

# Exceptional genetic differentiation at a micro-geographic scale in *Apistogramma agassizii* (Steindachner, 1875) in the Peruvian Amazon: sympatric speciation?

Guillain Estivals<sup>1</sup>, Fabrice Duponchelle<sup>2</sup>, Carmen Garcia-Davila<sup>3</sup>, Uwe Römer<sup>4</sup>, Cedric Mariac<sup>2</sup>, and Jean-François Renno<sup>2</sup>

<sup>1</sup>MNHN

<sup>2</sup>IRD

<sup>3</sup>IIAP

<sup>4</sup>University of Trier

April 16, 2024

## Abstract

Mechanisms related to ecological or sexual selection have favoured sympatric speciation events in African and Central American lake cichlids. Allopatric divergence is the predominant speciation process observed in Amazonia, although, to the best of our knowledge, no study to date has attempted to determine whether speciation process could exist under sympatric conditions in Amazonian cichlids. The *Apistogramma agassizii* species complex is an excellent model for investigating the existence of sympatric divergence events in the Amazon, as it shares many common life history characteristics with African Haplochromine cichlids in which sympatric speciation mechanisms are well documented. The genetic structure of *A. agassizii* was analysed by genotyping 889 individuals with ten microsatellite loci, collected from 26 sites distributed among small streams in 11 micro-basins in a very small portion of the Peruvian Amazon. It revealed 22 genetic populations identified according to panmictic criteria ( $F_{IS}$  estimator) and strongly differentiated:  $F_{ST}$  estimator (0.034 to 0.356). Such a strong genetic structuring on such small geographical areas has never been demonstrated before in an Amazonian fish. Several of these populations may have diverged sympatrically and repeatedly in small stream networks. The results are discussed with respect to divergence processes, including sympatric speciation, that may be associated with the observed genetic structure.

## Exceptional genetic differentiation at a micro-geographic scale in *Apistogramma agassizii* (Steindachner, 1875) in the Peruvian Amazon: sympatric speciation?

Guillain Estivals<sup>1, 2,\*</sup>, Fabrice Duponchelle<sup>2,3</sup>, Carmen García-Dávila<sup>2,4</sup>, Uwe Römer<sup>2,5,6</sup>, Cédric Mariac<sup>2,7</sup>, Jean-François Renno<sup>2,7,\*</sup>

<sup>1</sup> Ecole Doctorale Sciences de la Nature et de l'Homme–ED 227. Sorbonne Universités, UPMC Univ Paris 06, Paris, France.

<sup>2</sup> Laboratoire Mixte International – Evolution et Domestication de l'Ichtyofaune Amazonienne (LMI–EDIA), Instituto de Investigaciones de la Amazonia Peruana (IIAP) – Peru, Universidad Autónoma Gabriel René Moreno (UAGRM) - Bolivia, IRD (Institut de Recherche pour le Développement) - France

<sup>3</sup> MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, Montpellier, France, Montpellier, France

<sup>4</sup> Instituto de Investigaciones de la Amazonia Peruana (IIAP), AQUAREC, Laboratorio de Biología y Genética Molecular, Iquitos, Perú

<sup>5</sup> Österwieher Straße 196, 33415 Verl; Germany

<sup>6</sup> University of Trier, Animal Research Group, Inst. Biogeography [UTIB], Dep. Geo-Sciences, Universitätsstraße 15, 54286 Trier, Germany

<sup>7</sup> IRD, UMR Diversité Adaptation et Développement des Plantes (DIADE), Montpellier, France

## Correspondance:

Guillain Estivals (*ichtyos3134@hotmail.fr*), Ecole Doctorale Sciences de la Nature et de l'Homme–ED 227. Sorbonne Universités, UPMC Univ Paris 06, Paris, France.

Jean-François Renno (*Jean-Francois.Renno@ird.fr*), Institut de Recherche pour le Développement (IRD), UMR Diversité Adaptation et Développement des Plantes (DIADE), Montpellier, France

## Abstract

Mechanisms related to ecological or sexual selection have favoured sympatric speciation events in African and Central American lake cichlids. Allopatric divergence is the predominant speciation process observed in Amazonia, although, to the best of our knowledge, no study to date has attempted to determine whether speciation process could exist under sympatric conditions in Amazonian cichlids. The *Apistogramma agassizii* species complex is an excellent model for investigating the existence of sympatric divergence events in the Amazon, as it shares many common life history characteristics with African Haplochromine cichlids in which sympatric speciation mechanisms are well documented. The genetic structure of *A. agassizii* was analysed by genotyping 889 individuals with ten microsatellite loci, collected from 26 sites distributed among small streams in 11 micro-basins in a very small portion of the Peruvian Amazon. It revealed 22 genetic populations identified according to panmictic criteria ( $F_{IS}$  estimator) and strongly differentiated:  $F_{ST}$  estimator (0.034 to 0.356). Such a strong genetic structuring on such small geographical areas has never been demonstrated before in an Amazonian fish. Several of these populations may have diverged sympatrically and repeatedly in small stream networks. The results are discussed with respect to divergence processes, including sympatric speciation, that may be associated with the observed genetic structure.

## Keywords

Microsatellites, population structure, speciation, allopatry, sympatry

## Introduction

Understanding the mechanisms involved in the initiation and persistence of genetic divergence in speciation process is a central concern in evolutionary and ecological studies (Funk, Nosil; & Etges, 2006; Maan & Seehausen, 2010; Seehausen & Magalhaes, 2010; Marques et al., 2017). The identification of these mechanisms and their role in speciation processes is key to understanding the driving force of biodiversity. Haplochromine cichlids from the African great lakes and particularly from Lake Victoria have been a model of choice for studying these mechanisms in vertebrates (Kornfield & Smith, 2000; Schluter, 2000; Kocher, 2004; Salzburger & Meyer, 2004; Meier et al., 2017). Theoretical and empirical evidence have accumulated on the prominent role of interactions between ecological and sexual selection in the rapid evolution of cichlids in the African great lakes (Gavrilets, 2004; Salzburger, 2009; Wagner, Harmon, & Seehausen, 2012; Poelstra, Richards, & Martin, 2018), but also for those living in Central American lakes (Wilson, Noack-Kunmann, & Meyer, 2000; Barluenga, Stolting, Salzburger, Muschick, & Meyer, 2006; Machado-Schiaffino et al., 2017). These interactions, when they promote linkage disequilibrium between genes involved in adaptation (e.g. trophic



and visual adaptation mechanisms) and those involved in mating, facilitate the evolution of reproductive isolation (Seehausen et al., 2008; Wagner et al., 2012; Richards, Servedio, & Martin, 2019). The genomic potential (gene pool), through the rearrangement and recombination of ancestral polymorphism, would also play an important role in speciation processes and would have conditioned and accelerated the most spectacular radiation observed to date, in Haplochromines cichlids of Lake Victoria (McGee et al., 2020).

The role of geographical conditions (allopatric, parapatric, sympatric) in divergence and speciation processes are often difficult to define in nature because they are actually distributed along a continuum (Rice & Hostert, 1993; Fitzpatrick, Fordyce, & Gavrillets; 2008). The conditions relating to sympatry are the most difficult in practice. Their definition was originally based on an exclusively geographical criterion, and populations or species found in the same geographical area were then considered sympatric (Poulton, 1904). However, on a small geographical scale, spatial structuring in the distribution of populations can limit gene flow and condition their divergence. A later approach considered that the speciation process operates from a panmictic population, or deme, splitting into new demes that can diverge to form new species (Fitzpatrick et al., 2008), but it only relied on population genetic mechanisms, without taking into account the effect of geography. In addition to the notion of "deme", the distribution of populations through their intrinsic dispersal capacities, or "normal cruising range of individuals", was subsequently proposed (Mallet, Meyer, Nosil, & Feder 2009). Lately, in an effort to standardise these notions, Richards et al. (2019) have proposed that a genetic divergence event could have taken place under sympatric conditions if four criteria are met: 1) the demes (panmictic populations) form a monophyletic clade (originating from the same ancestral and exclusive deme), 2) their distribution areas (or "cruising range") must overlap widely (geographical criterion), 3) the populations must not be the result of secondary contacts, 4) they must be reproductively isolated. Our study is based on this conceptual framework.

