Historical biogeography of South American freshwater fishes
Nicolas Hubert* and Jean-François Renno

ABSTRACT

Aim To investigate biogeographical patterns of the obligate freshwater fish order Characiformes.

Location South America.

Methods Parsimony analysis of endemicity, likelihood analysis of congruent geographical distribution, and partition Bremer support were used.

Results Areas of endemism are deduced from parsimony analysis of endemicity, and putative dispersal routes from a separate analysis of discordant patterns of distribution.

Main conclusions Our results demonstrate the occurrence of 11 major areas of endemism and support a preferential eastern–western differentiation of the characiforms in the Amazonian region, contrasting with the southern–northern differentiation of terrestrial organisms. The areas of endemism identified seem to be deeply influenced by the distribution of the emerged land during the late Miocene and allow us to hypothesize the existence of eight aquatic freshwater refuges at that time. The raw distribution of non-endemic species supports nine patterns of species distribution across the 11 areas of endemism, two of which support a southern–northern differentiation in the eastern part of the Amazon. This result shows that the main channel of the Amazon limited dispersal between tributaries from each bank of the river. The levels of endemism further demonstrate that the aquatic freshwater refuges promoted allopatric speciation and later allowed the colonization of the lowlands. By contrast, the biogeographical pattern found in the western part of the Amazon is identified as a result of the Miocene Andean foreland dynamic and the uplift of the palaeoarches that promoted allopatric divergence across several sedimentary basins by the establishment of disconnected floodplains. The assessment of conflicting species distributions also shows the presence of seven putative dispersal routes between the Amazon, Orinoco and Paraná rivers. Our findings suggest that, rather than there being a single predominant process, the establishment of the modern South American freshwater fish biotas is the result of an interaction between marine incursions, uplift of the palaeoarches, and historical connections allowing cross-drainage dispersal.

Keywords Area of endemism, characiformes, dispersal routes, maximum likelihood, Neotropics, parsimony analysis of endemicity, partition Bremer support.

INTRODUCTION

Dealing with complex palaeogeographical histories is a problem of major importance in biogeographical studies. Since the superposition of palaeogeographical events may produce multiple changes in species range distributions, highly complex patterns of animal and plant distributions are to be expected (Nelson & Platnick, 1981; Brown & Gibson, 1983; Myers &
Giller, 1988). As changes may interact and produce an apparent discordance between species distributions, the biogeographical relationships of areas are often difficult to trace back in time (e.g. McLennan & Brooks, 2002; Brooks & Van Veller, 2003). The Neotropical region exhibits such a complex biogeographical pattern, and several hypotheses based on different processes have been proposed to explain the origin of tropical species richness and patterns of distribution in South America. These include the gradient hypothesis (steep environmental gradient; Endler, 1977, 1982), the palaeogeography hypothesis (geological changes resulting from the Andean foreland dynamic during the Tertiary; Räsänen et al., 1990, 1992; Patton et al., 1994; Hoorn et al., 1995; Patton & Da Silva, 1998), the river hypothesis (fragmentation of terrestrial biota subsequent to the final establishment of the Amazon drainage during the late Tertiary; Wallace, 1852; Patton et al., 1994; Bates et al., 1998; Hall & Harvey, 2002), the disturbance–vicariance hypothesis (climatic cooling resulting in an ecologically unsuitable habitat; Colínvaux, 1987, 1993; Colínvaux et al., 1996), the refuge hypothesis (climatological fluctuations of the Pleistocene leading to alternative fragmentations and coalescences of the forestry ecosystem; Haffer, 1969, 1997; Prance, 1982, Whitmore & Prance, 1987), the river-refuge hypothesis (refuges enhancing allopatric differentiation across rivers; Ayres & Clutton-Brock, 1992; Haffer, 1997), and the museum hypothesis (species originating by allopatric differentiation in stable mountain forests during marine highstand and later accumulating by dispersal in the lowlands, which act as 'museums'; Fjeldså, 1994; Roy et al., 1997; Nores, 1999). Although arising from different fields of biogeography and involving processes at different scales of space and time (see Bush, 1994; Haffer, 1997; Tuomisto & Ruokolainen, 1997; Bates et al., 1998; Nores, 1999; Racheli & Racheli, 2004 for a review), all the previous authors recognize some concordances between known geological and climatic events and distribution patterns.

