

Natural selection in the water: freshwater invasion and adaptation by water colour in the Amazonian pufferfish

G. M. COOKE*, N. L. CHAO† & L. B. BEHEREGARAY*‡

*Molecular Ecology Laboratory, Department of Biological Sciences, Macquarie University, Sydney, NSW, Australia

†Departamento de Ciências Pesqueiras, Universidade Federal do Amazonas, Manaus, Brazil

‡Molecular Ecology Laboratory, School of Biological Sciences, Flinders University, Adelaide, SA, Australia

Keywords:

adaptive divergence;
Amazonia;
Genome scan;
phylogeography;
population genomics.

Abstract

Natural selection and ecological adaptation are ultimately responsible for much of the origin of biodiversity. Yet, the identification of divergent natural selection has been hindered by the spatial complexity of natural systems, the difficulty in identifying genes under selection and their relationship to environment, and the confounding genomic effects of time. Here, we employed genome scans, population genetics and sequence-based phylogeographic methods to identify divergent natural selection on population boundaries in a freshwater invader, the Amazonian pufferfish, *Colomesus asellus*. We sampled extensively across markedly different hydrochemical settings in the Amazon Basin and use 'water colour' to test for ecological isolation. We distinguish the relative contribution of natural selection across hydrochemical gradients from biogeographic history in the origin and maintenance of population boundaries within a single species and across a complex ecosystem. We show that spatially distinct population structure generated by multiple forces (i.e. water colour and vicariant biogeographic history) can be identified if the confounding effects of genetic drift have not accumulated between selective populations. Our findings have repercussions for studies aimed at identifying engines of biodiversity and assessing their temporal progression in understudied and ecologically complex tropical ecosystems.

Introduction

Natural selection and its role in ecological adaptation is quintessential in understanding the origin of biodiversity. Yet, despite this, the role of ecological adaptation and divergence in the face of homogenizing gene flow remains controversial (Coyne & Orr, 2004; Bolnick & Fitzpatrick, 2007; Coyne, 2007; Nosil, 2008; Schluter, 2009). One reason for such controversy is the difficulty in disentangling ecological divergence from vicariant biogeographic history. This distinction usually requires a system in which the existence of an allopatric phase is very unlikely in the context of evolutionary history (Endler, 1982; Coyne, 2007; Nosil, 2008). In fact,

spatially defined speciation processes have dominated discussions until recently and allopatry has generally been used as the null (Coyne & Orr, 2004). Yet, many of the global centres of biodiversity are highly dynamic and poorly studied tropical ecosystems for which even large-scale vicariant events are under documented (Beheregaray, 2008). Although the role of ecological divergence in the accumulation of tropical diversity is becoming increasingly recognized (e.g. Smith *et al.*, 1997, 2001; Schneider *et al.*, 1999; Garcia-Paris *et al.*, 2000; Ogden & Thorpe, 2002; Cooke *et al.*, 2012a), the engines of biodiversity in the tropics remain controversial and largely identified within the context of allopatric isolation (for review see Moritz *et al.*, 2000; Hoorn *et al.*, 2010). Fortunately, recent advances in population genomics are enabling scientists to incorporate information from adaptive variation into ecological and evolutionary studies of ecologically important nonmodel species (Travers

Correspondence: Luciano B. Beheregaray, Molecular Ecology Laboratory, School of Biological Sciences, Flinders University, Adelaide, SA 5001, Australia. Tel.: +61 882 015 243; fax: +61 882 013 015; e-mail: luciano.beheregaray@flinders.edu.au

et al., 2007; Ellegren, 2008a,b; Mardis, 2008). This has opened up an avenue of research opportunities for surveys of ecological divergence in understudied, species-rich ecosystems.

One such advance is the 'genome scan' whereby loci associated with selected genomic regions either directly or via hitchhiking are identified (Jensen et al., 2007). This method is based on the understanding that genetic drift, inbreeding and migration have genome wide effects, whereas selection leaves signatures only at those loci that are adaptive to a particular condition (Luikart et al., 2003), a phenomenon known as 'heterogeneous genomic divergence' (reviewed in Nosil et al., 2009). Here, differentiation will accumulate in regions under selection, whereas in other regions, genetic drift will require extended periods of time to accumulate (Wu, 2001; Gavrilets & Vose, 2005; Nosil et al., 2009). Thus, examining heterogeneous genetic divergence via a 'genome scan' can be particularly informative in populations at an early stage of divergence. In such cases, loci associated with selected genomic regions (genomic islands) still reflect the evolutionary history of adaptive divergence (Via & West, 2008; Via, 2009). An F_{ST} -based genome scan (*sensu* Beaumont & Balding, 2004) can be used to assess the role of selection in the origin and maintenance of diversity across ecological gradients or divides (Nosil et al., 2009). Simultaneously, loci identified as putatively neutral can be employed to address more classical questions of population genetics (Luikart et al., 2003; Nosil et al., 2009).

Freshwater fishes within Amazonia offer an interesting model for hypothesis testing across the speciation spectrum. Fish diversity in the Neotropics is unrivalled, with around 18% of the Earth's known fish species concen-

trated in < 0.003% of the Earth's water (Vari & Malabarba, 1998; Albert et al., 2011). Within this system, fish dispersal depends on the complex formation and fusion of anabranching riverine environments. Because of this, tracing the chronology of palaeogeographic events has improved our understanding of how allopatric scenarios have shaped Amazonian fish diversity (e.g. Hubert & Renno, 2006; Ready et al., 2006; Hubert et al., 2007; Willis et al., 2007; Cooke et al., 2009, 2012b; Siström et al., 2009; Piggott et al., 2011). Yet, the Amazonian aquatic environment also sustains marked hydrochemical and environmental discontinuities 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; Crampton, 2011). Based largely on optical and sedimentary characteristics, three water types within the basin have been identified (Sioli, 1984): (i) white water (Andean origin), which is turbid in nature and characterized by large amounts of dissolved solids; (ii) clear water (drains Brazil's Precambrian shields), which is comparatively transparent containing low content of dissolved solids and (iii) black water, which is transparent yet stained by tannins and humic acids leached from vegetation, differing most dramatically from the latter by its low pH (Fig. 1; Table 1). Considering that ecological gradients or divides can influence population structure via the action of divergent natural selection (Endler, 1977; Beheregaray & Sunnucks, 2001; Hendry, 2004; Nosil et al., 2005; Crispo & Chapman, 2008), hydrochemical discontinuities may therefore be a powerful force in the accumulation of Amazonian fish diversity.

Here, we employ a combination of genome scans, population genetics and sequence-based phylogeographic



Fig. 1 Sampling localities of *Colomesus asellus* in the Amazon Basin. Each site is distinguished by a unique colour and site label. Inset (i) shows the meeting of the black waters of the Negro River with the white waters of the Amazon River, (ii) shows the meeting of the Amazon River with the clear waters of the Tapajós River and (iii) shows the sampling area within northern South America.

Table 1 Water 'colour', sampling locations, sample size, and average hydrochemical variables collected from 2005 to 2008 (turbidity, cm; temperature, °C; pH; dissolved oxygen (mg/L), OD; oxygen saturation, O₂ %).

River	Site	Coordinates	<i>n</i>	cm	°C	pH	OD	O ₂ %
Black								
Negro	N1	3°4'44.00" S/60°14'44.00" W	3	76.0	29.7	5.2	6.4	82.3
White								
Madeira	M1	3°28'14.00" S/58°52'5.00" W	28	5.5	29.7	7.1	5.7	82.4
Amazon	A1	3°20'40.00" S/60°7'10.00" W	31	12.3	28.8	7.2	6.7	86.3
Amazon	A2	3°6'56.00" S/59°32'19.00" W	28	18.8	29.6	7.1	6.3	85.2
Amazon	A3	3°4'39.00" S/58°13'13.00" W	43	18.3	28.7	7.1	4.8	84.0
Amazon	A4	2°33'7.00" S/57°1'59.00" W	40	10.5	29.2	7.2	6.4	85.6
Amazon	A5	2°10'21.00" S/54°58'21.00" W	45	12.5	29.0	7.2	6.3	82.0
Amazon	A6	2°28'10.00" S/54°30'5.00" W	52	15	29.7	7.2	6.6	87.9
Clear								
Rio Tapajós	T1	2°52'17.00" S/55°9'38.00" W	21	118	29.5	6.7	7.1	97.5

methods to assess the relative roles of divergent natural selection and vicariant biogeographic history in a freshwater invader, the Amazonian pufferfish, *Colomesus asellus* (Tyler, 1964). The Amazonian puffer is distributed throughout the central Amazon Basin and is relatively abundant in the shallow habitat of the white and clear water rivers; however, it has also been reported in the brackish waters of the Amazon estuary (Araujo-Lima, 1994; Monks, 2006). The sister species, *Colomesus psittacus*, is distributed around the north of South America in brackish and coastal waters. *Colomesus asellus* are thought to have split from *C. psittacus* and invaded South America during the Miocene marine transgressions in northern Venezuela associated with the northward-draining palaeolake, Lake Pebas [25–11 Million years ago (Ma)] (Lovejoy *et al.*, 1998; Monsch, 1998; Hoorn *et al.*, 2010; Yamanoue *et al.*, 2011). Our study system consists of four large river systems (Amazon, Negro, Tapajós and Madeira Rivers), three major hydrochemical settings (white, black and clear water), two ecotones (where black meets white and white meets clear water) and a 'control' (Madeira River) in which two river systems of the same water type meet (Fig. 1i, ii). Further, the white waters of the Madeira River intersect the white waters of the Amazon River mid-transect.