The environmental conditions in which Amazonian cichlids live are markedly different from those of the African or central American lake cichlids. The Amazon is the world largest hydrological basin comprising a fractal network of thousands of interconnected watercourses ranging from kilometres-wide rivers to tiny forest brooks, large floodplains and myriads of temporary or permanently connected lakes (Sioli, 1984; Goulding, Barthem, & Ferreira, 2003). A further complexity comes for the chemically contrasted natures of the waters flowing through this network: from Andean-born, nutrient-rich and basic white waters to the nutrient-poor, acidic black- or neutral clear- waters (Sioli, 1984). These environmental conditions are accompanied by highly variable levels of connectivity between ecosystems and a marked seasonality rhythm by the flood pulse (Junk, Bayley, & Sparks, 1989; Junk & Wantzen, 2004). The Amazon basin thus offers innumerable ecological niches that could be favourable to the appearance of adaptive radiations, sympatric and ecological speciation mechanisms, as observed in African rifts and Central American lake environments.

More than a third of the 258 cichlid species described for the Amazon basin (Dagosta & Pinna, 2019) belong to the genus *Apistogramma*. Endemic to South America, it currently totals 94 species, usually small (often < 8cm), mainly distributed in the Amazon basin (Römer, 2000, 2006). The dwarf cichlids of the genus *Apistogramma* offer the opportunity to test the existence of speciation or sympatric divergence events in the Amazon. Indeed, they share many similarities with the Haplochromine cichlids of the African great lakes: high morphological variability, marked sexual dimorphism, colour polymorphism associated with differential mate choice, parental care, high endemism, and trophic adaptations (Römer, 2000, 2006; Römer & Beseinherz, 2005; Ready et al., 2006; Engelking, Römer, & Beisenherz, 2010). Within this genus, the morphological species *Apistogramma agassizii* has the widest geographical distribution in the Amazon basin. It has colonised many aquatic environments in fluvial or lacustrine systems, can live in all types of water (white, clear, black), although it is found preferentially in areas where plant debris accumulate. This species shows a marked sexual dimorphism, a significant colour polymorphism on a local and regional scale and provides parental care for its offspring (Römer, 2000, 2006). In the Peruvian Amazon, in a small portion of the Loreto region alone, Estivals et al (2020) showed that *A. agassizii* is actually composed of at least 3 biological species reproductively isolated by prezygotic barriers through preferential mate choice and the barrier effect of rivers. These three species occur in an extremely reduced geographical area compared to the range originally described for *A. agassizii*. The authors suggest that the important genetic variations

within each of these three species may in fact be composed of diverging populations, or even constitute a species complex. Here, the geographical distribution of the genetic polymorphism of *A. agassizii* Sp1, the most widely distributed species within this complex, was analysed with the objectives of 1) assessing the link between the genetic structuring of the species identified as *A. agassizii* Sp1 and the organisation of the hydrographic network into micro-basins and streams, and 2) identifying the geographical conditions that may have favoured the observed genetic structure.

## Materials and Methods

### Sampling

The biological species *Apistogramma agassizii* (Sp1) was sampled between 2015 and 2018 at 26 geographical sites spread over 11 different micro-basins (stream networks) (Figure 1). Eight of the micro-basins are tributaries of the Ucayali River (A-H), two of the Marañon River (J,K) and one of the Amazon River (I). The sampling sites are all located outside the floodplain of the Ucayali, Marañon and Amazon rivers. A total of 889 individuals were collected (mean per site [?] 34,  $\sigma$  [?] 12).

Fish were caught by professional aquarium fishers using beach seines or dip nets. Fish caught in the same river and along a transect of less than 100 m were considered to belong to the same sampling site. At the camp, all individuals were anaesthetised with eugenol following the protocol of Chanseau, Bosc, Galiay, & Oules (2002) until respiratory arrest, then preserved in a tube with its own identification code and filled with 96% ethanol. The ethanol in each of the tubes was renewed after 24 hours, before conservation in the collection of the "Laboratorio de Biología y Genética Molecular del Instituto de Investigaciones de la Amazonía Peruana" (LBGM-IIAP, Iquitos, Peru).

### Genotyping

DNA from the samples was extracted from 10 mg of tail fin using a modified protocol from Doyle & Doyle (1987). The 10 microsatellite loci (SSRs) were organised and amplified in 4 multiplexes using the Qiagen Multiplex polymerase chain reaction kit and the primers defined by Quéroutil, Vela Diaz, García-Dávila, Römer, & Renno et al. (2015) for *Apistogramma*. All selected SSRs correspond to highly polymorphic dinucleotide motifs. For more details, the DNA extraction methods, amplification conditions and genotyping are identical to those presented in Estivals et al. (2020). The ten SSR markers were genotyped on 889 individuals and each run was composed of samples from at least two different geographical locations and positive controls consisting of reference genotypes to correct for possible shift problems in allelic size assignment. The negative controls used in the amplifications were analysed for possible contamination.

Allele sizes were obtained using the Geneious 9.1.8 program (Kearse et al., 2012). For each locus, when the chromatic signal of an allele was difficult to interpret, a new amplification and reading was performed. Individuals with missing values for more than three loci were removed from the analyses.

### Population structure

The program STRUCTURE version 2.3.4 (Pritchard, Stephens, & Donnelly, 2000) was used as it allows to minimise deviation from panmixia to define clusters of individuals independently from their geographic location and then to evaluate the admixture between geographic populations, most of which are close to each other.

The chosen model accepted admixture (i.e. gene flow), correlated allele frequencies and did not take into account the geographical origin of individuals. In order to search for panmictic units integrating the largest possible number of individuals, the analysis was performed in several steps. The first step was performed on all individuals. The 'optimal' K value was determined from the  $\Delta K$  measure (Evanno, Regnaut, & Goudet,

2005). The deviations from panmixia within these clusters were estimated by calculating the theta value of the Weir & Cockerham (1984)  $F_{IS}$  and considered significant when  $P < 0.025$ . At the end of this first step, the clusters that did not show significant deviations from panmixia were considered as representative of genetic populations (demes). In a second step, a new STRUCTURE analysis was carried out independently for each cluster with remaining significant deviation to panmixia. This process was repeated until obtaining panmictic clusters or until STRUCTURE was no longer able to partition the analysed cluster.

For each STRUCTURE analysis 15 independent iterations were performed for each partitioning K (K from 1 to 27), with a burn-in period of 100,000, followed by a number of Markov chain Monte Carlo (MCMC) repetitions of 500,000. At each step analysis, each individual was placed in the cluster for which it had the highest percentage assignment value (inferred ancestry).

## Allelic and genetic diversity

Allelic and genetic diversities were evaluated for each cluster highlighted by structure, considering: total number of alleles (k), allelic richness (Ar), total number of private alleles (priv) and their proportion (priv / k), expected heterozygosity (He), unbiased expected heterozygosity (Hn.b.) and observed heterozygosity (Hobs.), using the GENETIX software (Belkhir, Borsa, Chikhi, Raufaste, & Bonhomme, 2004) and the 'PopGenKit' R package (Paquette, 2012).

## Population descriptors: measuring differentiation

The degrees of genetic differentiation between clusters were calculated for each pair of clusters from the  $F_{ST}$  theta estimator (Weir & Cockerham, 1984) using the GENETIX program (Belkhir et al., 2004). Then, genetic distances (Nei, 1972) were calculated for each pair of clusters from which an unrooted dendrogram was constructed using the Neighbour Joining method. The robustness of the nodes of the dendrogram was tested by a bootstrap test following 1000 draws. Genetic distance calculations were performed using the R package 'ade4' (Jombart, 2008).