**Palaeogeographical context**

Marine incursions constitute some of the important events that shaped the Amazonian biotas (Bates et al., 1998; Lovejoy et al., 1998; Nores, 1999, 2004; Hall & Harvey, 2002; Boeger & Kritsky, 2003; Donato et al., 2003). The last event of great marine incursion, before the final establishment of the Amazon, was previously dated between 15 and 10 Ma (Fig. 1a) and was postulated to lead to a 150-m marine highstand (Haq et al., 1987; Hoorn, 1996; Räsänen & Linna, 1996). At least one continental sea, the Paranean Sea in southern Argentina, was formed (Marshall & Lundberg, 1996). Evidence of freshwater was recorded from the upper part of the Amazon c. 11 Ma (Hoorn, 1994; Potter, 1994; Wesselingh et al., 2002) in the location of the hypothesized Lago Pebas (Lundberg et al., 1998). At this time, the direction of the streams in the Magdalena Valley shifted to the west in relation to an uplift stage of the north-western Andes, and this new range was high enough to isolate the Magdalena basin (Hoorn et al., 1995). Marine regressions and Andean foreland dynamics between 10 and 8 Ma are associated with the final establishment of the Amazon basin. The Paraná–Paraguay split from the proto-Amazon at 10 Ma (Fig. 1b), but there is evidence of headwater-capture events of the Paraná system by the Amazon occurring within the last 10 Myr (Räsänen et al., 1995; Lundberg et al., 1998). Separation of the Orinoco occurred on the Vaupés arch between 8 and 5 Ma (Hoorn, 1993; Hoorn et al., 1995). The final breakthrough of the Amazon River towards its modern course occurred with the final uplift of the central Andean cordillera (Gregory-Wodzicky, 2000), related to the rise of the Purus arch (Fig. 1c) (Lundberg et al., 1998). Concomitantly, the final uplift of the north-western Andes led to the establishment of Maracaibo Lake (Hoorn, 1993; Hoorn et al., 1995). Evidence of sea-level fluctuations provides an estimate of a 100-m marine highstand at 5 Ma (Fig. 1d) with a duration of c. 800,000 years (Haq et al., 1987, 1990; Nores, 1999, 2004 for a review). The Upper Amazon was isolated from the remainders of the Orinoco and Paraná rivers located in the Brazilian and Guyana shields (Irion et al., 1995; Nores, 1999). After marine regressions, Andean foreland dynamics led to the fragmentation of the Upper Amazon, with several arches forming and the final establishment of the main tributaries during the last 4 Myr (Fig. 1e) (Räsänen et al., 1987, 1990, 1992; Hoorn et al., 1995).