On the basis of the contemporary patterns of hydrochemistry within our study system, in concert with the putative biogeographic history of the puffer and the geomorphological history of the Amazon Basin, we have two predictions. Firstly, if the ancestor of *C. asellus* colonized South America via the northward-flowing Lake Pebas, we predict a west-to-east trajectory of colonization with signals of demographic and range expansion events associated with the modern formation of the Amazon River. Secondly, if water colour represents a strong mechanism of divergent natural selection, signatures of selection, identified by outlier loci, and overall population divergence should be greater between selective environments than within them. Here, we show evidence for both ecological divergence due to water

colour and historical vicariance powering population diversification in Amazonia. Using both genome scans and phylogeographic approaches, we identify the contribution of divergent natural selection and genetic drift in generating population boundaries within *C. asellus* over different time scales. We show that our research framework can address fundamental evolutionary questions about biodiversity in Amazonia, a highly complex and poorly studied ecosystem. Findings from this study also have implications on how we identify the engines of biodiversity in other species-rich, complex ecosystems.

Materials and methods

Sampling

Field expeditions in the Amazon Basin took place between January and February in 2005 and in 2008. We explored a vast area of approximate riverine distance of 2200 km that encompasses four major river systems and the three hydrochemical settings of the Amazon Basin. These included the Amazon River (white water), Negro River (black water), Madeira River (white water) and Tapajós River (clear water). The 290 puffers were sampled at nine sites (Fig. 1; Table 1), and from most populations, we obtained up to 25 individuals in each sampling year.

Fish were caught in sandy and shallow beaches using a seine net, euthanized, and muscle tissue was preserved in 95% ethanol. Geographic coordinates of sampling sites were recorded using a global positioning system, and temperature, pH, visibility, dissolved oxygen and oxygen saturation were measured at each site (Table 1). We also included geographically distant samples from Peru and Colombia for improving our phylogeographic analysis of the direction and timing of freshwater invasion and colonization of central Amazonas. Peruvian samples were collected from the western end of the species distribution (Amazon River, near Iquitos – an area representing the palaeolake Lake Pebas). Colombian

samples came from the Orinoco Basin (Cano Macera), the northern end of the species distribution.

Laboratory procedures

DNA was extracted using a salting-out method (Sunnucks & Hales, 1996) and data obtained from the mitochondrial (mtDNA) and nuclear genomes. The mtDNA adenosine triphosphatase subunits 6 and 8 (ATPase 6 and 8 genes) were amplified via polymerase chain reaction (PCR) using primers and conditions specified in Corrigan *et al.* (2008). PCR products were sequenced using BigDyeTM terminator conditions in an AB3730xl (Applied Biosystems, Foster City, CA, USA) sequencer.

Our nuclear data set was based on 460 polymorphic loci. These were amplified fragment length polymorphisms (AFLP) obtained using a modified protocol after Zenger *et al.* (2006). Here, restriction fragments were amplified in a single selective PCR using fluorescently terminally labelled *Eco*RI selective primer carrying three selective nucleotides (5'-GAC TGC GTA CCA ATT C + AGT or AAC-3') and unlabelled *Mse*I selective primers carrying four selective nucleotides (5'-GAT GAG TCC TGA GTA A + CAAC, CTGC, CAGC or CTTC-3'). Fragments were denatured and separated on an AB3130xl sequencer using Liz500 labelled 500 bp size standard.

The presence and absence of AFLP fragments at 1-bp intervals were determined using GENEMAPPER 4.0 (Applied Biosystems, Foster City, CA, USA), within a fragment of 50–500 bp. Peaks with < 50 relative fluorescent units were not scored, and every bin fragment position identified using GENEMAPPER was checked manually. AFLPSCORE 1.4 (Whitlock *et al.*, 2008) was employed to normalize and score our data within the predefined fragment locations as well as calculate mismatch error rates for each primer combination. AFLPSCORE determines optimal thresholds that both minimize genotyping error and maximize the numbers of retained loci using a mismatch error rate analysis (Whitlock *et al.*, 2008).

Phylogeography and demographic history

The mtDNA sequence data were aligned using SEQUENCHER 4.1 (Gene Codes Corporation, Ann Arbor, MI, USA), and genealogical relationships within and among sampled populations were investigated by constructing a haplotype network in Tcs (Templeton *et al.*, 1992; Clement *et al.*, 2000). Demographic history was assessed by computing mismatch distributions in ARLEQUIN3.01 (Excoffier *et al.*, 2005) for the entire ATPase 6 and 8 data set, and for three geographic regions: (i) populations before and including the confluence of the Negro River (A1, N1 and A1), (ii) populations after the confluence of the Negro River and before the confluence of the Tapajós River (M1, A3, A4 and A5), and (iii) populations after and including the Tapajós River (T1 and A6). Mismatch analysis tests agreement of the data set with a model of

demographic growth (Rogers & Harpending, 1992; Excoffier *et al.*, 2005), by calculating the estimator of time to expansion (τ) and the mutation parameter (θ) (Schneider & Excoffier, 1999). From this, we used $t = \tau/(2\mu)$ to estimate the timing of possible population expansions where t is given in generations and μ is mutation rate for ATPase 6 and 8. We used an ATPase 6 and 8 mutation rate of 1.4% per million years proposed for geminate fishes across the Panama Isthmus (Bermingham *et al.*, 1997). Additionally, Fu's (1997) test of selective neutrality was employed to assess the signal of expansion in the data set. In the event of demographic expansion, large negative F_S values are generally observed (Fu, 1997).

Detecting natural selection

We used both a frequentist method, DFDIST (Beaumont & Nichols, 1996), and a Bayesian method, BAYESCAN (Foll & Gaggiotti, 2008), to detect selection in the AFLP data. BAYESCAN was used to directly estimate the probability of each locus being subject to selection for the whole data set (<http://www-leca.ujf-grenoble.fr/logiciels.htm>). 'Outlier loci' were assessed based on the posterior distribution of α_i where a positive value is indicative of positive selection and a negative value indicates stabilizing selection. The significance of α_i for each locus is estimated in BAYESCAN using a Reversible Jump MCMC algorithm and loci were classified according to Jeffreys (1961) scale of evidence.

Secondly, we used DFDIST following the approach of Beaumont & Balding (2004) to detect 'outlier loci' in pairwise population comparisons. This methodology is based on the premise that loci under divergent selection will display significantly higher F_{ST} values than the majority of neutral loci in a sample. For each analysis, a critical frequency for the most common alleles of 0.98 was chosen and those loci with allele frequencies higher than 0.98 were excluded from the empirical distribution. Next, using a trimmed mean F_{ST} of 30%, 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. Using DFDIST, we searched for (positive) directional selection between nine adjacent population comparisons (see Fig. 3i). Pairwise comparisons were made between sites A1 and A2 (N1 samples were pooled with A2 samples because N1 = 3 fish collected near the Negro River outflow into the Amazon River), A2 and M1, M1 and A3, A2 and A3, A3 and A4, A4 and A5, A5 and A6, A5 and T1, and A6 and T1.

Loci detected as positive outliers were divided into two groups: (i) nonrepeat outliers that were detected in just one pairwise comparison and (ii) repeat outliers that were detected in multiple comparisons. The remaining loci that were not positive outliers in any comparison were classified as putatively neutral. To minimize the chances of type I errors, nonrepeat outliers were

considered potential false positive irrespective of their P -values. Thus, only repeat-outlier loci and loci detected using both methods were classified as loci under selection and consequently removed from the total data set for subsequent data analyses.

Summary statistics of population diversity and genetic structure

Sequence (mtDNA) diversity was estimated as haplotypic diversity and nucleotide diversity (Nei, 1987) per population in ARLEQUIN 3.11 (Excoffier *et al.*, 2005). AFLP 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 equilibrium. The number and proportion of polymorphic loci, expected heterozygosity and average gene diversity were calculated, with the unbiased estimator of Lynch & Milligan (1994).

Genetic structure between adjacent populations was calculated for mtDNA (Φ_{ST} : Excoffier *et al.*, 1992) in ARLEQUIN, and for both the total and the neutral AFLP (F_{ST} : Lynch & Milligan, 1994) data sets in AFLP-SURV (Vekemans, 2002). Due to the dominant nature of AFLPs, the inbreeding coefficient (f) cannot be calculated directly. Therefore, an alternative Bayesian approach was employed to estimate an F_{ST} analogue (Θ^B) for dominant markers that accounts for uncertainty in f (Holsinger *et al.*, 2002). Using the program HICKORY 1.1 (Holsinger & Lewis, 2005), models were fitted to the data under three hypotheses: (i) both f and Θ^B are unknown and ≥ 0 (full model); (ii) $f = 0$ but Θ^B is unknown ($f = 0$ model); (iii) f is unknown and there is no genetic structure ($\Theta^B = 0$ model). Depending on the model, f , Θ^B or both are estimated. A uniform prior distribution of f was assumed for both the ‘full’ and ‘ $\Theta^B = 0$ ’ models. Default HICKORY values for sampling and chain length parameters were used (burn-in = 50 000, sample = 250 000, thin = 50), and model selection was based on the deviance information criterion (DIC; see Holsinger & Wallace, 2004).

The probability of connectivity between populations is likely to be influenced by the unidirectional river flow and geographic distance (McGlashen *et al.*, 2001). Therefore, we tested for the association between genetic (Φ_{ST}) and geographic distance (Isolation by distance, IBD; Wright, 1943) in the mtDNA and AFLP data sets for (i) all sampled populations and (ii) sampled populations along the Amazon River (A1–A6). The significance of the association was tested using a Mantel permutation test (Mantel, 1967) in ARLEQUIN 3.1 (Excoffier *et al.*, 2005) for mtDNA and GENALX 6.1 (Peakall & Smouse, 2006) for AFLPs.

The degree of population structure was further explored with a hierarchical analysis of molecular variance (AMOVA) using ARLEQUIN for mtDNA (Excoffier *et al.*, 1992) and GENALX for neutral AFLP (Peakall & Smouse, 2006). Two AMOVAs were ran for each data set:

(i) populations before and including the confluence of the Negro River (A1, N1 and A1), (ii) populations after the confluence of the Negro River and before the confluence of the Tapajós River (M1, A3, A4 and A5), and (iii) Populations after and including the Tapajós River (T1 and A6). The second AMOVA was conducted to control for the confluence of a river system into the Amazon River by looking for structure associated with the white water Madeira River. Using only the populations situated along the Amazon River between the Negro and Tapajós, we created two white water groups: (i) the Madeira and (ii) Amazonas sites A2–A5.

