihood of secondary contact and admixture after a period of allopatric isolation or whether this is the product of gene flow in the face of divergent natural selection.

Isolation by water 'colour'

The expansion into the modern Amazon River and the transition of *T. albus* from black or clear water to sediment rich white water environments present an opportunity for ecological speciation, where divergent selection between water colours leads to the evolution of reproductive isolation (Schluter 2000). Indeed, it has been shown that adaptive evolution can be facilitated by range expansion (Beheregaray & Sunnucks 2001; Zayed & Whitfield 2008; Freedman *et al.* 2010). Within the Amazon Basin, the greater nutrient richness of the white waters supports significantly greater aquatic plant and phytoplankton development than the comparatively depauperate clear and black waters (Fink & Fink 1973). This plant life in turn supports extraordinarily diverse animal communities, the main diet of most *Triportheus* species (Malabarba 2004). However, characins are generally visually oriented diurnal fish with large eyes, often preferring transparent well-lit surface waters (Tejerina-Garro *et al.* 1998). Thus, the transparency of the black and clear waters compared with white waters (Table 1) probably offers different ecological opportunities for *T. albus* that may have impacted colonization and range expansion into the modern nutrient-rich Amazon River.

During ecologically based divergent natural selection, selection will act in opposing directions between environments, such that fixation of alleles occurs separately in each environment (Schluter & Conte 2009). Our DFDIST genome scan detected 8% of the total loci in multiple population comparisons as repeat outliers. Because it is doubtful that these repeat outlier loci represent false positives (Campbell & Bernatchez 2004; Bonin *et al.* 2007) and that BAYESCAN also identified the same loci as outliers, they probably represent loci that are directly under selection or loci closely linked to causal loci via 'hitchhiking' (Jensen *et al.* 2007). The detection of these loci enabled us to directly assess the association of these genomic regions with major hydrochemical gradients within Amazonia. We found that the majority of DFDIST outlier loci ($n = 21$) were detected in pairwise comparisons between black and white water and between white and clear water populations. Very few loci ($n = 8$) were identified in pairwise comparisons of the same water colour. Further, very few loci were detected in pairwise comparisons between black and clear water ($n = 3$), both of which are craton-derived water types (Table 4). Considering the associa-

tion of the outlier loci to specific contrasting environments, we were able to rule out mutation-order divergence as a possible cause for the outlier loci detected. Accordingly, we argue that strong selective gradients exist between Andean- and craton-derived aquatic environments, and contemporary gene flow between ecotypes is unlikely a product of secondary contact after allopatric isolation.

While divergent natural selection is likely operating between *T. albus* from Andean (white water)- and craton (black and clear water)-derived habitats, the effects of this process have also filtered throughout the functionally neutral loci, leaving a signature of neutral divergence between selective environments, similar to IBA (without phenotypic comparison). Indeed, when outlier loci were removed from population genetic analyses leaving only putatively neutral loci, black and clear water populations remained distinct from white water populations irrespective of geographic distance (Fig. 5; Tables 3). There are, however, two potential but not necessarily exclusive biological explanations for this. First, reproductive isolation encouraged by adaptive divergence to water colour may have enhanced barriers to neutral gene flow enabling genome-wide divergence via genetic drift (i.e. incipient 'ecological speciation', Mayr 1963; Funk 1998; Schluter 2000; Rundle & Nosil 2005; Nosil *et al.* 2008). Supporting this inference is our mtDNA data that show a relationship between haplotype and water colour, irrespective of distance or river system (Fig. 2). Because it is generally thought that the mitochondrial and nuclear genomes are not linked to one another, this relationship has likely arisen via restricted gene flow in concert with increasing adaptive divergence between ecotypes. As such, the mtDNA pattern is in essence 'ecologically allopatric' and is probably responding independently between lineages via either genetic drift or selection within each environment (Via 2009; Thibert-Plante & Hendry 2010). The second explanation for the signal of neutral divergence is that selection may reduce gene flow in neutral regions of the genome that is physically linked, albeit weakly, to selected regions (Endler 1973; Lewontin & Krakauer 1973; Barton 2000). Because of linkage, the effect of natural selection on the genome is not discrete between selected and neutral loci. Rather, there is a spectrum between these two conditions. Using a genome scan, only those loci that are diverging from empirical or simulated neutral expectations are identified as outliers (Luikart *et al.* 2003; Bonin *et al.* 2006). While we have not quantified this within our data, it has been suggested that up to 10% of neutral loci exhibiting such isolation may be weakly linked to selected loci (Charlesworth *et al.* 1997; Nosil *et al.* 2008).