## Isolation by distance

The hypothesis of isolation by distance between clusters predominant at one site, was tested using their genetic distances (Rousset, 1997) calculated with the program Genetix (Belkhir et al., 2004) and their geographical distances (km), measured by following the streamlines calculated from ArcMap version 10.4.1 (ArcMap, 1995 - 2015) and ImageJ version 1.49 (Rasband, 1997-2012). A Mantel test (1967) was first performed on the entire dataset from  $10^4$  replicates using the mantel.randtest function of the 'ade4' package (Dray & Dufour, 2007; Bougeard & Dray, 2018; Chessel, Dufour, & Thioulouse et al., 2004; Dray, Dufour, & Chessel, 2007). A Mantel correlogram test (package 'vegan', Oksanen et al., 2015), was then performed to search for the presence of isolation by distance at different geographical scales, defining kilometre classes at the intra-micro-basin (0 - 5 km) and inter-micro-basin (5-82 km, 82-200 km, 200-400 km) levels, considering their positioning on the hydrographic network (Ucayali, Marañon, Amazon).

## Genetic relatedness/Kinship

Genetic relatedness (kinship) was searched for in the clusters for which a deviation from panmixia could not be rejected at the end of the STRUCTURE analyses. The null hypothesis (absence of relatedness) was tested by permutations (1000), by comparing the average value of genetic identity (Mathieu, Autem, Roux, & Bonhomme, 1990) obtained on all pairs of comparisons between individuals belonging to the tested cluster, with the distribution of values of genetic identity expected in an unstructured population, using the identix version 1.1 program (Belkhir, Castric, & Bonhomme, 2002).

# Results

## Genetic structure

Four steps and 16 STRUCTURE analyses were required to identify the set of genetic clusters closest to panmixia and that best approximate the demes (or genetic populations) (Figure 2, Table S1). A total of 22 clusters were identified, with 17 for which panmixia could not be rejected ( $F_{IS}$  estimator min = -0.093,  $p < 0.072$ ;  $F_{IS}$  estimator max = 0.059  $p < 0.045$ ; at significance level 0.025 (Table S1). The remaining 5 clusters (3, 11, 13, 15 and 18) all have significant heterozygote deficits (Table S1).

## Geographical distribution of clusters

Clusters have a geographic distribution that is highly dependent on micro-basins and their streams (Figure 3). Indeed, at least 76% and up to 100% of the individuals within a cluster come from a single micro-basin, 12 clusters out of 22 are endemic to a single micro-basin and the remaining 10 are overwhelmingly present in one micro-basin (median = 100; mean 95.6,  $\sigma = 7.2$ ). When multiple clusters are endemic or predominant in a single micro-basin, each cluster predominates in a single stream that harbours at least 64% to 100% of the total cluster size (median = 94; mean 90.1,  $\sigma = 9.7$ ) (Table 1). Of the 26 streams sampled, 18 have at least 2 clusters.

## Allelic and genetic diversity

The numbers of the 22 clusters are heterogeneous, ranging from 16 to 106 individuals with a mean value of 40 and a standard deviation of 19 (Table 2). Allelic richness is also variable, ranging from 3.3 to 10.0 (mean = 5.9,  $\sigma = 1.9$ ) approximately. Private alleles are present in low frequencies and generally correspond to rare alleles. Cluster 14, in addition to having the highest allelic richness (10.0), has the highest number of private alleles. Clusters 1 and 2 show the lowest genetic diversity compared to the other clusters.

## Genetic differentiation

The  $F_{ST}$  values between cluster pairs range from 0.034 to 0.356, all of which are significant ( $p$ -value  $< 0.001$ ). The median value of the theta estimator of  $F_{ST}$  is high: 0.151 (Table 3). Clusters 1 and 2 are the two most differentiated and correspond to individuals from micro-basins J (cluster 2) and K (clusters 1), both tributaries of the Marañon River.

## Genetic relationship between clusters (genetic distance)

The dendrogram of genetic distances reveals an organization of clusters by both micro-basins and geographic distances (Figure 4). The nodes are all supported by bootstrap values between 6 and 100. The grouping of clusters 1 and 2 is supported by a bootstrap value of 100 and gathers all individuals sampled in micro-basins K and J (tributaries of the Marañon) plus 2 individuals from micro-basins B and G.

Clustering of clusters 14, 12, 9, 8, and 13 corresponds primarily to individuals sampled in micro-basins A (31/35), B (98/123), F (38/39), H (41/47), and I (60/60), respectively. These micro-basins are tributaries of the Ucayali (A, B, F and H) and the Amazon (I). Clusters 13 and 14 are among the least genetically differentiated ( $F_{ST} = 0.092$ ), yet they mainly gather individuals from the most geographically distant micro-basins.

Clusters 19, 20, 21, and 22 form a set supported by a bootstrap value of 64, which primarily collects individuals sampled in micro-basin C (C1 = 42/77; C2 = 25/30; C3 = 26/30; C4 = 27/77, respectively). Stream C1 gathers 42 of the individuals from cluster 19 ( $N = 43$ ) and 27 of the individuals from cluster 22 ( $N = 28$ ), with these two clusters predominating and coexisting in the same stream (Figure 3).

The grouping of clusters 3, 4, 5, 6 and 7 is supported by a bootstrap value of 65 and gathers mainly individuals sampled in micro-basin D (D3 = 50/52; D2 = 12/12; D4 = 34/40; D5 = 35/39; D1 = 27/30, respectively; Figure 4).

Clusters 16, 17, and 18 form a set supported by a bootstrap value of 64 and groups mainly individuals sampled in micro-basin E (E4 = 33/38, E2 = 26/26; E3 = 26/32, respectively).

Clusters that are endemic or predominant in a micro-basin form a single genetic set. This means that clusters such as C, D or E are genetically closer to each other than they are to clusters in other micro-basins.

Nevertheless, cluster 15, which gathers all the individuals sampled in stream E1, does not group together with the other endemic or predominant clusters of micro-basin E. Similarly, clusters 10 and 11, which correspond mainly to individuals sampled in stream G2 (42 individuals of cluster 11; N = 43) and G1 (23 individuals of cluster 10; N = 28), do not group together.

## Isolation by distance

A significant ( $p\text{-value} = 0.003$ ) isolation by distance was observed only for the smallest geographic distance class (0-5 km; Table 4), which corresponds to intra-micro-basin level. For all other comparisons, isolation by distance was rejected ( $p\text{-value} = 0.385$ ).

## Genetic relatedness/Kinship

Among the 5 clusters with significant heterozygote deficits at the end of the analysis (3, 11, 13, 15 and 18), only clusters 11 and 15 could correspond to genetic relatedness (11:  $p\text{-value} = 0.006$ ; 15:  $p\text{-value} = 0.005$ ).

## Discussion

### Cluster reality / signification?

The identified clusters respond to both a geographical and biological logic. Indeed, the information carried by the clusters is consistent with the geographical distribution of individuals by stream or by micro-basins. Most of the clusters (17 out of 22) did not show significant deviations from panmixia. They would thus correspond to geographical units in which individuals reproduce randomly, forming reproductive units representing genetic populations in *A. agassizii* and that we will consider hereafter as being the representation of demes (Fitzpatrick et al., 2008). For five (3, 11, 13, 15 and 18) out of the 22 clusters, the deviation from panmixia could not be rejected. Two of them (cf. clusters 11 and 15) would correspond to family structures (related individuals). However, the significant heterozygote deficits observed in these “clusters” could be the consequence of a mixing of different demes (Wahlund effect) not detectable by STRUCTURE owing to the small size of the data matrix to analyse. This hypothesis could be tested by increasing our sampling sites (including new streams) within the micro-basins.