Genetic subdivision was also investigated using the nonparametric, model-based clustering method implemented in STRUCTURE v.2.3.1, which accommodates AFLP data (Falush *et al.*, 2003, 2007). Using the admixture model, we determined the number of populations, K , by comparing log-likelihood ratios across three independent runs by varying the assumed number of K for the total AFLP and neutral AFLP data sets (Evanno *et al.*, 2005). Each run consisted of a burn-in phase of 100 000 iterations, followed by 1 000 000 iterations.

Results

We generated 797 bp of mtDNA ATPase 6 and 8 genes for 238 samples including Colombia and Peruvian ‘outgroups’. Of these, 88 unique haplotypes were identified (GenBank accession numbers JQ818261–JQ818350). These were composed of 90 variable characters, of which 42 were parsimony informative. AFLP profiles were resolved for 290 individuals, with a total of 460 polymorphic loci (Dryad repository doi:10.5061/dryad.m6k89833). Mismatch error rate calculated by AFLPScore was on average 5% ($\pm 2.9\%$) per primer combination. This is in the lower threshold of the acceptable range of 5–10% AFLP error rates (Bonin *et al.*, 2007).

Phylogeography and demographic history

A strong phylogeographic signal of haplotypes clustered in an easterly (upstream to downstream) pattern was generally observed in the mtDNA haplotype network (Fig. 2). Indeed, within the network there appeared to be three predominant groups. The first consists largely of westernmost sampled haplotypes that predominantly originate west of the confluence of the Negro and Amazon River. This group also includes the westernmost sample from Iquitos (Peru) and the north-western sample from the Orinoco Basin (Colombia). The second grouping is arranged in a star shape with many recently derived haplotypes. It is composed mostly of fish from the central sites in the Amazon and Madeira Rivers. Moving downstream, the final grouping is dominated by fish from the Tapajós River and the two easternmost Amazon River sites.

Results from mismatch analysis and Fu’s test of neutrality recovered important insights into the

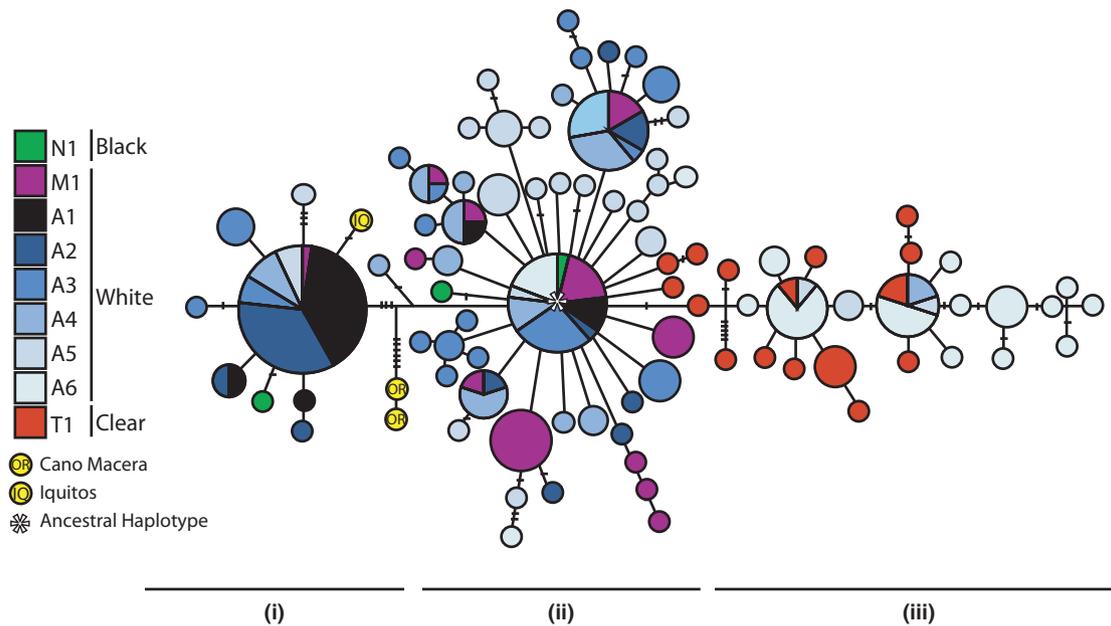


Fig. 2 Statistical parsimony network for mtDNA ATPase 6 and 8 haplotypes. Relationships among haplotypes are estimated using the parsimony method of Templeton *et al.* (1992). Each circle denotes a unique haplotype and the area of the circle is proportional to its frequency. The colour/s of the circle represents the sampling locality as in Fig. 1 and the lines along the bottom of the network refer to 'haplogroups' 1–3.

Table 2 Demographic analysis for the total data set and for all the data and for (i) populations before and including the confluence of the Negro River (A1, N1 and A2), (ii) Populations after the confluence of the Negro River, and before the confluence of the Tapajós River (M1, A3, A4, A5) and (iii), populations after and including the Tapajós River (T1 and A6). Results are summarized by the Sum of Squared Deviations (SSD), Raggedness Index (R), and Fu's test of neutrality (F_S). Estimates of time since expansion are shown for analyses exhibiting evidence of demographic expansion.

	SSD	P	R	P	F_S	P	Time since expansion	Range ($\alpha = 0.05$)
All Data	0.007	0.54	0.012	0.681	-24.898	0.000	~3 Ma	1 to 4.3 My
(i)	0.132	0.371	0.197	0.058	0.368	0.593	NA	NA
(ii)	0.017	0.349	0.028	0.476	-25.617	0.000	~2.5 Ma	500 ka to 4 My
(iii)	0.005	0.708	0.018	0.795	-10.243	0.001	~1.5 Ma	500 ka to 2.8 My

demographic history of *C. asellus* (Table 2) consistent with the west-to-east trajectory of colonization inferred from the haplotype network (Fig. 2). Mismatch analysis of our westernmost group did not fit to the model of demographic expansion, a result typical of older populations (Rogers & Harpending, 1992). Further, Fu's test of neutrality gave a positive F_S value. In marked contrast, mismatch analysis of our central and eastern haplogroups did not deviate from a model expected under demographic expansion, a result corroborated by large negative F_S values. While date estimates calculated using this method must be interpreted with caution, the estimated τ values suggest the demographic expansion occurred first in the central Amazonian sites (approximately 2.5 Ma) and later in the most easterly ones (approximately 1.5 Ma). For the entire data set, there was a nonsignificant deviation from a model expected under demographic expansion, and a large negative F_S value was found. Estimated τ values

suggest that this demographic expansion occurred approximately 3 Ma.

Detecting natural selection and population structure

A total of 54 outlier AFLP loci (11.7% of the total) were detected across eight of the nine pairwise comparisons at the 99% quantile (Fig. 3i; Table 5). Of these, 16 (3.5%) were repeat-outlier loci. Of these repeat-outlier loci, the majority were located in the more easterly populations from A4 to A6 and T1 (Fig. 3), near the confluence of the clear waters of the Tapajós and white waters of the Negro River. Fifteen of these repeat loci were detected in population comparisons including A6. In contrast, only four repeat loci were found west of A3, and none were found in A1.

The hierarchical Bayesian approach of BAYESCAN performed on all data identified 166 loci with positive α_i values. Applying Jeffreys (1961) scale of evidence, six

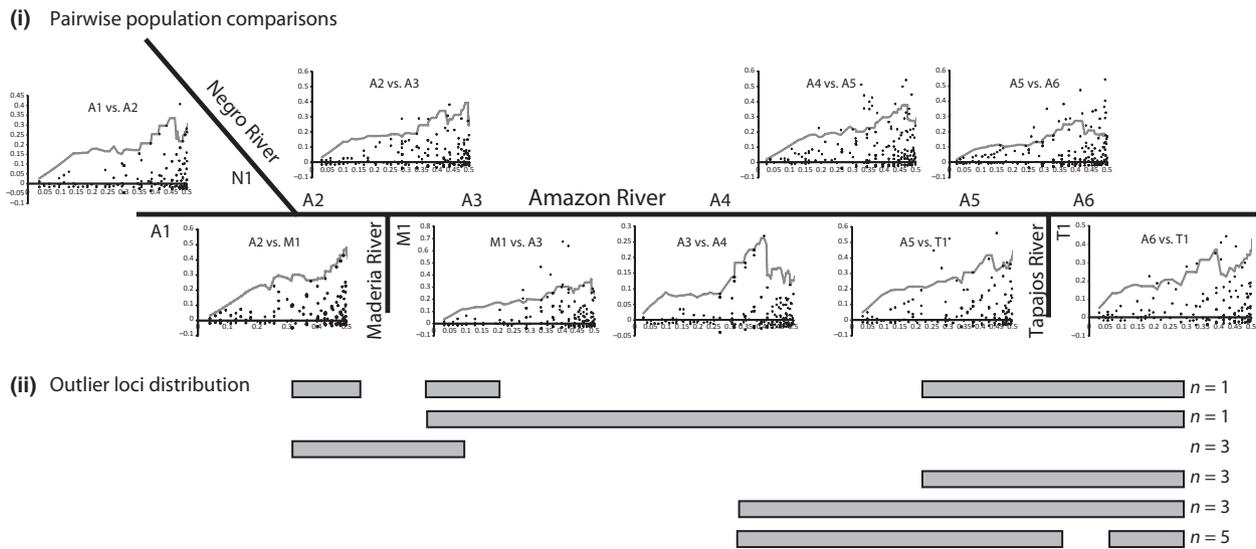


Fig. 3 Results from DFDIST analysis: (i) the empirical distributions of F_{ST} values for amplified fragment length polymorphisms (AFLP) loci for each population pairwise comparison overlain on a schematic representation of the sampling localities. The x-axes are heterozygosity and the y-axes are F_{ST} . The solid line represents the 99th quantile, and dots exceeding the 99th quantile are ‘outlier loci’, (ii) a schematic representation of the distribution of repeat-outlier loci between sample sites. Each bar represents the area shared by n loci.