Genetic isolation driven by adaptations to environments (herein referred to as IBE) however is a function of time and the stochastic nature of genetic drift (Nosil *et al.* 2008). Because heterogeneous genomic divergence is particularly prevalent during population divergence and speciation, genetic differentiation will accumulate in some selected regions, while the homogenizing effect of gene flow may preclude divergence in other nonselected regions (Wu 2001; Gavrilets & Vose 2005; Nosil *et al.* 2009). In this way, IBE will often vary between selective gradients depending on time because divergence and factors that affect drift such as stochasticity and gene flow. We observed variation in the extent of IBE in our putatively neutral nuclear data set. After outlier loci were removed, some degree of population structure between water colours was still maintained but was considerably reduced (Table 4). A marked reduction of population structure was observed at the interface of the black waters of the Negro and white waters of the Amazon River (F_{ST}: total loci = 0.106, putatively neutral loci = 0.62). A similar, but less pronounced reduction in F_{ST} was also observed between the Amazon and Tapajós rivers after removal of outlier loci. Most affected, however, were the results of Structure analysis between total and putatively neutral data sets (Fig. 5). Here, there was an almost entire elimination of the population boundary between the Negro and Amazon Rivers, and increased admixture within the Tapajós River after the removal of outlier loci. Considering that the signal of IBE is a product of time and may be weakened with gene flow (Wu 2001; Gavrilets & Vose 2005), it is reasonable to infer that divergent natural selection between Andean- and craton-born ecotypes is relatively recent, a conclusion also supported by the phylogeographic scenario of divergence outlined above.

The variable degree of IBE observed between the Branco, Negro and Tapajós River populations is consistent with the corresponding gene flow estimates and geographic locations. This is because high levels of gene flow can erase the effects of adaptive divergence (Saint-Laurent *et al.* 2003; Hendry & Taylor 2004; Crispo *et al.* 2006), while spatial scales greater than that at which gene flow can occur enable neutral divergence to persist via drift (Nosil *et al.* 2009). The Branco River, which exhibits the most IBE (Fig. 5), is also the most geographically isolated from the Amazon River (Fig. 1). This isolation probably escapes the homogenizing effects of gene flow, enabling neutral genetic divergence to accumulate faster throughout the nuclear genome. Interestingly however, Branco River has a relatively high number of nonrepeat outliers in pairwise comparisons to the Negro River. This may indeed be a product of the seasonal hydrochemical changes that occur in the

Branco River. The Negro River in contrast exhibits the least IBE. Samples from the Negro River were collected near the meeting of Amazon waters, and there are numerous channels and small rivers connecting these systems. Thus, adaptive divergence here has probably not yet entirely isolated black and white water ecotypes to a point that it overcomes the homogenizing influence of gene flow. The Tapajós River actually exhibits an intermediate condition of IBE. Some gene flow was detected from Tapajós into the Amazon, but not in the reverse (Fig. 3), enabling neutral divergence to accumulate more rapidly than at the Negro River site. However, total IBE is not seen here because the geographic proximity of our Tapajós sampling site to the Amazon River affords reduced isolation compared with the Branco River. Thus, within *T. albus*, we have actually identified a spectrum of IBE within the nuclear genome, ranging from near-complete isolation of neutral loci in the Branco River to practically no isolation of neutral loci in populations of the Negro River. The significance of this spectrum is that variable perspectives of IBE within the one species can actually reveal more about the process of divergent natural selection between water colours than comparisons between purely reproductively isolated ecotypes (Via 2009).