**Diversification of Amazonian fishes**

The South and Central American rivers host the most diverse freshwater fish fauna of the world, with c. 6000 of the world’s approximated 13,000 species (Reis et al., 2004). Although the increase of anthropogenic perturbations of both terrestrial and aquatic ecosystems has made the Amazon a priority for conservation (Myers et al., 2000; Bates & Demos, 2001; Laurance et al., 2002; Veríssimo et al., 2002; Ferraz et al., 2003), biogeographical patterns among Neotropical fishes and the underlying forces that generated them are still far from understood. Morphologically based studies of the phylogenetic biogeography of South American freshwater fishes began two decades ago (Weitzman & Weitzman, 1982; Vari, 1988, 1989a,b; Vari & Weitzman, 1990). Such studies supported the idea that an important diversification of the Neotropical fishes predated the Pleistocene and thereby challenged the refuge hypothesis as a major explanation of biogeographical patterns for taxonomic levels higher than population. Recent insights from a palaentontological perspective confirmed that the fauna was essentially modern across a wide taxonomic and ecological range during the late Miocene (25–5.2 Ma) (Gayet & Meunier, 1998; Lundberg et al., 1998; Malabarba, 1998). Lundberg et al. (1998) highlighted that the establishment of the major drainages of the Neotropics (Amazon, Orinoco, Paraná) occurred during the last 10 Myr, and supported the role of the Andean foreland dynamic and palaearcôes in shaping modern watersheds by vicariance and headwater-capture events (hydrogeological hypothesis; Montoya-Burgos, 2003). This first set of hypotheses has recently been tested by...
Figure 1 Evolution of South American river systems during the last 15 Myr according to Gregory-Wodzicky (2000), Haq et al. (1987), Hoorn (1993, 1996), Hoorn et al. (1995), Lundberg et al. (1998), Marshall & Lundberg (1996), Nores (1999), Potter (1994), Räsänen et al. (1987, 1990, 1992, 1995), Räsänen & Linna (1996), and Wesselingh et al. (2002). (a) Marine incursions and continental lakes from 15 to 10 Ma. (b) Marine regressions and continental rivers from 10 to 8 Ma. (c) Final establishment of the Amazon, Paraguay and Orinoco rivers from 8 to 5 Ma. (d) Marine incursions from 5 to 4.2 Ma. (e) Modern South American geomorphology and hydrologic systems. Sedimentary basins are in bold characters.
an increasing number of molecular phylogenetic and phylo-
geographical studies (Bermingham & Martin, 1998; Lovejoy &
De Araújo, 2000; Sivasundar et al., 2001; Montoya-Burgos,
2003; Turner et al., 2004; Moyer et al., 2005). Such studies
confirm that an important diversification stage, related to
multiple dispersal events, preceded the Pleistocene climatic
fluctuations, and further emphasize that the biogeographical
history of the South American freshwater fishes might be far
more complex than expected. Although fossil records support
the idea that the characiform fauna was almost modern at
10 Ma (Gayet & Meunier, 1998; Lundberg et al., 1998;
Malabarba, 1998), the relative impacts of the palaeogeographi-
cal events are still debated at the community level, and the
biogeographical patterns of Amazonian fishes are still to be
eucidated.

Detailed and robust phylogenetic hypotheses for most of the
South American fishes are still scarce. However, there is an
increasing amount of distribution data available, which may be
of practical use for conservation as well as biogeographical
purposes (Rosen, 1988; Cracraft, 1991; Morrone, 1994). Among
the 68 families of teleost fishes occurring in South
American rivers, at least 14 belong to the primary and obligate
freshwater order Characiformes (Reis et al., 2005). Most of the
lineages of this order exhibit a high level of endemism (Ortí &
Meyer, 1997; Calcagnotto et al., 2005; Hubert et al., 2005a,b),
and, thereby, constitute a model of choice for the study of the
historical biogeography of South American freshwater fishes.

Of the hypotheses of diversification in South America, we
especially focussed here on the hypotheses relying on vicariant
and dispersal processes. The palaeogeography hypothesis, river
hypothesis and museum hypothesis involve either the raising
of the palaeoarches or hydro-morphological changes or the
Miocene marine incursions as major vicariant events promot-
ing allopatric speciation (Table 1). Hence, we applied parsimi-
ony analysis of endemicity (PAE) and a likelihood approach
to a data set of characiform species distribution to investigate:
(1) how many areas of endemicity (sensus Rosen, 1988;
Morrone, 1994, 1998) for the Characiformes occur in South
American rivers, (2) which predictions from the hypotheses
proposed for the origin of the South American diversity are
corroborated, (3) whether non-endemic widespread species
across the areas of endemisms follow general patterns of
distribution across South America, and (4) whether the general
patterns of non-endemic species distribution reflect recent
dispersal events and connections between the major drainages.

We compiled distributional data of characiform taxa from a
broad taxonomic and ecological range that have recently
undergone a systematic revision to produce general historical
biogeographical hypotheses.

**METHODS**

**Area**

Since we aimed at identifying areas of endemicity and dispersal
routes, operational hydrological units (Fig. 2) were defined
using the following criteria.