Table 3 Genetic diversity at mtDNA and putatively neutral amplified fragment length polymorphisms (AFLP) markers. Mitochondrial diversity is summarized by the number of haplotypes (N), haplotypic diversity (h) and nucleotide diversity (π). AFLP diversity is summarized by the number of polymorphic loci per population (NPL) and expected heterozygosity (H_e).

Site	mtDNA diversity				AFLP diversity		
	n	N	h	π	n	NPL	H_e
N1	3	3	1 (0.272)	0.008 (0.007)	3	51	0.136
A1	23	5	0.451 (0.121)	0.003 (0.002)	31	267	0.269
A2	27	11	0.692 (0.097)	0.005 (0.003)	28	223	0.230
M1	28	11	0.855 (0.046)	0.003 (0.002)	27	252	0.240
A3	33	18	0.934 (0.027)	0.005 (0.003)	43	290	0.257
A4	29	12	0.921 (0.026)	0.005 (0.003)	40	292	0.261
A5	35	23	0.961 (0.018)	0.006 (0.003)	45	313	0.264
A6	35	15	0.918 (0.025)	0.007 (0.004)	52	244	0.218
T1	21	17	0.967 (0.030)	0.007 (0.004)	21	186	0.192
Total	238	88		Total	290	444	

of these loci that were supported with \log_{10} (Bayes Factor) > 2 can be considered as ‘decisive’ evidence for selection, two that were supported \log_{10} (Bayes Factor) > 1.5 can be considered ‘very strong’ evidence for selection, three that were supported \log_{10} (Bayes Factor) > 1 can be considered ‘strong’ evidence for selection and five that were supported with \log_{10} (Bayes Factor) > 0.5 can be considered ‘substantial’ evidence for selection. Of the 54 outliers detected using DFDIST, 20 of these were also detected using BAYESCAN. After removing repeat outliers, our data set consisted of 444 putatively neutral polymorphic loci.

For the mtDNA sequence data, nucleotide diversity ranged from 0.003 to 0.007 and haplotypic diversity from 0.451 to 0.967 (excluding N1 for which there were only 3 samples). From our putatively neutral AFLP loci, expected heterozygosity ranged between 0.192 and 0.269 (also excluding N1; Table 3).

High global AFLP population structure was observed across puffer populations ($\Theta^B = 0.439055$, Table 4). The

Table 4 Amplified fragment length polymorphisms data fitted to models of population structure using the Bayesian method in HICKORY. Parameters estimated are the inbreeding coefficient, f , and the F_{ST} analogue Θ^B . Dbar is a measure of how well the model fits the data, pD is the number of parameters being estimated, and deviance information criterion (DIC) is the model selection criteria.

Model	F		Θ^B		Dbar	pD	DIC
	Mean (SD)	95% CI	Mean (SD)	95% CI			
Full*	0.135 (0.023)	0.092–0.183	0.439 (0.013)	0.414–0.464	10185.3	2017.06	12202.4
$f = 0$	–	–	0.413 (0.012)	0.389–0.438	10152.6	2127.72	12280.3
$\Theta^B = 0$	0.057 (0.019)	0.022–0.097	0.370 (0.014)	0.343–0.398	19160.3	424.459	19584.8

*The optimal model selected according to this criterion.

Table 5 Number of outlier and repeat-outlier amplified fragment length polymorphisms (AFLP) loci found in each pairwise comparison, the total and neutral AFLP F_{ST} values, and the mtDNA Φ_{ST} value.

Comparison	No. 99% Outliers	No. repeat outliers	Total F_{ST}	Neutral F_{ST}	mtDNA Φ_{ST}
A1 vs. A2	2	0	0.056*	0.056*	0.047*
A2 vs. A3	6	4	0.054*	0.049*	0.232*
A2 vs. M1	0	0	0.049*	0.041*	0.406*
M1 vs. A3	9	3	0.061*	0.047*	0.082*
A3 vs. A4	1	1	0.008	0.007	0.016
A4 vs. A5	18	9	0.074*	0.059*	0.010
A5 vs. A6	24	8	0.064*	0.055*	0.211*
A5 vs. T1	5	4	0.080*	0.069*	0.270*
T1 vs. A6	9	7	0.069*	0.059*	0.064*

* $P \leq 0.05$.

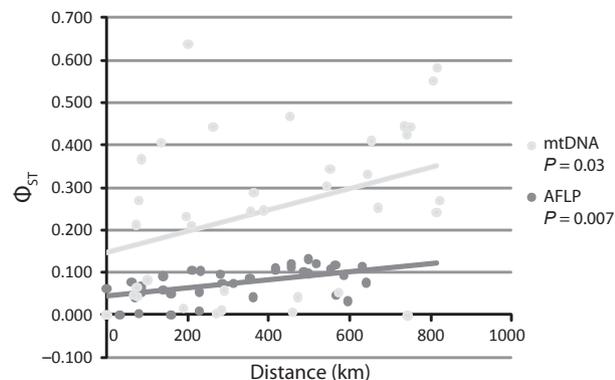


Fig. 4 Results from the Mantel test for both amplified fragment length polymorphisms (AFLP) data and mtDNA ATPase 6 and 8, based on pairwise population genetic differentiation (Φ_{ST}) and geographic distance (km).

full model, which assumes both population structure and $f \geq 0$, fitted our data best having the lowest DIC value (DIC = 12202.4). A strong signal of pairwise population

genetic structure was observed for both F_{ST} and Φ_{ST} comparisons, indicating relatively low levels of dispersal between populations. Pairwise comparisons that fall at the confluence of either the Negro or Tapajós Rivers into the Amazon River had higher F_{ST}/Φ_{ST} values (e.g. A2 vs. M1; $\Phi_{ST} = 0.406$, $F_{ST} = 0.041$; A5 vs. T1, $\Phi_{ST} = 0.270$, $F_{ST} = 0.068$). In contrast, pairwise comparisons between white water sites, not interrupted by the confluence of the Negro or Tapajós Rivers, generally had lower values (e.g. A3 vs. A4 $\Phi_{ST} = 0.016$, $F_{ST} = 0.007$; Table 5). In general, neutral pairwise F_{ST} values were lower than total F_{ST} values, most notably in easterly population comparisons.

Mantel tests revealed a positive correlation between all sampled populations and geographic distance for both mtDNA ($r = 0.357$, $P = 0.03$) and neutral AFLP data ($r = 0.295$, $P = 0.007$; Fig. 4). Further, Mantel tests of populations along the Amazon River only (excluding tributary samples) had a positive correlation for AFLP data ($r = 0.505$, $P = 0.001$). The latter was marginally nonsignificant for mtDNA data ($r = 0.361$, $P = 0.07$), a result potentially influenced by the small number of sample comparisons.

The AMOVAS for both mtDNA and AFLP identified consistent patterns of regional population structure associated with water colour (Table 6). AMOVA 1, which tested for genetic structure associated with ecotones of the Negro and Tapajós Rivers, was significant between regions in both cases (mtDNA: $\Phi_{CT} = 0.3350$, $P \leq 0.000$, AFLP: $\Phi_{RT} = 0.03$, $P < 0.001$). In contrast, AMOVA 2, which tested for genetic structure associated with the confluence of the Madeira River into the Amazon River (i.e. white vs. white water), was not significant for either DNA data set (mtDNA: $\Phi_{CT} = -0.014$, $P = 0.398$, AFLP: $\Phi_{RT} = 0.4$, $P = 0.235$).

STRUCTURE analysis for the AFLP data sets ($n = 460$, includes neutral and outlier loci) identified a prominent phylogeographic break between A2 and sites downstream, coinciding with the confluence of the Negro River into the Amazon River (Fig. 5i, ii). Using the ad hoc statistic ΔK , this STRUCTURE analysis estimated the most likely number

Table 6 Analysis of molecular variance (AMOVA) for mtDNA and amplified fragment length polymorphisms (AFLP) data. In AMOVA 1, regions include: (i) populations before and including the confluence of the Negro River (A1, RN and A2), (ii). Populations after the confluence of the Negro River, and before the confluence of the Tapajós River (M1, A3, A4, A5) and (iii), populations after and including the Tapajós River (T1 and A6). In AMOVA 2 regions include: (i) the Madeira River, and (ii) white water sites A2–A5. FI stands for Fixation Index.

Data type	AMOVA 1				AMOVA 2			
	Source of variation	% Variation	FI	P	Source of variation	% Variation	FI	P
mtDNA	Among regions	33.50	Φ_{CT} : 0.3350	0.000	Among regions	0.00	Φ_{CT} : -0.014	0.398
	Among populations	3.29	Φ_{SC} : 0.0494	0.002	Among populations	15.36	Φ_{SC} : 0.1514	0.000
	Among individuals	63.21	Φ_{ST} : 0.3679	0.000	Among individuals	86.09	Φ_{ST} : 0.1391	0.000
AFLP	Among regions	3.00	Φ_{RT} : 0.030	0.000	Among regions	0.00	Φ_{RT} : 0.4	0.235
	Among populations	7.00	Φ_{PR} : 0.067	0.000	Among populations	8.00	Φ_{PR} : 0.081	0.000
	Among individuals	91.00	Φ_{PT} : 0.095	0.000	Among individuals	92.00	Φ_{PT} : 0.081	0.000