Conclusions and perspectives

The role of divergent natural selection, as proposed by Darwin, in The Origin of Species is becoming increasingly accepted. Indeed, an assortment of conditions that enable ecological speciation in the face of gene flow is now well described (Rice & Hostert 1993; Via 2001). These generally include strong ecologically based selection against migrants and/or hybrids and divergent selection on multiple traits coupled with resource and habitat use (Via 2009). However, spatially defined models of speciation, in which allopatric divergence is the null, still remain dominant in theses on speciation (Coyne & Orr 2004). We presented a puzzling phylogeographic pattern for *T. albus* in which nuclear and mitochondrial genetic relationships are not correlated with distance or the spatial arrangement of rivers. Rather, genotypes appear to be associated with water colour, irrespective of the system they are in, a pattern consistent with genetic isolation resulting from ecological adaptation. Based on phylogeographic analyses and the current spatial distribution of genotypes in this system, we find that a simple vicariant model of divergence is not sufficient to describe the evolutionary history of *T. albus* in the Amazon Basin. Using genome scans, we identified powerful selection occurring between the craton-derived waters (black and clear) and the Andean-derived waters (white). To ascertain the precise selective mechanisms that isolate white water *T. albus* from black and clear water *T. albus* however, complementary research approaches that involve ecological and behavioural metrics, gene mapping and the identification of functionally important genes using next-generation sequencing technology would be invaluable.

Tropical regions like the Amazon Basin exemplify some of the most valuable yet threatened and understudied ecosystems on Earth (Lopez-Osorio & Miranda-Esquivel 2010). Population genetic surveys aimed at describing the distribution of adaptive genetic variation over the landscape and identifying ecological gradients are critically important in such regions. After all, in the face of ongoing environmental and climate change, these are the types of studies that will prove vital for assessing the potential of populations to persist in their current state.

Acknowledgements

The authors thank C. Moritz and A. Hendry for their helpful comments on an earlier version of the manuscript. This study was funded by the Discovery Program of the Australian Research Council (ARC grant DP0556496 to L. Beheregaray) and by Macquarie University through a postgraduate travel grant and research award to G. Cooke. Logistics and local arrangements were supported in part through the Brazilian National Council of Research and Technology CNPq-SEAP No. 408782/2006-4 Collection permit is under IBAMA No. 1920550, and ethical approval was received from Macquarie University, Approval number: 2007/033 under G. Cooke.

References

Achere V, Favre JM, Besnard G, Jeandroz S (2005) Genomic organization of molecular differentiation in Norway spruce (*Picea abies*). *Molecular Ecology*, **14**, 3191–3201.

Barton NH (2000) Genetic hitchhiking. *Philosophical Transactions of the Royal Society of London Series B – Biological Sciences*, **355**, 1553–1562.

Barton N, Bengtsson BO (1986) The barrier to genetic exchange between hybridizing populations. *Heredity*, **57**, 357–376.

Beaumont MA (2005) Adaptation and speciation: what can F-st tell us? *Trends in Ecology & Evolution*, **20**, 435–440.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London Series B – Biological Sciences*, **263**, 1619–1626.

Beheregaray LB (2008) Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology*, **17**, 3754–3774.

Beheregaray LB, Sunnucks P (2001) Fine-scale genetic structure, estuarine colonization and incipient speciation in

the marine silverside fish *Odontesthes argentinensis*. *Molecular Ecology*, **10**, 2849–2866.

Bermingham E, Martin AP (1998) Comparative mtDNA phylogeography of Neotropical freshwater fishes: testing shared history to infer the evolutionary landscape of lower Central America. *Molecular Ecology*, **7**, 499–517.

Bermingham E, McCafferty SS, Martin AP (1997) Fish biogeography and molecular clocks: perspectives from the Panamanian isthmus. In: *Molecular Systematics of Fishes* (eds Kocher TD and Stepien CA), pp. 113–128. Academic Press, San Diego.

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 of London Series B – Biological Sciences*, **365**, 1783–1800.

Berner D, Grandchamp A-C, Hendry AP (2009) Variable progress toward ecological speciation in parapatry: stickleback across eight lake-stream transitions. *Evolution*, **63**, 1740–1753.

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.

Bonin A, Ehrich D, Manel S (2007) Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Molecular Ecology*, **16**, 3737–3758.

Bradshaw HD, Schemske DW (2003) Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature*, **426**, 176–178.

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.

Campbell KE, Frailey CD, Romero-Pittman L (2006) The Pan-Amazonian Ucayali Peneplain, late Neogene sedimentation in Amazonia, and the birth of the modern Amazon River system. *Palaeogeography Palaeoclimatology Palaeoecology*, **239**, 166–219.

Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, **70**, 155–174.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Cooke GM, Chao NL, Beheregaray LB (2012) Marine incursions, cryptic species and ecological diversification in Amazonia: the biogeographic history of the croaker genus Plagioscion (Sciaenidae). *Journal of Biogeography*, doi: 10.111/i.1365-2699.2011.02635.x.

Cooper ML (2000) Random amplified polymorphic DNA analysis of southern brown bandicoot (*Isoodon obesulus*) populations in Western Australia reveals genetic differentiation related to environmental variables. *Molecular Ecology*, **9**, 469–479.

Cope ED (1872) On the fishes of the Ambyiacu River. *Proceedings of the National Academy of Sciences of Philadelphia*, **23**, 250–294.

Corrigan S, Huveneers C, Schwartz TS, Harcourt RG, Beheregaray LB (2008) Genetic and reproductive evidence for two species of ornate wobbegong shark *Orectolobus* spp. on the Australian east coast. *Journal of Fish Biology*, **73**, 1662–1675.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, MA.

Crispo E, Bentzen P, Reznick DN, Kinnison MT, Hendry AP (2006) The relative influence of natural selection and geography on gene flow in guppies. *Molecular Ecology*, **15**, 49–62.

Darwin C (1859) *On the Origin of Species*, 6th edn. Murray, London.

Emelianov I, Marec F, Mallet J (2004) Genomic evidence for divergence with gene flow in host races of the larch budmoth. *Proceedings of the Royal Society of London Series B – Biological Sciences*, **271**, 97–105.

Endler JA (1973) Gene flow and population differentiation. *Science*, **179**, 243–250.

Excoffier L, Laval LG, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure: extensions to linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Figueiredo J, Hoorn C, van der Ven P, Soares E (2009) Late Miocene onset of the Amazon River and the Amazon deep-sea fan: evidence from the Foz do Amazonas Basin. *Geology*, **37**, 619–622.

Fink WL, Fink SV (1973) Central Amazonia and its fishes. *Comparative Biochemistry and Physiology Part A: Physiology*, **62**, 13–29.

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.

Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Funk DJ (1998) Isolating a role for natural selection in speciation: host adaptation and sexual isolation in *Neochlamisus bebbianae* leaf beetles. *Evolution*, **52**, 1744–1759.

Garcia-Paris M, Good DA, Parra-Olea G, Wake DB (2000) Biodiversity of Costa Rican salamanders: implications of high levels of genetic differentiation and phylogeographic structure for species formation. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 1640–1647.

Gardner TA, Barlow J, Chazdon R *et al.* (2009) Prospects for tropical forest biodiversity in a human-modified world. *Ecology Letters*, **12**, 561–582.

Gavrilets S, Vose A (2005) Dynamic patterns of adaptive radiation. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 18040–18045.

Henderson PA, Crampton WGR (1997) A comparison of fish diversity and abundance between nutrient-rich and nutrient-poor lakes in the Upper Amazon. *Journal of Tropical Ecology*, **13**, 175–198.

Hendry AP, Taylor EB (2004) How much of the variation in adaptive divergence can be explained by gene flow? An evaluation using lake-stream stickleback pairs. *Evolution*, **58**, 2319–2331.

Hey J (2005) On the number of New World founders: a population genetic portrait of the peopling of the Americas. *PLoS Biology*, **3**, 965–975.

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**, 747–760.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 2785–2790.

Hoorn C, Guerrero J, Sarmiento GA, Lorente MA (1995) Andean tectonics as a cause for changing drainage patterns in Miocene northern South America. *Geology*, **23**, 237–240.

Hoorn C, Roddaz M, Dino R *et al.* (2010a) The Amazon Craton and its influence on the past fluvial systems (Mesozoic-Cenozoic, Amazonia). In: *Amazonia: Landscape and Species Evolution, a Look Into the Past* (eds Hoorn C and Wesselingh F), pp. 103–122. Wiley-Blackwell, West Sussex.

Hoorn C, Wesselingh FP, ter Steege H *et al.* (2010b) Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*, **330**, 927–931.

Jeffreys H (1961) *Theory of Probability*. Clarendon Press, Oxford.

Jensen JD, Wong A, Aquadro CF (2007) Approaches for identifying targets of positive selection. *Trends in Genetics*, **23**, 568–577.

Junk WJ, Soares MG, Carvalho FM (1983) Distribution of fish species in a lake of the Amazon River floodplain near Manaus (Lago Camaleao), with special reference to extreme oxygen conditions. *Amazoniana*, **VII**, 397–431.