The demes are all endemic or predominant to a micro-basin or to a stream within a micro-basin. If we consider only individuals from demes endemic to a stream and those that represent the predominant fraction at a given site, it totals 797 individuals out of 889 analysed. The average assignment value obtained for these individuals for their respective clusters is 0.923 ( $\sigma = 0.11$ ) while that obtained for the remaining, minority fraction at one site (92 individuals) is 0.74 ( $\sigma = 0.175$ ). This lower value of the assignment rate could be explained by current or past gene flow (introgressions) between some demes.

## Deme structure in the study area

Differentiation values ( $F_{ST}$ ) between the 22 demes vary between 0.034 and 0.356 (median = 0.151; mean = 0.157,  $\sigma = 0.064$ ). These very high values are comparable to those found by Amado, Farias, & Hrbek (2011) but at an interspecific level in cichlids of the genus *Symphysodon* (range 0.02-0.38; mean = 0.242,  $\sigma = 0.110$ ) or comparable to those found by Willis et al. (2015) between geographically distant populations in the cichlid species *Cichla temensis* (0.157 vs 0.186). Furthermore, it should be noted that this genetic structuring into 22 demes is observed in a much more restricted geographical area (49,000 km<sup>2</sup>) than those of the previous two studies, carried out at the scale of the Amazon and Orinoco basins (Figure 5).

## Role of geographical factors on genetic differentiation

It is interesting to note that the genetic distance between the 22 demes distributed in the different streams is not correlated with their geographical distance. This lack of correlation could be explained by the presence of barriers to dispersal in the study area. Indeed, the demes of micro-basins J (cluster 2) and K (cluster 1) on the left bank of the Mara  n correspond to the most differentiated from all the others demes, even those geographically close to them but located in micro-basins distributed on the right bank of the Ucayali and the Amazon. The courses of the large rivers could therefore play the role of barrier to dispersal and to gene flow in *A. agassizii* between the left bank of the Mara  n and the right bank of the Ucayali and Amazon rivers.

Moreover, even when the micro-basins are located on the same bank and geographically close, all the demes are endemic or predominant in a single micro-basin. The course of large rivers could therefore also play the role of barrier to the dispersal of *A. agassizii* even between micro-basins located on the same bank.

Additionally, we have observed that some streams partially dried out during the dry season, which could entail severe population decrease. These environmental fluctuations could therefore lead to multiple bottlenecks or founding effects, accelerating genetic drift, and increasing genetic differentiation.

## Deme structure in the micro-basins

The high values of genetic differentiation ( $F_{ST}$ ) observed in the Jenaro Herrera region (up to 0.223) have never been observed before in such a small region of the Amazon for a fish species. The  $F_{ST}$  values between the demes that predominate in the sampled streams range between 0.048 and 0.091 in micro-basin C (clusters 19 to 22), between 0.065 and 0.223 in micro-basin D (clusters 3 to 7) and between 0.056 and 0.168 in micro-basin E (clusters 15 to 18). In these micro-basins, the geographical distances between the sampled streams do not exceed 5 km following the course of the river networks (mean (C,D,E) = 2.77 km,  $\sigma = 0.98$ ). The genetic differentiation observed between the predominant demes within streams is correlated with geographical distance, which may reflect poor dispersal abilities of the demes.

Between streams within the same micro-basin, the extreme  $F_{ST}$  values observed are of the same order of magnitude as those observed between demes of *Apistogramma geophyra* ( $F_{ST} = 0.128$ ) or *A. pertensis* ( $F_{ST} = 0.228$ ) that are at least 55 km apart and are thought to have diverged since the Middle or Late Pleistocene, as a result of the separation of their original watersheds (Leit  o et al., 2017). The differentiation values observed in *A. agassizii* are therefore extremely strong both at the scale of the micro-basin and between micro-basins and despite the proximity of their confluences (a few hundred meters for C, D and E), their demes did not mix.

It is worth noting that the demes corresponding to clusters 19 and 22 are abundant in stream C1, with 42 individuals for cluster 19 ( $N = 43$ ) and 27 individuals for cluster 22 ( $N = 28$ ). They are genetically differentiated ( $F_{ST} = 0.051$ ;  $p < 0.001$ ), but are more closely genetically related to each other than to other demes, suggesting they could be sister demes originating from a common ancestor deme. The individuals in stream C1 were sampled over a fraction of the stream less than 100 meters long and ~1 m width, so if there was even partial gene flow between these two demes, one would expect them to homogenise rapidly. These

perfectly sympatric but reproductively isolated demes would constitute biological species according to Mayr (1942).

## What would fall under allopatric, parapatric and sympatric conditions in the study area?

We do not know the extent of the distribution areas of the observed demes, nor those of their overlaps. However, we do know the geographical location of the individuals sampled from the 22 demes, and we can therefore understand their "normal cruising range of individuals" (Mallet, 2008) in relation to the geographical conditions in which they evolve. According to Mallet (2009)'s definition of allopatry: "Where groups of populations are separated by uninhabited space across which dispersal and gene flow occurs at very low frequency", micro-basins could correspond to distinct geographical units between which gene flow is very reduced, similar to allopatric or 'micro-allopatric' conditions. The genetic divergence between demes of different micro-basins could depend on ecological and physical constraints (barrier role of a river to dispersion inter-banks or along the same bank) which would favour genetic differentiation phenomena by vicariance.

In river systems, the dispersal of individuals tends to be asymmetrical, following the direction of the current (Hanfling & Weetman, 2006; Crispo, Bentzen, Reznick, Kinnison, & Hendry, 2006). Some fish species, particularly *Apistogramma*, may disperse by taking refuge in rafts of floating vegetation, dispersing passively over long distances (Schiesari et al., 2003). However, between micro-basins we do not observe any gradient of genetic polymorphism in the differentiation of the demes or in their number: on the contrary, each identified deme is endemic or predominant in a single micro-basin and no movement of dispersion can clearly be demonstrated between micro-basins. Under the assumption that the demes were mostly formed where they are now, there may be an ecological barrier to their dispersal. The morphological species *A. agassizii* prefers shallow areas with little current (lentic) where plant debris and especially dead leaves accumulate (Römer, 2000). This type of habitat is mostly found in small streams, such as those in micro-basins and in lakes. On the contrary, large rivers correspond to lotic and deep environments that are not very favourable to the presence of *A. agassizii*. It has been shown in cichlids from the African great lakes that a habitat discontinuity of only a few tens of metres can sometimes be sufficient to isolate two demes (Rico & Turner, 2002). The reduction in gene flow between micro-basins, whose mouths in the main river are separated by only a few hundred metres following a shoreline dispersal, could also be reinforced in lotic environments by inappropriate habitats and higher predation in large rivers (Dodrill, Yard, & Pine Iii, 2016; Deacon, Jones, & Magurran, 2018).

### *Sympatric conditions*

Within a micro-basin we would be under the conditions of sympatry as described by Mallet et al. (2009): "where individuals are physically capable of encountering one another with moderately high frequency. Populations may be sympatric if they are ecologically segregated, as long as a fairly high proportion of each population encounters the other along ecotones; and they may be sympatric, yet breed at different seasons". Indeed, habitats suitable for the presence of *A. agassizii* were observed along the networks of streams that make up these micro-basins. Each sampled stream is represented by a single collection site that groups individuals sampled over a section of stream of only about 100 metres. Our representation of the distribution of demes in the streams and *a fortiori* in the micro-basins is therefore very fragmented. However, although the endemic or predominant demes identified in the micro-basins are usually predominant in a single stream, they are also often present in other streams of the same micro-basin, which extends their actual distribution ("normal cruising range of individuals" according to Mallet et al., 2009) outside the stream where they are predominant. The cruising ranges of the demes may therefore overlap between streams of the same micro-basin, which supports the hypothesis of sympatry.