1. When consistent faunistic data were available, coastal
   drainages were treated as separated units [Atrato (Atr),
   Magdalena (Mag), Essequibo (Ess), Courantijn (Cou), Surin-
   nam (Sur), Maroni (Mar), Mana (Man), Sinnamary (Sin),
   Approuague (App), Opapock (Oya), Parnaiba (Prm), and
   grouped when not [drainages from Maranhão (Mrn), Planalto
   (Pla) and Paraíba do Sul (Pab)].

2. Since the tributaries of a given drainage may have
   undergone distinct headwater-capture events, each tributary
   was treated as a separated unit [Parguay (Parg), Paraná (Parn),
   Uruguay (Ur) for the Paraná–Paraguay River; Tocantins
   (Toc), Araguaia (Ara), Xingu (Xin), Tapajós (Tap), Guaporé
   (Gua), Mamoré (Mam), Béni (Be), Madré de Dios (Mdd),

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Events</th>
<th>Allopatric process</th>
<th>Predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Palaeogeography</td>
<td>Rise of palaeoarches due to the Andean foreland</td>
<td>Speciation by vicariance due to palaeoarches</td>
<td>Areas from each side of palaeoarches should harbour differentiated biotas and endemic species</td>
</tr>
<tr>
<td>River</td>
<td>Hydro-morphological changes</td>
<td>Speciation by vicariance due to impassable major rivers</td>
<td>Widespread species should occur on one of the banks of the river and not on the other</td>
</tr>
<tr>
<td>Museum</td>
<td>Miocene marine incursion</td>
<td>Speciation by vicariance due to marine incursion</td>
<td>The lowlands should harbour a high number of species and a lower level of endemicity than the emerged lands during the Miocene marine incursion Presence of freshwater refuges in the highlands</td>
</tr>
<tr>
<td>Hydrogeology</td>
<td>Headwater capture events and dispersal routes due to hydro-morphological changes</td>
<td>Dispersal and post-dispersal speciation</td>
<td>Different closely related species or the same species should occur on contiguous headwaters linked by actual or historical dispersal routes</td>
</tr>
</tbody>
</table>

**Table 1** Predictions evaluated in this study derived from four hypotheses of Amazonian diversification regarding the biogeographical differentiation of Characiform fish communities in South America
The geographical distribution of the characiform species included here was compiled from current revisions published by the specialists of each group: 67 species of Curimatae (Vari, 1982a, b, 1984, 1989a, b, 1991, 1992a, b; Vari & Reis, 1995), 7 of Chilodontidae (Vari et al., 1995; Vari & Ortega, 1997), 21 of Prochilodontidae (Vari, 2004), 132 of Anostomidae (Garavello & Britski, 2003), 28 of Hemiodontidae (Langeani, 2003), 23 of Parodontidae (Pavanelli, 2003), 69 of Crenuchidae (Buckup, 1993, 2003), 7 of Ctenoluciidae (Vari, 1995), 10 of Erythrinidae (Oyakawa, 2003), 59 of Lebiasinidae (Weitzman & Weitzman, 1982, 2003), 14 of Cynodontidae (Menezes & Lucena, 1998; Géry et al., 1999; Toledo-Piza et al., 1999; Toledo-Pizza, 2000), 14 of Aestrorhynchidae (Menezes, 1969, 2003), 9 of Gasteropelecidae (Weitzman, 1960; Weitzman & Palmer, 2003), 111 of the family Characidae, 12 of which belong to the subfamily Stethacanthinae (Reis, 1989), 6 to the genus Creagrutus, 1 to the genus Piabina (Vari & Harold, 2001), and 33 to the subfamily Serrasalminae (including only Serrasalmus, Pygocentrus, Catoprin, Pristobrycon and Pygopristis; Jegu et al., 1991; Fink & Machado-Allison, 1992, 2001; Fink, 1993; Jegu, 2003). The Fishbase (Foese & Pauly, 2000) and Neodat (http://neodat.org/) data bases provided complementary information. Although these two internet data bases include some doubtful localities owing to taxonomic and nomenclatural bias, citations outside of the known distributional ranges following the Check List of Freshwater Fishes of Central and South America (Reis et al., 2004) were discarded for the sake of security. Additional data were selected according to the availability of ichthyological atlases or inventories in order to refine some distributional ranges, and information was added from Dos Santos et al. (1984), Lauzanne et al. (1991), Machado-Allison et al. (1999), Navarro & Maldonado (2002), Ortega & Vari (1986), Planquette et al. (1996), Sarmiento & Barrera (1997), and Ten et al. (2001).