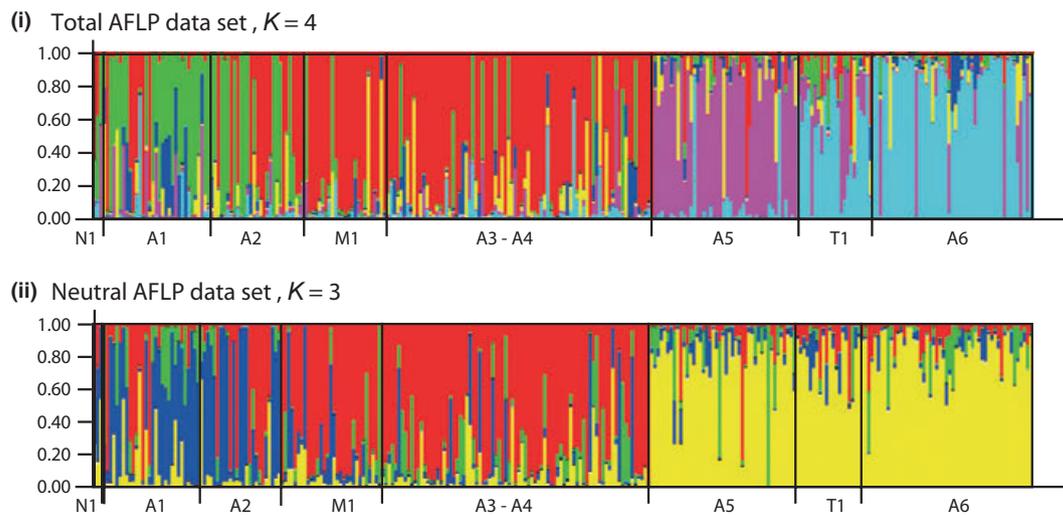


Fig. 5 STRUCTURE results of total amplified fragment length polymorphisms (AFLP) data set ($n = 460$) and the putatively neutral AFLP data set ($n = 444$). Individuals are grouped by sampling location, and each individual is represented by one vertical column. K = the number of inferred genetic populations for each data set.

of populations was four six [$\ln \Pr (X | K = 6) = -27827.7$], yet this did not resolve any further clusters from $K = 4$ [$\ln \Pr (X | K = 3) = -29243.2$]. Thus, we chose $K = 4$ as the most likely number of populations for our total AFLP data set (Fig. 5i), which coincides with our inferred phylogeographic breaks. Additionally, the break between A5 and T1, which coincided with the confluence of the Tapajós River, was also supported. An additional phylogeographic break was observed between white water sites A4 and A5. Such a break was not unexpected given the marked topographic feature that exists between these two sites, normally demarcated by the Obidos County. Here, the Amazon River channel is at its narrowest point in Brazilian territory (approximately 2.4 km), the current has high velocity water flow and the river channel is very deep (approximately 19 m; Mertes *et al.*, 1996). Because of this, the area between A4 and A5 is dominated by rocky beds instead of the preferred puffer habitat (i.e. sandy and shallow beaches), which is more typically found along margins of the Amazon River.

The STRUCTURE analysis based on putatively neutral AFLP loci ($n = 444$) and ΔK estimated that the most likely number of populations was six [$\ln \Pr (X | K = 6) = -27049.8$], yet this did not resolve any further clusters from $K = 3$ [$\ln \Pr (X | K = 3) = -28309.7$]. Thus, we chose $K = 3$ as the most likely number of populations for our putatively neutral loci. Again, the strong break downstream to the meeting of the Negro and Amazon (A2) was present. Removal of outlier loci appeared to have no effect on the presence of this barrier within our data set. Similar to our total AFLP data set (Fig. 5i), there was a strong phylogeographic break between A4 and A5. This barrier was also unaffected by the removal of outlier loci. Most interestingly however, the break between A5 and T1

was no longer present in our analysis when outlier loci were excluded from our data set. This suggests that any population structure generated in this region was a product of divergent selection acting on those loci. Finally, for both the total and the neutral AFLP data sets, our white water control sample (Madeira River) grouped with other typical white water samples from the Amazon River.

Discussion

This study provides evidence for the effect of divergent natural selection associated with water colour entwined in the biogeographic history of a freshwater invader, the Amazonian Puffer *C. asellus*. Using genome scans and phylogeographic approaches, we identified two major genetic breaks in *C. asellus* populations within the Amazon Basin. The first break occurs at the confluence of the black waters of the Negro River and the white waters of the Amazon River, whereas the second break occurs at the confluence of the Amazon River and the clear waters of the Tapajós River. Consistent with the prediction of a west-to-east trajectory of colonization associated with the last stages of formation of the Amazon River, we detected a strong signal of IBD with demographic expansion downstream. Evidence for divergent natural selection was detected most strongly around the confluence of the Tapajós into the Amazon River, yet there was reduced evidence for selection around the confluence of the Negro River. No evidence for selection was obtained when comparing populations from the same selective environment. Below, we discuss on the relative contribution of vicariant biogeographic history and then of natural selection in the origin and maintenance of population divergence.

A freshwater invasion and colonization

Although the precise timing of the Neogene marine incursions into South America is controversial, it has become generally accepted that a connection existed between the Upper Amazon and the Caribbean (for reviews see Lundberg *et al.*, 1998; Hoorn *et al.*, 2010). A growing body of multidisciplinary studies has provided strong evidence for such connection. This includes studies on fossilized marine fishes (Monsch, 1998), isotope data that confirm the presence of saline waters in the upper Amazon (Hoorn, 1993, 1994; Monsch, 1998; Vonhof *et al.*, 2003), and biogeographic and phylogenetic evidence for multiple freshwater invasions into Amazonia from marine ancestral lineages (Lovejoy *et al.*, 1998, 2006, 2010; Cooke *et al.*, 2012b). Of particular importance to the transition of marine to freshwater taxa has been the documentation of Lake Pebas, a semi-permanent (24–11 Ma) wetland complex in the upper Amazon. Lake Pebas was connected to the Caribbean and is thought to have behaved as a complex series of interconnected lakes of inconstant salinity that enabled the progression of taxa from marine to freshwater (Wesselingh *et al.*, 2002; Hoorn *et al.*, 2010).

The colonization of *C. asellus* into the freshwaters of South America via Lake Pebas during the Miocene is consistent not only with fossil data (Monsch, 1998), but also with our phylogeographic reconstruction. Our results strongly indicate a west-to-east trajectory of colonization along the modern Amazon River, a pattern supported by the genealogical distributions of matrilineal statistics of demographic expansion and Mantel tests

(Table 2, Figs 2 and 4). For example, mtDNA haplotypes generally cluster in a downstream pattern (west to east). In addition, there is evidence for sequential demographic expansions, one after the confluence of the Negro River approximately 2.5 Ma, and later, after the confluence of the Tapajós River approximately 1.5 Ma. Further, the more westerly (upstream) samples showed the genealogical architecture expected of older populations (Rogers & Harpending, 1992).

Although it is generally accepted that the modern drainage system of the Amazon River and its largest tributaries were established in the late Miocene (Hoorn *et al.*, 1995; Lundberg *et al.*, 1998), palaeogeographic work by Campbell *et al.* (2006) has dated the full establishment of the modern Amazon River drainage system to the late Pliocene approximately 2.5 Ma (Campbell *et al.*, 2006). Accordingly, the modern Amazon River may have formed as a compound response to the breaching of the eastern rim of the sedimentary basin known as the Içá Formation and erosion of the proto-Amazon River (Rossetti *et al.*, 2005; Campbell *et al.*, 2006). This is proposed to have occurred at the onset of the Plio-Pleistocene glacial climate cycle during the lowest sea level stand of the middle Miocene with frequent geomorphological changes that continued well into the Holocene (Latrubesse & Franzinelli, 2005; Rossetti *et al.*, 2005; Campbell *et al.*, 2006). Interestingly, our data show a strong phylogeographic break that parallels this region of geomorphological activity (between sites A2 and M1; Fig. 6iii). This is most evident in STRUCTURE analyses from the AFLP data (inclusive and exclusive of outlier loci; Fig. 5). Further, demographic

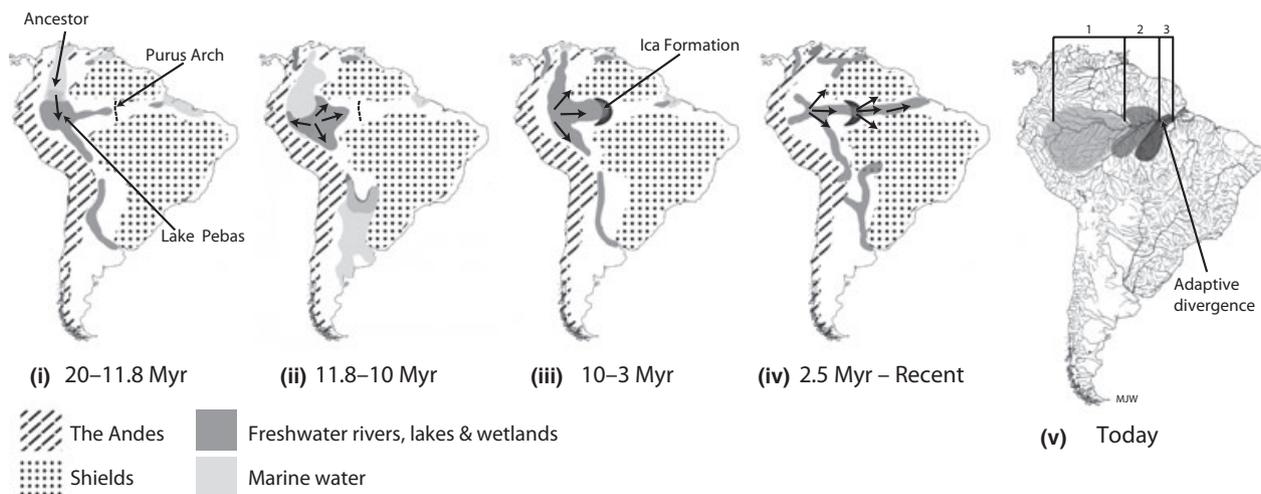


Fig. 6 Phylogeographic scenario depicting the invasion and colonization route of *Colomesus asellus* into South America: (i) Entry point into South America via the northward-flowing connection between Lake Pebas, the upper Amazon and the Caribbean, (ii) the evolution of freshwater tolerance encouraged by the dynamic saline/freshwater environment of Lake Pebas, (iii) termination of the marine incursions and the erosion of Içá Formation and colonization of western freshwater habitat (iv) the development of the modern transcontinental west-to-east flow and subsequent colonization of the Amazon River (v) contemporary population subdivision of *C. asellus* and the modern drainage systems of South America including adaptive divergence at the junction of the Amazon and Tapajós Rivers.

expansion within the haplogroup consisting predominantly of samples east of the Negro River, and west of the Tapajós River, has been dated to approximately 2.5 Ma.