## Evidence of a sympatric divergence process in cichlids in the Amazon?

In each of the micro-basins C, D and E in the Jenaro Herrera region, the demes are more closely related genetically to each other than to other micro-basins. The four endemic or predominant demes in micro-basin C are clustered on the dendrogram and are supported by a bootstrap value of 64. The same is observed for the five most abundant demes in micro-basin D (bootstrap value = 65) and for three of the four demes in micro-basin E (bootstrap value of 64). This means that for these three situations, the endemic or predominant demes in the same micro-basin would have diverged from the same ancestral deme. It seems very unlikely that on three occasions, the predominant demes within these micro-basins would have diverged into different micro-basins and then that they would have met following secondary contacts in the same micro-basin. The interpretation of our observations by multiple divergence events (from a same ancestral deme, but different in each micro-basin) having occurred independently in the three micro-basins seems to be the most likely interpretation. Furthermore, for the demes corresponding to clusters 19 and 22, quasi-endemic to stream C1, as well as for the demes that have an overlapping "normal cruising range of individuals", the observed differentiation seems to have operated in perfect sympatry.

The presence of strongly differentiated demes between micro-basins, between streams within the same micro-basin and within the same stream, suggests the existence of pre-zygotic or post-zygotic barriers favouring divergence between these demes. Otherwise, gene flow between demes would oppose the forces of divergence and tend towards genetic homogenisation. However, the evolution of intrinsic postzygotic isolation among fish species usually takes millions of years (review in Seehausen & Wagner, 2014). It is also a rare phenomenon in cichlid species, where divergence is usually maintained through prezygotic barriers such as behavioural mate choice (Seehausen, Van Alphen, & Witte, 1997; Wilson et al., 2000; Römer & Beisenherz, 2005). Behavioural mate choice has been demonstrated in many species of *Apistogramma* (Römer & Beisenherz, 2005; Ready et al., 2006; Engelking et al., 2010), including *A. agassizii* (Estivals et al., 2020) and more widely in many African (Seehausen, Van Alphen, & Witte, 1997; Danley & Kocher, 2001; Wagner, McCune, & Lovette, 2012) and Central American cichlid species (Barlow & Siri, 1997; Wilson et al., 2000; Barluenga & Meyer, 2004). It is therefore likely that the maintenance of genetic structure in *A. agassizii* is favoured by a pre-zygotic barrier through mate choice, which would play an important role in the maintenance and divergence of these demes in sympatry. However, we cannot reject the hypothesis that disruptive ecological selection, without mate choice, could be involved in reproductive isolation (Schluter, 2000; Van Doorn, Edelaar, & Weissing, 2009; Elmer, Lehtonen, Kautt, Harrod, & Meyer, 2010). An interaction between sexual and ecological selection 'facilitating' the evolution towards reproductive isolation is also possible (Seehausen et al., 2008; Wagner et al., 2012; Richard et al. 2019; McGee et al. 2020).

In addition to the criterion of sharing the same geographical area, the demes present in micro-basins C, D and E would meet the three conditions required for sympatric speciation (Mallet et al., 2009): 1) they would diverge from an ancestral deme, 2) they would not result from secondary contacts, but would be distributed where they formed, 3) they would be reproductively isolated.

### *Apistogramma*, the model for Amazonian cichlids?

In this Amazonian cichlid species, we show here a complex genetic structure, compatible with multiple divergence events that could have taken place in different geographical situations, from allopatry to sympatry and under the constraint of different ecological characteristics ranging from those of large rivers to those of small streams. Although speciation processes in Amazonian ichthyofauna are thought to be mainly based on allopatric conditions (Hubert et al., 2007; Albert & Reis, 2011; Dias, Cornu, Oberdorff, Lasso, & Tedesco, 2013), part of this biodiversity could also be based on sympatric speciation. In addition to sharing many characteristics with the Haplochromines of the African great lakes, *A. agassizii* presents an important genetic structuring that tends to be organised according to different geographical scales, thus allowing us to study the influence of geographical, but also ecological and historical conditions on the divergence process. The important levels of differentiation observed between demes and their maintenance in sympatry raise questions about the status of the species *Apistogramma agassizii*. Our observations reinforce the hypothesis that *A.*



*agassizii* would be a complex of species and demes in the process of divergence at diverse geographical scales (Estivals et al., 2020). It seems very likely that other species of the same genus, such as *A. bitaeniata* and *A. cacatuoides*, which have a wide distribution in the Amazon basin, could present such a genetic structuring and would reinforce the hypothesis of sympatric speciation. The potential number of biological species grouped in the morphological species *A. agassizii*, and more broadly in the genus *Apistogramma*, could then be several hundred species (Ready et al., 2006; Estivals et al., 2020). In order to unravel the mechanisms involved in their evolution, a particular effort should be made to study mechanisms related to ecological and sexual selection that have played an important role in the rapid evolution of cichlids in the African great lakes and Central American lakes (Van Doorn & Weissing, 2001; Gavrillets, 2004; Salzburger, 2009; Wagner, Harmon, & Seehausen, 2012; Machado-Schiaffino et al., 2017). Additionally, several recent studies have shown that secondary gene flow, hybridization and introgressions events are common in cichlids (Malinsky et al., 2015; Kautt, Machado-Schiaffino et al., 2017; Meier et al., 2017; Brock & Wagner, 2018; Poelstra et al., 2018; Richards, Poelstra, & Martin, 2018), and are thought to have conditioned and accelerated species divergence (McGee et al., 2020). Given the paradigm constituted by the cichlids of African and Central American lakes, the genus *Apistogramma* thus appears to be an excellent model for the study of specific and adaptive radiation events in cichlids at the scale of Amazonian dendritic networks.

## Acknowledgements

Guillain Estivals was supported by a scholarship of the Franco-Peruvian School of Life Sciences. Research was funded by and carried out within the framework of the LMI-EDIA (International Mixt Laboratory - Evolution and Domestication of Amazonian Ichthyofauna), IIAP-IRD-UAGRM. The authors declare to have no conflict of interest regarding the publication of this article.

## References

- Albert, J., & Reis, R. (2011). *Historical Biogeography of Neotropical Freshwater Fishes*. University of California Press.
- Amado, M., Farias, I., & Hrbek, T. (2011). A Molecular Perspective on Systematics, Taxonomy and Classification Amazonian Discus Fishes of the Genus *Symphysodon*. *International journal of evolutionary biology*, 2011, 1-16. doi:10.4061/2011/360654
- ArcGIS® software by Esri. ArcGIS® and ArcMap are the intellectual property of Esri (1995 – 2015)
- Barlow, G. W., & Siri, P. (1997). Does sexual selection account for the conspicuous head dimorphism in the Midas cichlid? *Animal Behaviour*, 53, 573-584. doi:10.1006/anbe.1996.0332
- Barluenga, M., & Meyer, A. (2004). The Midas cichlid species complex: incipient sympatric speciation in Nicaraguan cichlid fishes? *Molecular Ecology*, 13, 2061-2076. doi:10.1111/j.1365-294X.2004.02211.x
- Barluenga, M., Stolting, K. N., Salzburger, W., Muschick, M., & Meyer, A. (2006). Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature*, 439, 719-723. doi:10.1038/nature04325
- Belkhir, K., Castrie, V., & Bonhomme, F. (2002). identix, a software to test for relatedness in a population using permutation methods. *Molecular Ecology Notes*, 2, 611-614. doi:10.1046/j.1471-8286.2002.00273.x
- Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., & Bonhomme, F. (2004). Genetix, logiciel sous Windows TM pour la genetique des populations. In *Laboratoire Genome, populations, interactions CNRS UMR 5000*. Montpellier (France): Universite de Montpellier II.
- Bougard S., & Dray S. (2018). Supervised Multiblock Analysis in R with the ade4 Package. *Journal of Statistical Software*, 86, 1-17. doi:10.18637/jss.v086.i01.
- Brock, C. D., & Wagner, C. E. (2018). The smelly path to sympatric speciation? *Molecular Ecology*, 27, 4153-4156. doi:10.1111/mec.14845