The data collected formed a data matrix compiling the presence (marked 1) or the absence (marked 0) in each hydrological unit of 601 species belonging to the 14 Neotropical families of characiforms (Buckup, 1998). The data matrix is available in Appendix S1 (see Supplementary Material).

Analyses

Three distinct analyses were performed (Fig. 3). Vicariant events are likely to produce some substantial discontinuities in species distributions, and endemic species might be expected to support clades of areas, which are considered as areas of endemism assessed using the PAE in analysis 1 (Fig. 3; analysis 1: characterization of areas of endemism). Similarly, historical connections and dispersal routes are likely to affect a great number of species and produce some general patterns of distributions crossing the areas of endemism. Once endemic species were detected, we further checked the presence of congruent geographic distribution (CGD) within species.
shared by several areas of endemism. The detection of the CGD was assessed by means of a likelihood analysis in analysis 2 (Fig. 3; analysis 2: detection of congruent geographic distributions). Dispersal routes are likely to induce concomitant dispersal events between areas of endemism, thereby leading to species with contradictory distributions and conflicting area-branching support. By contrast, the oldest connections could precede vicariant events, and subsequent differential extinctions might be expected to produce no highly supported alternative area branching, since extinction corresponds to the loss of information about historical area connections. Hence, shared species within the study area may be considered to be analogous to common derived characters (synapomorphies) in phylogenetic systematics, indicating a common history (Zandee & Roos, 1987). In a phylogenetic framework, the recognition of common derived character states relies on a two-step process: definition of primary homology (H0: *a priori* assumption that characters have the same ontological origin among taxa), and the secondary homology hypothesis (H1: *a posteriori* assumption that derived character states are shared among taxa by ancestry, namely synapomorphy) (Tassy & Darlu, 1993). Although H0 relies on an *a priori* statement, H1 is inferred *a posteriori* by means of the cladistic analysis, which allows for discrimination between ancestral and derived character states using a taxonomic outgroup. For biogeographical purposes, H0 requires that species used are well recognized and constitute monophyletic units (Bates et al., 1998; Unmack, 2001), and H1 requires the recognition of shared species among areas by ancestry, which is achieved by rooting the area cladogram with a hypothetical ancestral area where all the species are absent (Cracraft, 1991).

The presence of shared taxa would reflect how areas were historically connected to one another when vicariance is the predominant process. Actually, species might be shared among areas as a result of recent dispersal events, and faunistic affinities might reflect recent rather than ancient connections precluding vicariant events. Similarly, historical connections might not be detected if species distributions were partially shaped by local extinction events. As a result of this statement, Brooks & Van Veller (2003) and Humphries & Parenti (1999) dismissed PAE as an historical method, since it is susceptible to being misled by shared episodes of dispersal or extinction. Although it is worth noting that PAE is not designed to be an alternative to historical phylogeny-based methods (Porzecanski & Cracraft, 2005), it constitutes a useful first step for historical biogeographical analyses when phylogenetic information is scarce (e.g. Bates et al., 1998; Ron, 2000; Trejo-torres & Ackerman, 2001; Unmack, 2001; García-Barros et al., 2002; Morrone & Escalante, 2002; Aguilar-Aguilar et al., 2003; Manrique et al., 2003; Racheli & Racheli, 2004). In addition, if patterns of area branching and grouping are compared with evidence from the Earth sciences, an historical interpretation may be given. Furthermore, if general and discordant patterns of distribution are to be detected, these apparent conflicting
data may be treated separately and a temporal connection may be detected (Bates et al., 1998).