On the basis of our results and of the large body of literature on palaeogeography mentioned above, we propose a convincing phylogeographic scenario for the freshwater invasion and colonization of *C. asellus* into the Amazon Basin (Fig. 6). Initiated by Miocene marine incursions, the puffer ancestor entered South America via the Caribbean and colonized the dynamic saline environment of the upper Amazon wetland system. The varying salinity levels and intermittent connectivity of Lake Pebas to the Caribbean would have encouraged the evolution of freshwater tolerance isolating an incipient freshwater lineage from sister marine lineages (Wesselingh *et al.*, 2002). Continuing until the late Miocene, *C. asellus* underwent range expansions, colonizing the available shallow freshwater habitats throughout western Amazonia and the early Andean foreland basin (west of the Purus Arch). These events likely extended into the Palaeo-Orinoco system in central Colombia via the northward-flowing Palaeo-Amazon (Hoorn *et al.*, 1995). Evidence for this early colonization can be seen in the relationship of the Orinoco samples with our older westerly haplotypes (Fig. 2). Thus, the colonization and expansions of puffers in the Palaeo-Amazon and Orinoco systems would have continued until the late Miocene after which the proto-Amazon River breached the Purus Arch. At that time, approximately 2.5 Ma, the modern west-to-east transcontinental flow of the Amazon River established and puffers subsequently colonized downstream.

Adaptation through natural selection

Following the establishment of the west-to-east-flowing Amazon River, the subsequent west-to-east IBD (Fig. 4) and genetic structure along the Amazon River and its tributaries could be explained by certain migratory constraints of *C. asellus* including its overall limited mobility and low larval dispersal (Araujo-Lima, 1994). Further, the complex anabranching Amazonian aquatic environment could have created numerous micro- and macro-allopatric events allowing genetic drift to accumulate between populations (e.g. between sites A4 and A5). However, our data, both the mitochondrial and nuclear DNA, consistently show significant phylogeographic breaks at the confluence of the black waters of the Negro River into the white waters of the Amazon River and after the confluence of the clear waters of the Tapajós into the Amazon River (Table 6, Fig. 5). Yet, we see no significant structure associated with the confluence of the white waters of the Madeira River into the white waters of the Amazon River. This finding is important because it rules out the possibility of population structure generated by the confluence of a tributary alone (e.g. Fernandes *et al.*, 2004). Indeed, hydrochemical variables

such as oxygen availability have been identified as key factors influencing the distribution of *C. asellus* larvae (Junk *et al.*, 1983; Araujo-Lima, 1994), and sensory adaptations such as optical sensitivity (Appleby & Muntz, 1979) and prey detection have shown to play a major role in the determination of Amazonian fish community structure (Tejerina-Garro *et al.*, 1998). Thus, the striking hydrochemical discontinuities of the Amazonian aquatic environment likely drive adaptive divergence across these selective environments in *C. asellus*.

Here, using genome scans and phylogeographic analyses, we have identified outlier loci geographically clustered around the junction of 'white' and 'clear' waters (Fig. 3) and higher levels of genetic differentiation between selective environments than within (Table 6). Most interestingly however, we have also identified the contribution of these outlier loci to population structure (Fig. 5i, ii).

The *DFDIST* outlier analysis revealed that approximately 3.5% of AFLP loci repeatedly deviated from neutral expectations, signifying them as likely candidates for markers affected by divergent natural selection. Because these loci were repeatedly identified as positive outliers, they are less likely to be false positives (Cooper, 2000; Campbell & Bernatchez, 2004). Due to the anonymous nature of AFLP loci detection, the outlier loci represent regions of the genome under selection but probably not the precise mutation under selection themselves (Jensen *et al.*, 2007). Nonetheless, their presence enabled us to test the role of selection in the maintenance of population boundaries across hydrochemical discontinuities in the Amazon Basin. We found that the majority of positive outliers were identified in population pairwise comparisons involving the easterly populations around the junction of the white waters of the Amazon River and the clear waters of the Tapajós River (Fig. 3), whereas few were found west of A4. Most striking was that after the removal of all repeat-outlier loci, population structure associated with the confluence of the Tapajós River was lost from our AFLP data (e.g. when comparing *STRUCTURE* results between total and putatively neutral data sets; Fig. 5i, ii). In contrast however, population structure associated with the confluence of the Negro River was unaffected after the removal of outlier loci (Table 4; Fig. 5i, ii). Thus, our data suggest strong ecologically driven divergent selection on multiple traits perhaps associated with resource or habitat use coinciding with hydrochemical differences of white and clear water.

There is a chance, however, that the easterly pattern of outlier loci observed may have been a product of 'allele surfing' (Edmonds *et al.*, 2004; Klopstein *et al.*, 2006; Excoffier & Ray, 2008). Under this scenario, the downstream range expansion would have led to the spread of rare alleles that, over time, reached high frequencies in the populations located on the edge of the expansion (A5, T1, A6). Although allele surfing may have

contributed in part to inflated estimates of positive selection between clear and white water habitats, it remains parsimonious that adaptive divergence is occurring across the ecotone. Indeed, a similar signature of divergent natural selection has been found in unrelated yet co-distributed species *Triportheus albus* (Cooke *et al.*, 2012a) and *Plagioscion squamosissimus* (Cooke *et al.*, 2012b). Furthermore, allele surfing is more likely to occur in small and fast growing populations (Klopfstein *et al.*, 2006), and based on diversity estimates and personal observation, this is unlikely the case in the Amazonian pufferfish.

When previously interbreeding populations are isolated from the homogenizing effects of gene flow, neutral DNA will diverge via genetic drift. In the same way, neutral gene flow will decrease and genetic drift will increase as adaptive divergence progresses through time (Nosil *et al.*, 2008). Therefore, we would also expect a neutral signature of structure associated with the junction of the Amazon and Tapajós Rivers if divergent natural selection has persisted long enough for drift to accumulate. Given the loss of nuclear population structure at this aquatic interface after removal of outlier loci, we can conclude that divergent natural selection in this region is a powerful, yet relatively recent phenomenon. Thus, niche preference, niche adaptation and assortative mating are in their infancy across this ecotone. Because of this, pinpointing ecologically driven divergent selection as the mechanism for population subdivision here appears convincing. Under this scenario, genetic changes resulting from selection between spatially and ecologically explicit young lineages have not yet become confounded further by genetic differences that accumulate after isolation (Schluter, 2000; Via, 2009).

In striking contrast, the phylogeographic barrier between the Amazon River and the Tapajós River is well represented in our mtDNA data (Fig. 2). Due to the lower effective population size of the mtDNA molecule and its independence from the nuclear genome, the incipient population subdivision observed in the mtDNA data is in essence 'ecologically allopatric'. Assuming it is unassociated with regions of adaptive divergence or divergence hitchhiking, it has responded independently between the parapatric lineages via either genetic drift or selection within each ecotone (Via, 2009). In this way, 'ecological allopatry' can leave a signature of IBD and restricted gene flow in which divergence increases with adaptation, in the same way that divergence increases with geographic distance (Nosil *et al.*, 2008).

There are however two plausible but not necessarily exclusive explanations for the maintenance of population structure after the confluence of the black waters of Negro River into the white waters of the Amazon River, despite the removal of loci under selection. Firstly, the westerly populations appear older, and barrier-induced divergence of neutral loci via genetic drift will accumulate through time once natural selection has initiated

(ecological allopatry; Nosil *et al.*, 2008; Via, 2009). Therefore, even after the removal of outlier loci, we would expect to see neutral population genetic structure here if natural selection was indeed occurring as a product of hydrochemistry. On the contrary however, few outlier loci were actually identified around the junction of the black and white waters, and *C. asellus* are not typically found in the Negro River (Araujo-Lima, 1994). Therefore, it is more likely that population structure here was induced allopatrically resulting from geomorphological changes in the region since 2.5 Ma. Subsequently, moderate isolation has been maintained via the dynamic geomorphological activity that has persisted at the confluence of the Negro River and Amazon River into the Holocene (Latrubesse & Franzinelli, 2005; Rossetti *et al.*, 2005).

Our phylogeographic scenario now has an additional dimension that includes divergent natural selection resulting from hydrochemistry (Fig. 6). Following the final erosion of the western portion of the Amazon River, *C. asellus* colonized eastward downstream. The consequent migration would have been constrained by limited mobility and larval dispersal (Araujo-Lima, 1994). Most recently however, divergent natural selection has established at the interface of the white waters of the Amazon River and the clear waters of the Tapajós River.

Conclusions

By distinguishing the contribution of vicariant biogeographic history from the power of divergent natural selection in *C. asellus*, we have expanded upon a classical phylogeographic approach. In doing so, we have demonstrated that diversity and spatially distinct population structure can be generated and maintained via multiple forces within a single species across a complex environment. Importantly however, we have shown that time may be the critical factor disabling the identification of natural selection in the wild. Because ecological divergence between white and clear water environments was relatively recent in *C. asellus* populations, the confounding effects of ecological allopatry (genetic drift) have not accumulated within the nuclear genome. This provided us with a unique opportunity to assess the contribution of ecologically mediated divergence versus allopatry.

Because drift will accumulate with time between ecotypes, disguising natural selection as allopatric divergence within the genome (Via, 2009), it is likely that the contribution of natural selection and ecological speciation to biodiversity has been largely underestimated within the literature. This has very important implications for the identification of engines of biodiversity and consequently the conservation of evolutionary processes (Schneider *et al.*, 1999), especially within species-rich ecosystems.

Although there is no certainty that ecologically diverged lineages will eventuate as reproductively

isolated species (Futuyma, 1987; Coyne & Orr, 2004), future work on divergent selection should focus on population-level interactions of genetics and ecology, rather than interspecific studies in which the confounding genetic differences have accumulated (Schluter, 2001; Via, 2009). We are currently entering an exciting era of evolutionary biology in which complimentary research approaches combining gene mapping, population genomics and the identification of functionally important genetic variation promise to define the interface between environment and selection. In this way, studies endeavouring to understand the precise mechanisms and interactions that couple genotype and environment are nurturing a renewed support for Darwin's theory of evolution by natural selection.