- Chanseau, M., Bosc, S., Galiay, E., & Oules, G. (2002). L'utilisation de l'huile de clou de girofle comme anesthésique pour les smolts de saumon atlantique (*Salmo salar* L.) et comparaison de ses effets avec ceux du 2-phénoxyéthanol. *Bulletin Français de la Pêche et de la Pisciculture* , 365-366, 579–589. doi:doi.org/10.1051/kmae:2002054
- Chessel, D., Dufour, A.-B., & Thioulouse, J. (2004). The ade4 package-I: One-table methods. *R News*, 4 , 5-10.
- Crispo, E., Bentzen, P., Reznick, D. N., Kinnison, M. T., & Hendry, A. P. (2006). The relative influence of natural selection and geography on gene flow in guppies. *Molecular Ecology* , 15 , 49–62. doi:10.1111/j.1365-294X.2005.02764.x
- Dagosta, F., & de Pinna, M. (2019). The fishes of the Amazon: Distribution and biogeographical patterns, with a comprehensive list of species. *Bulletin of the American Museum of Natural History* , 431 , 1–163. doi:10.1206/0003-0090.431.1.1
- Danley, P.D., & Kocher, T.D. (2001). Speciation in rapidly diverging systems: lessons from Lake Malawi. *Molecular Ecology*, 10, 1075–1086. doi:10.1046/j.1365-294x.2001.01283.x
- Deacon, A. E., Jones, F. A. M., & Magurran, A. E. (2018). Gradients in predation risk in a tropical river system. *Current zoology* , 64 , 213-221. doi:10.1093/cz/zoy004
- Dias, M. S., Cornu, J.-F., Oberdorff, T., Lasso, C. A., & Tedesco, P. A. (2013). Natural fragmentation in river networks as a driver of speciation for freshwater fishes. *Ecography*, 36 , 683-689. doi:10.1111/j.1600-0587.2012.07724.x
- Dodrill, M. J., Yard, M., & Pine Iii, W. E. (2016). Assessing predation risks for small fish in a large river ecosystem between contrasting habitats and turbidity conditions. *American Midland Naturalist*, 17 , 206-221. doi:10.1674/0003-0031-175.2.206
- Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry Bulletin* , 19 , 11–15.
- Dray, S., & Dufour, A.B. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software* , 22 , 1-20. doi:10.18637/jss.v022.i04.
- Dray, S., Dufour, A.B. & Chessel, D. (2007). The ade4 package - II: Two-table and K-table methods. *R News* , 7 , 47-52.
- Elmer, K. R., Lehtonen, T. K., Kautt, A. F., Harrod, C., & Meyer, A. (2010). Rapid sympatric ecological differentiation of crater lake cichlid fishes within historic times. *BMC Biology*, 8 , 60. doi:10.1186/1741-7007-8-60
- Engelking B, Romer U, & Beisenherz W. (2010). Intraspecific colour preference in mate choice by female *Apistogramma cacatuoides* HOEDEMAN, 1951 (Teleostei: Perciformes: Cichlidae). *Vertebrate Zoology* , 60 , 199–208.
- Estivals, G, Duponchelle, F., Romer, U., Garcia-Davila, C., Airola, E., Deleglise, M., & Renno (2020). The Amazonian dwarf cichlid *Apistogramma agassizii* (Steindachner, 1875) is a geographic mosaic of potentially tens of species: conservation implications. *Conservation: Marine and Freshwater Ecosystems* , 30 , 1521-1539. doi:10.1002/aqc.3373
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* , 14 , 2611-2620. doi:10.1111/j.1365-294X.2005.02553.x
- Fitzpatrick, B. M., Fordyce, J. A., & Gavrillets, S. (2008). What, if anything, is sympatric speciation? *Journal of Evolutionary Biology*, 21 , 1452-1459. doi:10.1111/j.1420-9101.2008.01611.x

- Funk, D. J., Nosil, P., & Etges, W. J. (2006). Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 3209–3213. doi:10.1073/pnas.0508653103
- Gavrilets, S. (2004). *Fitness landscapes and the origin of species*. Princeton, NJ, USA: Princeton University Press.
- Goulding, M., Barthem, R., & Ferreira, E. J. G. (2003). *The Smithsonian Atlas of the Amazon*. Smithsonian Institution.
- Hanfling, B., & Weetman, D. (2006). Concordant genetic estimators of migration reveal anthropogenetically enhanced source-sink population structure in the river sculpin, *Cottus gobio*. *Genetics*, *173*, 1487–1501. doi:10.1534/genetics.105.054296
- 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. *Molecular Ecology*, *16*, 2115–2136. doi:10.1111/j.1365-294X.2007.03267.x
- Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–1405. doi:10.1093/bioinformatics/btn129
- Junk, W. J., Bayley, P. B., & Sparks, J. S. (1989). *The flood pulse concept in river floodplain systems* (D. P. Dodge, Ed.; Vol. 106, pp. 110–127). Canadian Special Publication of Fisheries and Aquatic Sciences.
- Junk, W. J., & Wantzen, K. M. (2004). *The flood pulse concept: New aspects, approaches and applications—An update: Vol. Volume I* (R. L. Welcomme & T. Petr, Eds.; pp. 117–149). Food and Agriculture Organization and Mekong River Commission, FAO Regional Office for Asia and the Pacific.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*, 1647–1649. doi:10.1093/bioinformatics/bts199
- Kocher, T. D. (2004). Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews Genetics*, *5*, 288–298. doi:10.1038/nrg1316
- Kornfield I. & Smith P.F. (2000) African cichlid fishes: model systems for evolutionary biology. *Annual Review of Ecology, Evolution, and Systematics*, *31*, 163–196. doi:10.1146/annurev.ecolsys.31.1.163
- Leitao, C. S. S., Santos, C. H. A., Souza, E. M. S., Vilarinho, G. C. C., Paula-Silva, M. N., Val, P., . . . de Almeida-Val, V. M. F. (2017). Development and characterization of microsatellite loci in Amazonian dwarf cichlids *Apistogramma* spp. (Perciformes: Cichlidae): Uncovering geological influence on Amazonian fish population. *Journal of Applied Ichthyology*, *33*, 1196–1199. doi:10.1111/jai.13490
- Maan, M. E., & Seehausen, O. (2010). Mechanisms of species divergence through visual adaptation and sexual selection: Perspectives from a cichlid model system. *Current Zoology*, *56*, 285–299. doi.org/10.1093/czoolo/56.3.285
- Machado-Schiaffino, G., Kautt, A. F., Torres-Dowdall, J., Baumgarten, L., Henning, F., & Meyer, A. (2017). Incipient speciation driven by hypertrophied lips in Midas cichlid fishes? *Molecular Ecology*, *26*, 2348–2362. doi:10.1111/mec.14029
- Malinsky, M., Challis, R. J., Tyers, A. M., Schiffels, S., Terai, Y., Ngatunga, B. P., . . . & Turner, G. F. (2015). Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science*, *350*, 1493–1498. doi:10.1126/science.aac9927
- Mallet, J. 2008. Hybridization, ecological races, and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society of London. Series B*, *363*, 2971–2986. doi:10.1098/rstb.2008.0081