Knowles LL, Maddison WP (2002) Statistical phylogeography. *Molecular Ecology*, **11**, 2623–2635.

Kondrashov AS (2003) Accumulation of Dobzhansky-Muller incompatibilities within a spatially structured population. *Evolution*, **57**, 151–153.

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of theory of selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.

Lopez-Fernandez H, Winemiller KO, Honeycutt RL (2010) Multilocus phylogeny and rapid radiations in Neotropical cichlid fishes (Perciformes: Cichlidae: Cichlinae). *Molecular Phylogenetics and Evolution*, **55**, 1070–1086.

Lopez-Osorio F, Miranda-Esquivel DR (2010) A phylogenetic approach to conserving Amazonian biodiversity. *Conservation Biology*, **24**, 1359–1366.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.

Lundberg JC, Marshall LG, Guerrero J *et al.* (1998) The stage for Neotropical fish diversification: a history of tropical South American rivers. In: *Phylogeny and Classification of Neotropical Fishes* (eds Malabarba LR, Reis RE, Vari RP, Lucene ZMS and Lucena CAS), pp. 13–48. Edipucrs, Porto Alegre.

Lundberg JG, Sabaj Perez MH, Dahdul WM, Aguilera OA (2010) The Amazonian Neogene fish fauna. In: *Amazonia: Landscape and Species Evolution. A Look Into the Past* (eds Hoorn C and Wesselingh F), pp. 281–301. Wiley-Blackwell, West Sussex.

Lynch M, Milligan BG (1994) Analysis of population genetic structure with rapid markers. *Molecular Ecology*, **3**, 91–99.

Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics*, **10**, 565–577.

Malabarba MCSL (2004) Revision of the Neotropical genus *Triportheus* Cope, 1872 (Characiformes: Characidae). *Neotropical Ichthyology*, **2**, 167–204.

Mani GS, Clarke BC (1990) Mutational order: a major stochastic process in evolution. *Proceedings of the Royal Society of London Series B – Biological Sciences*, **240**, 29–37.

Mayr E (1963) *Animal Species and Evolution*. Harvard University Press, Cambridge, UK.

Moritz C, Patton JL, Schneider CJ, Smith TB (2000) Diversification of rainforest faunas: an integrated molecular approach. *Annual Review of Ecological Systematics*, **31**, 533–563.

Murray MC, Hare MP (2006) A genomic scan for divergent selection in a secondary contact zone between Atlantic and Gulf of Mexico oysters, *Crassostrea virginica*. *Molecular Ecology*, **15**, 4229–4242.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nosil P, Flaxman SM (2011) Conditions for mutation-order speciation. *Proceedings of the Royal Society of London Series B – Biological Sciences*, **278**, 399–407.

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.

Ogden R, Thorpe RS (2002) Molecular evidence for ecological speciation in tropical habitats. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 13612–13615.

Orr HA, Turelli M (2001) The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution*, **55**, 1085–1094.

Orti G, Meyer A (1997) The radiation of characiform fishes and the limits of resolution of mitochondrial ribosomal DNA sequences. *Systematic Biology*, **46**, 75–100.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Petry P, Bayley PB, Markle DF (2003) Relationships between fish assemblages, macrophytes and environmental gradients in the Amazon River floodplain. *Journal of Fish Biology*, **63**, 547–579.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 37–45.

Reis RE, Kullander SO, Ferraris CJ (2003) *Checklist of Freshwater Fishes of South and Central America*. Edipucrs, Porto Alegre.

Rice WR, Hostert EE (1993) Laboratory experiments on speciation – what have we learned in 40 years. *Evolution*, **47**, 1637–1653.

Rodgers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.

Rodriguez MA, Lewis WM (1997) Structure of fish assemblages along environmental gradients in floodplain lakes of the Orinoco River. *Ecological Monographs*, **67**, 109–128.

Rull V (2008) Speciation timing and Neotropical biodiversity: the Tertiary-Quaternary debate in the light of molecular phylogenetic evidence. *Molecular Ecology*, **17**, 2722–2729.

Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters*, **8**, 336–352.

Rundle HD, Nagel L, Boughman JW, Schluter D (2000) Natural selection and parallel speciation in sympatric sticklebacks. *Science*, **287**, 306–308.