Acknowledgments

The authors thank Nathan R. Lovejoy for providing samples from the Orinoco Basin, Colombia and Iquitos, Peru and C. Moritz and A. Hendry for their helpful comments on 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

- Albert, J.S., Petry, P. & Reis, R.E. 2011. Major biogeographic and phylogenetic patterns. In: *Historical Biogeography of Neotropical Freshwater Fishes* (J.S. Albert & R.E. Reis, eds), pp. 21–58. University of California Press, Berkeley.
- Appleby, S.J. & Muntz, R.A. 1979. Occlusable yellow corneas in Tetraodontidae. *J. Exp. Biol.* **83**: 249–259.
- Araujo-Lima, C.A.R.M. 1994. Drift of *Colomesus asellus* (Teleostei: Tetraodontidae) larvae in the Amazon River. *Rev. Hydrobiol. Trop.* **27**: 33–38.
- Beaumont, M.A. & Balding, D.J. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Mol. Ecol.* **13**: 969–980.
- Beaumont, M.A. & Nichols, R.A. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. Lond. B Biol. Sci.* **263**: 1619–1626.
- Beheregaray, L.B. 2008. Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Mol. Ecol.* **17**: 3754–3774.
- Beheregaray, L.B. & Sunnucks, P. 2001. Fine-scale genetic structure, estuarine colonization and incipient speciation in the marine silverside fish *Odontesthes argentinensis*. *Mol. Ecol.* **10**: 2849–2866.
- Bermingham, E., McCafferty, S.S. & Martin, A.P. 1997. Fish biogeography and molecular clocks: perspectives from the Panamanian isthmus. In: *Molecular Systematics of Fishes* (T.D. Kocher & C.A. Stepien, eds), pp. 113–128. Academic Press, San Diego.
- Bolnick, D.I. & Fitzpatrick, B.M. 2007. Sympatric speciation: models and empirical evidence. *Annu. Rev. Ecol. Evol. Syst.* **38**: 459–487.
- Bonin, A., Ehrlich, D. & Manel, S. 2007. Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Mol. Ecol.* **16**: 3737–3758.
- 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. *Mol. Biol. Evol.* **21**: 945–956.
- Campbell, K.E., Frailey, C.D. & Romero-Pittman, L. 2006. The Pan-Amazonian Ucayali Peneplain, late Neogene sedimentation in Amazonia, and the birth of the modern Amazon River system. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **239**: 166–219.
- Clement, M., Posada, D. & Crandall, K.A. 2000. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* **9**: 1657–1659.
- Cooke, G.M., Chao, N.L. & Beheregaray, L.B. 2009. Phylogeography of a flooded forest specialist fish from central Amazonia based on intron DNA: the cardinal tetra *Paracheirodon axelrodi*. *Fresh. Biol.* **54**: 1216–1232.
- Cooke, G.M., Chao, N.L. & Beheregaray, L.B. 2012a. Divergent natural selection with gene flow along major environmental gradients in Amazonia: insights from genome scans, population genetics and phylogeography of the characin fish *Triporthus albus*. *Mol. Ecol.* in press, doi: 10.1111/j.1365-294X.2012.05540.x.
- Cooke, G.M., Chao, N.L. & Beheregaray, L.B. 2012b. Marine incursions, cryptic species and ecological diversification in Amazonia: the biogeographic history of the croaker genus *Plagioscion* (Scaenidae). *J. Biogeogr.* **4**: 724–738.
- Cooper, M.L. 2000. Random amplified polymorphic DNA analysis of southern brown bandicoot (*Isoodon obesulus*) populations in Western Australia reveals genetic differentiation related to environmental variables. *Mol. Ecol.* **9**: 469–479.
- Corrigan, S., Huvneers, C., Schwartz, T.S., Harcourt, R.G. & Beheregaray, L.B. 2008. Genetic and reproductive evidence for two species of ornate wobbegong shark *Orectolobus* spp. on the Australian east coast. *J. Fish Biol.* **73**: 1662–1675.
- Coyne, J.A. 2007. Sympatric speciation. *Curr. Biol.* **17**: R787–R788.
- Coyne, J.A. & Orr, H.A. 2004. *Speciation*. Sinauer Associates, Sunderland, MA.
- Crampton, W.G.R. 2011. An ecological perspective on diversity and distributions. In: *Historical Biogeography of Neotropical Freshwater Fishes* (J.S. Albert & R.E. Reis, eds), pp. 165–192. University of California Press, Berkeley.
- Crispo, E. & Chapman, L.J. 2008. Population genetic structure across dissolved oxygen regimes in an African cichlid fish. *Mol. Ecol.* **17**: 2134–2148.
- Edmonds, C.A., Lillie, A.S. & Cavalli-Sforza, L.L. 2004. Mutations arising in the wave front of an expanding population. *Proc. Natl Acad. Sci. USA* **101**: 975–979.
- Ellegren, H. 2008a. Comparative genomics and the study of evolution by natural selection. *Mol. Ecol.* **17**: 4586–4596.
- Ellegren, H. 2008b. Sequencing goes 454 and takes large-scale genomics into the wild. *Mol. Ecol.* **17**: 1629–1631.
- Endler, J.A. 1977. *Geographic Variation, Speciation, and Clines*. Princeton University Press, Princeton.

- Endler, J.A. 1982. Problems in distinguishing historical from ecological factors in biogeography. *Am. Zool.* **22**: 441–452.
- Evanno, G., Regnaut, S. & J. G. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**: 2611–2620.
- Excoffier, L. & Ray, N. 2008. Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol. Evol.* **23**: 347–351.
- Excoffier, L., Smouse, P.E. & Quattro, J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479–491.
- Excoffier, L., Laval, L.G. & Schneider, S. 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol. Bioinf. Online* **1**: 47–50.
- Falush, D., Stephens, M. & Pritchard, J.K. 2003. Inference of population structure: extensions to linked loci and correlated allele frequencies. *Genetics* **164**: 1567–1587.
- Falush, D., Stephens, M. & Pritchard, J.K. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* **7**: 574–578.
- Fernandes, C.C., Podos, J. & Lundberg, J.G. 2004. Amazonian ecology: tributaries enhance the diversity of electric fishes. *Science* **305**: 1960–1962.
- 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.
- Fu, Y.-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**: 915–925.
- Futuyma, D.J. 1987. On the Role of Species in Anagenesis. *Am. Nat.* **130**: 465–473.
- Garcia-Paris, M., Good, D.A., Parra-Olea, G. & Wake, D.B. 2000. Biodiversity of Costa Rican salamanders: implications of high levels of genetic differentiation and phylogeographic structure for species formation. *Proc. Natl Acad. Sci. USA* **97**: 1640–1647.
- Gavrilets, S. & Vose, A. 2005. Dynamic patterns of adaptive radiation. *Proc. Natl Acad. Sci. USA* **102**: 18040–18045.
- Henderson, P.A. & Crampton, W.G.R. 1997. A comparison of fish diversity and abundance between nutrient-rich and nutrient-poor lakes in the Upper Amazon. *J. Trop. Ecol.* **13**: 175–198.
- Hendry, A.P. 2004. Selection against migrants contributes to the rapid-evolution of reproductive isolation. *Evol. Ecol. Res.* **6**: 1219–1236.
- Holsinger, K.E. & Lewis, P.O. 2005. *HICKORY: A Package for Analysis of Population Genetic Data*. Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT.
- Holsinger, K.E. & Wallace, L.E. 2004. Bayesian approaches for the analysis of population genetic structure: an example from *Platanthera leucophaea* (Orchidaceae). *Mol. Ecol.* **13**: 887–894.
- Holsinger, K.E., Lewis, P.O. & Dey, D.K. 2002. A Bayesian approach to inferring population structure from dominant markers. *Mol. Ecol.* **11**: 1157–1164.
- Hoorn, C. 1993. Marine incursions and the influence of Andean tectonics on the Miocene depositional history of northwestern Amazonia - Results of a palynostratigraphic study. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **105**: 267–309.
- Hoorn, C. 1994. An environmental reconstruction of the Paleo-Amazon river system (middle-late Miocene, NW Amazonia). *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **112**: 187–238.
- Hoorn, C., Guerrero, J., Sarmiento, G.A. & Lorente, M.A. 1995. Andean tectonics as a cause for changing drainage patterns in Miocene northern South America. *Geology* **23**: 237–240.
- Hoorn, C., Wesselingh, F.P., ter Steege, H., Bermudez, M.A., Mora, A., Sevink, J. et al. 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* **330**: 927–931.
- Hubert, N. & Renno, J.-F. 2006. Historical biogeography of South American freshwater fishes. *J. Biogeogr.* **33**: 1414–1436.
- Hubert, N., Duponchelle, F., Nunez, J., Garcia-Davila, C., Paugy, D. & Renno, J.F. 2007. Phylogeography of the piranha genera *Serrasalmus* and *Pygocentrus*: implications for the diversification of the Neotropical ichthyofauna. *Mol. Ecol.* **16**: 2115–2136.
- Jeffreys, H. 1961. *Theory of Probability*. Clarendon Press, Oxford.
- Jensen, J.D., Wong, A. & Aquadro, C.F. 2007. Approaches for identifying targets of positive selection. *Trends Genet.* **23**: 568–577.
- Junk, W.J., Soares, M.G. & Carvalho, F.M. 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.
- Klopfstein, S., Currat, M. & Excoffier, L. 2006. The fate of mutations surfing on the wave of a range expansion. *Mol. Biol. Evol.* **23**: 482–490.
- Latrubesse, E.M. & Franzinelli, E. 2005. The late Quaternary evolution of the Negro River, Amazon, Brazil: implications for island and floodplain formation in large anabranching tropical systems. *Geomorphology* **70**: 372–397.
- Lovejoy, N.R., Bermingham, E. & Martin, A.P. 1998. Marine incursion into South America. *Nature* **396**: 421–422.
- Lovejoy, N.R., Albert, J.S. & Crampton, W.G.R. 2006. Miocene marine incursions and marine/freshwater transitions: evidence from Neotropical fishes. *J. S. Am. Earth Sci.* **21**: 5–13.
- Lovejoy, N.R., Willis, S.C. & Albert, J.S. 2010. Molecular signatures of Neogene biogeographical events in the Amazon fish fauna. In: *Amazonia: Landscape and Species Evolution. A Look Into the Past* (C. Hoorn & F.P. Wesselingh, eds), pp. 405–420. Wiley-Blackwell, West Sussex.
- Luikart, G., England, P.R., Tallmon, D., Jordan, S. & Taberlet, P. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nat. Rev. Genet.* **4**: 981–994.
- Lundberg, J.C., Marshall, L.G., Guerrero, J., Horton, B., Claudia, M., Malabarba, L.R. et al. 1998. The stage for Neotropical fish diversification: a history of tropical South American rivers. In: *Phylogeny and Classification of Neotropical Fishes* (L.R. Malabarba, R.E. Reis, R.P. Vari, Z.M.S. Lucena & C.A.S. Lucena, eds), pp. 13–48. Edipucrs, Porto Alegre.
- Lynch, M. & Milligan, B.G. 1994. Analysis of population genetic structure with rapid markers. *Mol. Ecol.* **3**: 91–99.
- Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**: 209–220.
- Mardis, E.R. 2008. Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* **9**: 387–402.
- McGlashen, D.J., Hughes, J.M. & Bunn, S.E. 2001. Within-drainage population genetic structure of the freshwater fish *Pseudomugil signifer* (Pseudomugilidae) in northern Australia. *Can. J. Fish. Aquat. Sci.* **58**: 1842–1852.
- Mertes, L.A.K., Dunne, T. & Martinelli, L.A. 1996. Channel-floodplain geomorphology along the Solimoes-Amazon River, Brazil. *Geol. Soc. Am. Bull.* **108**: 1089–1107.