- Mallet, J., Meyer, A., Nosil, P., & Feder, J. L. (2009). Space, sympatry and speciation. *Journal of Evolutionary Biology*, *22*, 2332-2341. doi:10.1111/j.1420-9101.2009.01816.x
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, *27*, 209-220.
- Marques, D. A., Lucek, K., Haesler, M. P., Feller, A. F., Meier, J. I., Wagner, C. E., . . . & Seehausen, O. (2017). Genomic landscape of early ecological speciation initiated by selection on nuptial colour. *Molecular Ecology*, *26*, 7-24. doi:10.1111/mec.13774
- Mathieu, E., Autem, M., Roux, M., Bonhomme, F. (1990) Epreuves de validation dans l'analyse de structures genetiques multivariees: comment tester l'equilibre panmictique? *Revue de Statistique Appliquee*, *38*, 47-66.
- McGee, M., Borstein, S., Meier, J., Marques, D., Mwaiko, S., Taabu, A., . . . & Seehausen, O. (2020). The ecological and genomic basis of explosive adaptive radiation. *Nature*, 1-5. doi:10.1038/s41586-020-2652-7
- Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, *8*, 1-11. doi:10.1038/ncomms14363
- Nei, M. (1972). Genetic distance between populations. *The American Naturalist*, *106*, 283-292. doi:10.1086/282771
- Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P., O'Hara, B., . . . Wagner, H. (2015). *Vegan: Community Ecology Package. R Package Version 2.2-1*, *2*, 1-2.
- Paquette, S.R. (2012). PopGenKit: Useful functions for (batch) file conversion and data resampling in microsatellite datasets. R Package Version 1.0. *Molecular Ecology Notes*, *2*, 611-614. doi:10.1046/j.1471-8278.2002.00273.x
- Poelstra, J. W., Richards, E. J., & Martin, C. H. (2018). Speciation in sympatry with ongoing secondary gene flow and a potential olfactory trigger in a radiation of Cameroon cichlids. *Molecular Ecology*, *27*, 4270-4288. doi:10.1111/mec.14784
- Poulton, E.B. (1904). What is a species? *Proceedings of the Entomological Society of London*, *1903*, lxxvii-cxvi.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945-959.
- Querouil, S., Vela Diaz, A., Garcia-Davila, C., Romer, U., & Renno, J. F. (2015). Development and characterization of polymorphic microsatellite markers in neotropical fish of the genus *Apistogramma* (Perciformes: Labroidei: Cichlidae). *Journal of Applied Ichthyology*, *31*, 52-56. doi:10.1111/jai.12975
- Rasband, W. S., 1997-2012: ImageJ. US National Institutes of Health, Bethesda.
- Ready, J. S., Sampaio, I., Schneider, H., Vinson, C., Dos Santos, T., & Turner, G. F. (2006). Colour forms of Amazonian cichlid fish represent reproductively isolated species. *Journal of Evolutionary Biology*, *19*, 1139-1148. doi:10.1111/j.1420-9101.2006.01088.x
- Rice, W.R. & Hostert, E.E. (1993). Laboratory experiments on speciation: what have we learned in forty years? *Evolution*, *47*, 1637-1653. doi: 10.1111/j.1558-5646.1993.tb01257.x
- Richards, E. J., Poelstra, J. W., & Martin, C. H. (2018). Don't throw out the sympatric speciation with the crater lake water: fine-scale investigation of introgression provides equivocal support for causal role of secondary gene flow in one of the clearest examples of sympatric speciation. *Evolution letters*, *2*, 524-540. doi:10.1002/evl3.78
- Richards, E. J., Servedio, M. R., & Martin, C. H. (2019). Searching for Sympatric Speciation in the Genomic Era. *Bioessays*, *41*, 1-10. doi:10.1002/bies.201900047

- Rico, C., & Turner, G.F. (2002) Extreme microallopatric divergence in a cichlid species from Lake Malawi. *Molecular Ecology*, 11 , 1585–1590. doi:10.1046/j.1365-294X.2002.01537.x
- Romer, U. (2000). *Cichlid atlas*. In *Natural history of South American dwarf cichlids* (Vol. 1 ). Melle: Mergus Verlag.
- Romer, U. (2006). *Cichlid atlas*. In *Natural history of South American dwarf cichlids* (Vol. 2). Melle: Mergus Verlag.
- Romer, U., & Beisenherz, W. (2005). Intra- and interspecific mate choice of female *Apistogramma cacatuoides* (Teleostei: Cichlidae). *Ichthyological Exploration of Freshwaters* , 16 , 339–345.
- Salzburger, W. (2009). The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Molecular Ecology*, 18 , 169–185. doi:10.1111/j.1365-294X.2008.03981.x
- Salzburger, W., & Meyer, A. (2004). The species flocks of East African cichlid fishes: recent advances in molecular phylogenetics and population genetics. *Naturwissenschaften* , 91 , 277–290. doi:10.1007/s00114-004-0528-6
- Schluter, D. (2000). *The ecology of adaptive radiation* . New York: Oxford University Press.
- Schiesari, L., Zuanon, J., Azevedo-Ramos, C., Garcia, M., Gordo, M., Messias, M., & Vieira, E. (2003). Macrophyte rafts as dispersal vectors for fishes and amphibians in the Lower Solimes River, Central Amazon. *Journal of Tropical Ecology*, 19 , 333–336. doi:10.1017/S0266467403003365
- Seehausen, O., & Magalhaes, I. S. (2010). Geographical Mode and Evolutionary Mechanism of Ecological Speciation in Cichlid Fish. In Grant, P., Grant, R., (Eds.), *In search of the causes of evolution*. (pp. 282–308). Princeton, NJ, USA: Princeton University Press.
- Seehausen, O., Van Alphen, J.J.M., & Witte, F. (1997). Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science*, 277 , 1808–1811. doi:10.1126/science.277.5333.1808
- Seehausen, O., & Wagner, C. E. (2014). Speciation in Freshwater Fishes. *Annual Review of Ecology, Evolution, and Systematics*, 45 (1), 621–651. doi:10.1146/annurev-ecolsys-120213-091818
- Sioli, H. (1984). The Amazon and its main affluents: Hydrography, morphology of the river courses, and river types. In H. Sioli (Ed.), *The Amazon: Limnology and landscape ecology of a mighty tropical river and its basin*. (pp. 127–165). Dr W. Junk Publishers.
- Steindachner, F. (1875). Beitrage der Kenntniss der Chromiden des Amazonenstromes. Sitzungsberichte der kaiserlichen Akademie der Wissenschaften. *Mathematischnaturwissenschaftliche Classe* , 71 , 61–137.
- Van Doorn, G. S., Edelaar, P., & Weissing, F. J. (2009). On the Origin of Species by Natural and Sexual Selection. *Science*, 326 , 1704–1707. doi:10.1126/science.1181661
- Van Doorn G.S & Weissing F.J. (2001). Ecological versus sexual selection models of sympatric speciation: a synthesis. *Selection* .2 , 17–40. doi:10.1556/Select.2.2001.1-2.
- Wagner, C.E., Harmon, L.J. & Seehausen, O. (2012). Ecological opportunity and sexual selection together predict adaptive radiation. *Nature* , 487, 366–369. doi:10.1038/nature11144
- Wagner, C. E., McCune, A. R., & Lovette, I. J. (2012). Recent speciation between sympatric Tanganyikan cichlid colour morphs. *Molecular Ecology* , 21 , 3283–3292. doi:10.1111/j.1365-294X.2012.05607.x
- Wahlund, S. (1928). Zusammensetzung von populationen und korrelationserscheinungen vom standpunkt der vererbungslehre aus betrachtet. *Hereditas* , 11 , 65–106. doi:10.1111/j.1601-5223.1928.tb02483.x
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* , 38 , 1358–1370. doi:10.2307/2408641

Willis, S. C., Winemiller, K. O., Montana, C. G., Macrander, J., Reiss, P., Farias, I. P., & Orti, G. (2015). Population genetics of the speckled peacock bass (*Cichla temensis*), South America's most important inland sport fishery. *Conservation Genetics*, 16 , 1345-1357. doi:10.1007/s10592-015-0744-y

Wilson, A. B., Noack-Kunnmann, K., & Meyer, A. (2000). Incipient speciation in sympatric Nicaraguan crater lake cichlid fishes: sexual selection versus ecological diversification. *Proceedings of the Royal Society B: Biological Sciences* , 267 , 2133-2141. doi:10.1098/rspb.2000.1260

TABLES

Table 1.