Saint-Laurent R, Legault M, Bernatchez L (2003) Divergent selection maintains adaptive differentiation despite high gene flow between sympatric rainbow smelt ecotypes (*Osmerus mordax* Mitchill). *Molecular Ecology*, **12**, 315–330.

Saint-Paul U, Zuanon J, Correa MAV *et al.* (2000) Fish communities in central Amazonian white- and blackwater floodplains. *Environmental Biology of Fishes*, **57**, 235–250.

Schluter D (2000) *The Ecology of Adaptive Radiation*. Oxford University Press, Oxford.

Schluter D (2001) Ecology and the origin of species. *Trends in Ecology & Evolution*, **16**, 372–380.

Schluter D, Conte GL (2009) Genetics and ecological speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 9955–9962.

Schneider S, Excoffier L (1999) Estimation of demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites. *Genetics*, **152**, 1079–1089.

Schneider CJ, Smith TB, Larison B, Moritz C (1999) A test of alternative models of diversification in tropical rainforests: ecological gradients vs. rainforest refugia. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 13869–13873.

Scotti-Saintagne C, Mariette S, Porth I *et al.* (2004) Genome scanning for interspecific differentiation between two closely related oak species *Quercus robur L.* and *Q petraea* (Matt.) Liebl. *Genetics*, **168**, 1615–1626.

Sioli H (1984) *The Amazon Limnology and Landscape Ecology of a Mighty Tropical River and its Basin*. Dr. Junk Publisher, Dordrecht.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science*, **236**, 787–792.

Smith TB, Wayne RK, Girman DJ, Bruford MW (1997) A role for ecotones in generating rainforest biodiversity. *Science*, **276**, 1855–1857.

Smith TB, Schneider CJ, Holder K (2001) Refugial isolation versus ecological gradients. *Genetica*, **112**, 383–398.

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of multi-allele and multi-locus genetic microstructure. *Heredity*, **82**, 561–573.

Sunnucks P, Hales DF (1996) Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Molecular Biology and Evolution*, **13**, 510–524.

Swofford DL (2003) *PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates, Sunderland, MA.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.

Tejerina-Garro FL, Fortin R, Rodriguez MA (1998) Fish community structure in relation to environmental variation in floodplain lakes of the Araguaia River, Amazon Basin. *Environmental Biology of Fishes*, **51**, 399–410.

Thibert-Plante X, Hendry AP (2010) When can ecological speciation be detected with neutral loci? *Molecular Ecology*, **19**, 2301–2314.

Thorpe RS, Surget-Groba Y, Johansson H (2008) The relative importance of ecology and geographic isolation for speciation in anoles. *Philosophical Transactions of the Royal Society of London Series B – Biological Sciences*, **363**, 3071–3081.

Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, **3**, 1572–1578.

Vekemans X, Beauwens T, Lemaire M, Roldan-Ruiz I (2002) Data from amplified fragment length polymorphisms (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology*, **11**, 139–151.

Via S (1999) Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. *Evolution*, **53**, 1446–1457.

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 of the United States of America*, **106**, 9939–9946.

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 (1943) Isolation by distance. *Genetics*, **28**, 114–138.

Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.

Zayed A, Whitfield CW (2008) A genome-wide signature of positive selection in ancient and recent invasive expansions of the honey bee *Apis mellifera*. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 3421–3426.

Zenger KR, Stow AJ, Peddemors V, Briscoe DA, Harcourt RG (2006) Widespread utility of highly informative AFLP molecular markers across divergent shark species. *Journal of Heredity*, **97**, 607–611.

Zhivotovsky LA (1999) Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology*, **8**, 907–913.

This study forms part of G.C's PhD research supervised by L.B. and N.C.. The authors share similar research interests about understanding the processes underpinning the evolution of biodiversity.

Data accessibility

GenBank Accession numbers for DNA sequences: JQ622253-JQ62281. Sequence alignments for the 114 *T. albus* samples and AFLP data together with sampling locations have been deposited at Dryad: doi:10.5061/dryad.v5gp1233.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Summaries of the IMa marginal distribution values.

Table S2 The average log likelihood ratios from Structure analysis across multiple independent runs with *K* ranging from 1 to 10.

Table S3 Spatial autocorrelation analysis.

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