- Monks, N. 2006. Puffers. In: *Brackish Water Fishes: An aquarist's Guide to Identification, Care and Husbandry* (N. Monks, ed.), pp. 252–277. T.F.H., Neptune City, NJ.
- Monsch, K.A. 1998. Miocene fish faunas from the northwestern Amazon basin (Colombia, Peru, Brazil) with evidence of marine incursions. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **143**: 31–50.
- Moritz, C., Patton, J.L., Schneider, C.J. & Smith, T.B. 2000. Diversification of rainforest faunas: an integrated molecular approach. *Annu. Rev. Ecol. Syst.* **31**: 533–563.
- Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Nosil, P. 2008. Ernst Mayr and the integration of geographic and ecological factors in speciation. *Biol. J. Linn. Soc.* **95**: 26–46.
- Nosil, P., Vines, T.H. & Funk, D.J. 2005. Perspective: reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution* **59**: 705–719.
- Nosil, P., Egan, S.P. & Funk, D.J. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: “Isolation by adaptation” and multiple roles for divergent selection. *Evolution* **62**: 316–336.
- Nosil, P., Funk, D.J. & Ortiz-Barrientos, D. 2009. Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* **18**: 375–402.
- Ogden, R. & Thorpe, R.S. 2002. Molecular evidence for ecological speciation in tropical habitats. *Proc. Natl Acad. Sci. USA* **99**: 13612–13615.
- Peakall, R. & Smouse, P.E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **6**: 288–295.
- Petry, P., Bayley, P.B. & Markle, D.F. 2003. Relationships between fish assemblages, macrophytes and environmental gradients in the Amazon River floodplain. *J. Fish Biol.* **63**: 547–579.
- Piggott, M.P., Chao, N.L. & Beheregaray, L.B. 2011. Three fishes in one: cryptic species in an Amazonian floodplain forest specialist. *Biol. J. Linn. Soc.* **102**: 391–403.
- Ready, J.S., Ferreira, E.J.G. & Kullander, S.O. 2006. Discus fishes: mitochondrial DNA evidence for a phylogeographic barrier in the Amazonian genus *Symphysodon* (Teleostei: Cichlidae). *J. Fish Biol.* **69**: 200–211.
- Rodriguez, M.A. & Lewis, W.M. 1997. Structure of fish assemblages along environmental gradients in floodplain lakes of the Orinoco River. *Ecol. Monogr.* **67**: 109–128.
- Rogers, A.R. & Harpending, H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9**: 552–569.
- Rossetti, D.D., de Toledo, P.M. & Goes, A.M. 2005. New geological framework for Western Amazonia (Brazil) and implications for biogeography and evolution. *Quat. Res.* **63**: 78–89.
- Saint-Paul, U., Zuanon, J., Correa, M.A.V., Garcia, M., Fabre, N.N., Berger, U. *et al.* 2000. Fish communities in central Amazonian white- and blackwater floodplains. *Environ. Biol. Fish.* **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 Ecol. Evol.* **16**: 372–380.
- Schluter, D. 2009. Evidence for Ecological Speciation and Its Alternative. *Science* **323**: 737–741.
- 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, C.J., Smith, T.B., Larison, B. & Moritz, C. 1999. A test of alternative models of diversification in tropical rainforests: ecological gradients vs. rainforest refugia. *Proc. Natl Acad. Sci. USA* **96**: 13869–13873.
- Sioli, H. 1984. *The Amazon Limnology and Landscape Ecology of a Mighty Tropical River and its Basin*. Dr. Junk Publisher, Dordrecht.
- Sistrom, M.J., Chao, N.L. & Beheregaray, L.B. 2009. Population history of the Amazonian one-lined pencilfish based on intron DNA data. *J. Zool.* **278**: 287–298.
- Smith, T.B., Wayne, R.K., Girman, D.J. & Bruford, M.W. 1997. A role for ecotones in generating rainforest biodiversity. *Science* **276**: 1855–1857.
- Smith, T.B., Schneider, C.J. & Holder, K. 2001. Refugial isolation versus ecological gradients. *Genetica* **112**: 383–398.
- Sunnucks, P. & Hales, D.F. 1996. Numerous transposed sequences of mitochondrial cytochrome oxidase I–II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Mol. Biol. Evol.* **13**: 510–524.
- Tejerina-Garro, F.L., Fortin, R. & Rodriguez, M.A. 1998. Fish community structure in relation to environmental variation in floodplain lakes of the Araguaia River, Amazon Basin. *Environ. Biol. Fish.* **51**: 399–410.
- Templeton, A.R., Crandall, K.A. & Sing, C.F. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132**: 619–633.
- Travers, S.E., Smith, M.D., Bai, J.F., Hulbert, S.H., Leach, J.E., Schnable, P.S. *et al.* 2007. Ecological genomics: making the leap from model systems in the lab to native populations in the field. *Front. Ecol. Environ.* **5**: 19–24.
- Tyler, J.C. 1964. A diagnosis of the two species of South American Puffer fishes (Tetraodontidae, Plectognathi) of the genus *Colomesus*. *Proc. Natl Acad. Sci. Phila.* **116**: 119–148.
- Vari, R.P. & Malabarba, L.R. 1998. Neotropical ichthyology: an overview. In: *Phylogeny and Classification of Neotropical Fishes* (L.R. Malabarba, R.E. Reis, R.P. Vari, Z.M.S. Lucena & C.A.S. Lucena, eds), pp. 1–13. Edipucrs, Porto Alegre.
- Vekemans, X. 2002. *AFLP-SURV Version 1.0*. Distributed by the Author. Laboratoire de Genetique at Ecologie Vegetale, Universite Libre de Bruxelles, Belgium.
- Vekemans, X., T., B., 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. *Mol. Ecol.* **11**: 139–151.
- Via, S. 2009. Natural selection in action during speciation. *Proc. Natl Acad. Sci. USA* **106**: 9939–9946.
- Via, S. & West, J. 2008. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol. Ecol.* **17**: 4334–4345.
- Vonhof, H.B., Wesselingh, F.P., Kaandorp, R.J.G., Davies, G.R., van Hinte, J.E., Guerrero, J. *et al.* 2003. Paleogeography of Miocene Western Amazonia: isotopic composition of molluscan shells constrains the influence of marine incursions. *Geol. Soc. Am. Bull.* **115**: 983–993.
- Wesselingh, F.P., Rasanen, M.E., Irion, G., Vonhof, H.B., Kaandorp, R., Renema, W. *et al.* 2002. Lake Pebas: a palaeoecological reconstruction of a Miocene, long-lived lake complex in western Amazonia. *Cainoz. Res.* **1**: 35–81.
- Whitlock, R., Hipperson, H., Mannarelli, M., Butlin, R.K. & Burke, T. 2008. An objective, rapid and reproducible method

- for scoring AFLP peak-height data that minimizes genotyping error. *Mol. Ecol. Res.* **8**: 725–735.
- Willis, S.C., Nunes, M.S., Montana, C.G., Farias, I.P. & Lovejoy, N.R. 2007. Systematics, biogeography, and evolution of the Neotropical peacock basses *Cichla* (Perciformes: Cichlidae). *Mol. Phys. Evol.* **44**: 291–307.
- Wright, S. 1943. Isolation by distance. *Genetics* **28**: 114–138.
- Wu, C.I. 2001. The genic view of the process of speciation. *J. Evol. Biol.* **14**: 851–865.
- Yamanoue, Y., Miya, M., Doi, H., Mabuchi, K., Sakai, H. & Nishida, M. 2011. Multiple invasions into freshwater by pufferfishes (Teleostei: Tetraodontidae): a mitogenomic perspective. *PLoS One* **6**: e17410.
- Zenger, K.R., Stow, A.J., Peddemors, V., Briscoe, D.A. & Harcourt, R.G. 2006. Widespread utility of highly informative AFLP molecular markers across divergent shark species. *J. Heredity* **97**: 607–611.
- Zhivotovsky, L.A. 1999. Estimating population structure in diploids with multilocus dominant DNA markers. *Mol. Ecol.* **8**: 907–913.

Data deposited at Dryad: doi: 10.5061/dryad.m6k89833

Received 7 February 2012; revised 18 March 2012; accepted 20 March 2012