Distribution of each cluster, in percentage, according to micro-basins and streams (sample site).

	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster
<i>Micro-basin</i>	1	2	3	4	5	6	7	8	9	10	11
N	22	50	60	16	36	38	33	54	40	28	43
A								2			
B		2	7					17	5		
C			3					6			
D			88	100	100	100	100				
E											2
F									95		
G		2	2							100	98
H								76			
I											
J		96									
K	100										
<i>Stream</i>											
A								2			
B1			3					4	5		
B2		2	2					11			
B3			2					2			
C1											
C2			2					2			
C3			2					4			
D1					3	5	82				
D2				100							
D3			83		3						
D4					94	3	15				
D5			5			92	3				
E1											2
E2											
E3											
E4											
F									95		
G1			2							18	98
G2		2								82	
H1								43			
H2								33			
I1											
I2											
J1		48									

	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster
J2		48									
K	100										

Table 2  
Allelic diversity

Cluster	<i>N</i>	k	Ar	priv	priv/k	He	Hn.b.	Hobs.
1	22	34	3.253	0	0.000	0.425	0.435	0.436
2	50	71	4.747	2	0.028	0.409	0.413	0.424
3	60	89	6.157	3	0.034	0.687	0.692	0.650
4	16	37	3.700	0	0.000	0.544	0.562	0.613
5	36	45	4.185	0	0.000	0.618	0.627	0.625
6	38	52	4.526	2	0.038	0.541	0.548	0.521
7	33	45	4.011	0	0.000	0.557	0.566	0.579
8	54	121	8.266	3	0.025	0.729	0.736	0.715
9	40	88	6.806	3	0.034	0.688	0.696	0.695
10	28	81	6.813	4	0.049	0.653	0.665	0.653
11	43	67	5.432	1	0.015	0.632	0.640	0.579
12	106	184	9.520	16	0.087	0.753	0.757	0.736
13	62	126	8.127	19	0.151	0.723	0.729	0.701
14	39	147	9.951	19	0.129	0.759	0.769	0.751
15	27	49	4.616	1	0.020	0.610	0.622	0.548
16	39	60	5.040	0	0.000	0.624	0.632	0.658
17	27	58	5.346	0	0.000	0.656	0.668	0.700
18	31	53	4.827	0	0.000	0.614	0.624	0.574
19	43	62	5.404	0	0.000	0.644	0.651	0.670
20	39	99	7.808	1	0.010	0.720	0.729	0.710
21	28	60	5.385	0	0.000	0.647	0.659	0.621
22	28	63	5.659	0	0.000	0.611	0.622	0.637

(*N* = sample size, k = total alleles, Ar = allelic richness, priv = private allele, priv/k = proportion of private allele; He = expected heterozygotie, Hn.b.= heterozygotie without bias, Hobs. = observed heterozygotie).

Table 3  
Pairwise value of  $F_{ST}$  estimator of Weir & Cockerham (1984) between 22 clusters from *Apistogramma agassizii*.

Cluster	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
<i>N</i>	22	50	60	16	36	38	33	54	40	28	43	106	62	39	27
1	—														
2	0.149	—													
3	0.218	0.254	—												
4	0.356	0.347	0.087	—											
5	0.268	0.280	0.096	0.163	—										
6	0.340	0.356	0.126	0.223	0.092	—									
7	0.319	0.318	0.124	0.170	0.065	0.111	—								
8	0.195	0.235	0.111	0.199	0.108	0.169	0.167	—							

Cluster	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
9	0.220	0.253	0.122	0.221	0.132	0.160	0.177	0.058	—						
10	0.164	0.193	0.097	0.189	0.130	0.169	0.159	0.084	0.111	—					
11	0.217	0.252	0.115	0.224	0.142	0.168	0.179	0.112	0.138	0.057	—				
12	0.156	0.182	0.068	0.139	0.086	0.115	0.117	0.045	0.064	0.047	0.075	—			
13	0.214	0.236	0.143	0.204	0.177	0.228	0.195	0.084	0.119	0.121	0.169	0.091	—		
14	0.190	0.235	0.101	0.187	0.144	0.175	0.190	0.065	0.092	0.088	0.114	0.055	0.092	—	
15	0.267	0.308	0.156	0.241	0.165	0.188	0.220	0.124	0.148	0.138	0.120	0.110	0.193	0.132	—
16	0.292	0.300	0.158	0.184	0.165	0.221	0.190	0.139	0.169	0.143	0.183	0.106	0.176	0.154	0.1
17	0.230	0.253	0.112	0.171	0.116	0.198	0.172	0.104	0.117	0.095	0.155	0.080	0.139	0.121	0.1
18	0.278	0.283	0.144	0.183	0.158	0.205	0.193	0.130	0.151	0.113	0.152	0.095	0.182	0.138	0.1
19	0.202	0.215	0.136	0.231	0.186	0.227	0.218	0.122	0.130	0.090	0.150	0.095	0.135	0.116	0.2
20	0.185	0.197	0.080	0.149	0.120	0.153	0.133	0.083	0.084	0.052	0.085	0.034	0.107	0.085	0.1
21	0.206	0.217	0.114	0.202	0.159	0.230	0.194	0.107	0.133	0.091	0.129	0.074	0.148	0.111	0.1
22	0.212	0.210	0.168	0.245	0.187	0.248	0.222	0.135	0.161	0.098	0.166	0.116	0.154	0.148	0.2

The *p-value* for all comparisons are significant  $< 0.001$ .

Table 4

Mantel correlogram

kilometer class	class index (km)	number of distances	Mantel correlation	Pr(Mantel)	Pr(corrected)
[0 -5]	2.5	50	0.18628	0.003	0.003 **
[5 -82]	43.5	296	-0.14345	0.136	0.136
[82 -200]	141.0	118	0.15797	0.165	0.272
[200 - 400]	300.0	88	-0.12754	0.245	0.408

Test performed between genetic and geographic distance at different geographical scales, defining kilometer classes at the intra-micro-basin (0 - 5 km) and inter-micro-basin (5-82 km, 82-200 km, 200-400 km) levels, considering their positioning on the hydrographic network (Ucayali, Marañon, Amazon).

## FIGURE LEGEND

**Figure 1** . Elevation map of: **a)** the complete study area, Loreto (Peru) with the 26 sample sites, **b)** the sampling area in Jenaro Herrera region. Each stream network (micro-basin) connected to Ucayali, Marañon or Amazon rivers are identified by one letter from A to K and a colour code. Each sample site is identified by a number and preceded by the letter and the colour of the corresponding micro-basin.

**Figure 2.** Illustration of the 4 steps-analysis with STRUCTURE and the histograms of the corresponding clusters according to de higher value of  $\Delta K$ .  $Q$  = mean ancestry to cluster. The red stars below clusters indicate deviation to panmixia.  $[?]$  = sum of clusters found for each step analysis.

**Figure 3.** Distributions of the 22 clusters identified by STRUCTURE in the sampling sites of *Apistogramma agassizii* . The pie charts show the proportion of each cluster in a sampling site. Each cluster is identified by its number and its colour code.

**Figure 4** . Dendrogram of genetic distances between clusters using Nei (1972) genetic distance and a Neighbour joining construction. Numbers in dark blue, within the dendrogram, correspond to clusters. Only node values greater than 60 are shown. The pie charts illustrate in which sampling sites the clusters



were identified and in which proportion. Each sampling site is identified by its own colour code. Each large coloured triangle corresponds to a micro-basin (A to K).

**Figure 5.** Comparison of the geographical area covered by the present study in Peru (red ellipse) with those of Willis et al (2015, free-form) and Amado et al (2011, rectangle).