Parsimony analysis of endemity was applied to the characiform data matrix of the 345 parsimony-informative species among the 49 hydrological units to detect areas of endemism (Fig. 3; analysis 1). Maximum parsimony (MP) heuristic searches were performed by random addition of area and tree-bisection reconnection (TBR) branch swapping using PAUP* (Swofford, 1999). Trees were rooted using a hypothetical ancestral area with all species absent, and searches were replicated 100 times. Multiple apparition-extinction events can produce homoplasy in parsimony reconstruction, and the extent to which homoplasy affected the inferences was evaluated by means of the consistency index (CI), defined as the ratio of the theoretical minimum number of changes required to account for the number of observed character states in the data to the inferred number of changes. In order to produce inferences relying on the species less affected by extinction, MP analyses were replicated with characters weighted according to their CI. Following this procedure, the most homoplastic characters (species) were down-weighted in the analyses. Nodal support was assessed by performing bootstrap proportion (BP) analyses with 1000 pseudoreplicates (with 10 random additions of area and TBR branch-swapping). According to the topology observed, areas of endemism were identified as the most inclusive sets of areas identified with high statistical confidence. Following this criterion, only large areas supported by a substantial number of endemic species were retained, even if nested subsets were statistically well supported.

Analysis 2: Detection of congruent geographical distributions

As we aimed at identifying CGD among the non-endemic species distributed across the areas of endemism identified in analysis 1 and assigning species to patterns simultaneously, we applied the following likelihood algorithm (Fig. 3; analysis 2). Consider the case in which we are testing the hypothesis of two distinct clusters ($k = 2$) among $i$ species, with $n_i$ the number of sampled species from cluster 1 and $n_k$ that from cluster 2. Each species $i$ ($i = 1$ to $n$) has an a priori probability $n_i/n$ of belonging to cluster 1, and $n_k/n$ of belonging to cluster 2. This implies that, without any prior knowledge, a given species has an a priori higher probability of being sampled from the cluster with the higher number of species. Given the occurrence of the $n$ species of each cluster within the areas considered, the probability $a posteriori$ of a species $i$ being drawn from a cluster $k$ for an area $x$ is calculated as $P_i = n_i/n_x$, where $n_i$ is the number of species occurring in the area $x$, and $n_k$ the number of species in the cluster $k$. This implies that the probability of a species $i$ being observed in area $x$ to be sampled in a cluster $k$ is proportional to the number of species from cluster $k$ in area $x$. Hence, a species $i$ in area $x$ is more likely to be assigned to cluster $k$ with the highest number of species in area $x$. These values are calculated for each area and are then multiplied to give the likelihood of a species belonging to the species cluster $k$ ($L_k$). Thus the probability of drawing a species with a given distribution across the area is $L_i = (n_i/n \times L_{i1}) + (n_k/n \times L_{i2})$, where $L_{i1}$ and $L_{i2}$ are the probabilities of drawing a species from cluster 1 or 2, respectively. Then, the likelihood of the separation of the species pool in two distinct distribution patterns or species clusters is given by the sum of the $n$ species likelihood scores $L_i = \sum L_i$. The presence of $k$ species clusters is tested in the same way for $k > 2$.

This algorithm is currently available for individual assignment purposes based on multi-locus data in population genetics as implemented in PartitionML (Castric et al., 2002). The best assignment of the $i$ species to the $k$ clusters is obtained by a permutation procedure of species among clusters. Once likelihood scores for the best assignments were computed for an increasing number of clusters ($k = 1$ to $m$), the value of $k$ with the greatest likelihood score was detected by means of a likelihood ratio test (LRT; Smouse et al., 1990). In order to fit the input format of PartitionML, areas were considered as loci, and the presence and absence of species were coded as alternative alleles.

Analysis 3: Detection of contradictory distributions

Alternative schemes of relationships that might contradict the areas of endemism defined in analysis 1 were detected by quantifying the relative contribution of endemic and non-endemic species to the area cladograms obtained (Fig. 3; analysis 3). Area relationships inferred in PAE were reassessed among the areas hosting several distinct CGDs. The likelihood approach aforementioned was applied to the whole set of informative species distributed in the areas selected in order to define the local species clusters. The relative support of each cluster to the area branching was further assessed by means of the partition Bremer support (PBS, Baker & DeSalle, 1997; Baker et al., 1998), which determines the additive contribution of each data set (positive or negative) to the total Bremer support index (BSI) value at each node of the cladogram (Gatesy et al., 1999). To calculate the BSI, MP searches are constrained to exclude a given node from the search, and the number of extra steps inferred gives the relative support for the node. When several data sets are involved, the relative contribution of each data set is given by comparing the length of a data set on the tree(s) not containing the specified node. The difference in these two lengths provides the contribution of a given data set to the simultaneous analysis support at that node (Baker & DeSalle, 1997). We used TreeRot (Sorenson, 1999) and PAUP* to calculate these indices.

Once species clusters carrying conflicting signals were revealed by negative PBS values, separate PAEs were performed for each set of conflicting species clusters. Alternative area branching involving hydrological units from distinct areas of endemism were identified. When complex PBS patterns were found, a phenetic approach was used to detect general patterns of PBS. We used ADE-4 (Thioulouse et al., 1997) to compute the dendrogram (Ward divisive algorithm).
RESULTS

Analysis 1: Characterization of areas of endemism

The MP analysis of the parsimony-informative species (345 species) yielded six most-parsimonious trees of 1066 steps (CI = 0.324) from which four were retained when characters were weighted according to their CI (Fig. 4a). The majority-rule consensus consisted of a basal polytomy as a result of the lack of resolution for the positions of San Juan (SJ) and São Francisco (SF) + Parnaiba (Pab) (Fig. 4a). Among this set of 49 hydrological units, eight clades were well supported as Paraná–Paraguay (I); São Francisco (II); Upper Amazon (IIIA); Parnaiba (IIIBa); Guyana (IIIBb); Tocantins–Xingu (IIIBd); Lower Amazon (IIIBe); Orinoco–Upper Negro (IIIBf); Atrato–Maracaibo (IV); and San Juan (V) (Fig. 4b). Among this set of 11 areas of endemism at least five areas were closely related to the Precambrian shields: I, II, IIIBa, IIIBd in the Brazilian shield and IIIBf in the Guyana shield (Fig. 4b). The area IIIBe partially overlapped the Brazilian and Guyana shields.

The Amazon River was distributed across four clades, with the first grouping drainages from the upper Amazon (IIIA), the second with drainages from the Brazilian shield (IIIBd), the third with drainages from the lower course of the Amazon, and, finally, the Upper Negro nested with the three hydrological units from the Orinoco. Although most of these clades were well supported (except for IIIBf and IIIBe due to the only unsupported positions of Ama/Tro and Ama/Bra), relationships among these 11 areas of endemism were poorly supported, mostly as a result of the lack of support for the relationships among the areas of endemism from the Amazon. In contrast, all the tributaries from the Paraná–Paraguay, Orinoco, north-western drainages, São Francisco + Paraiba do sul and the coastal drainages from the Guyana shield constituted highly supported clades (I, IIIBf with Ori/Mat + Ori/Bar + Ori/Llan, IV and IIIBb), suggesting old area splitting. Among the Upper Amazon (area of endemism IIIA), at least three subclades were supported corresponding to the Bolivian drainages of the Upper Madera (Ama/Gua + Ama/MdD + Ama/Mam + Ama/Be; BP = 99), the Peruvian drainages of the Upper Solimões (Ama/Sol + Ama/Uca + Ama/Bru).

![Figure 4](image-url)

Figure 4 Characterization of areas of endemism using parsimony analysis of endemicity (analysis 1). (a) Majority-rule consensus of the four most-parsimonious trees with characters weighted according to their CI and trees rooted with a hypothetical ancestral area where all species are absent (all the nodes resolved in the consensus were observed in the four trees). Numbers above branches refer to bootstrap proportion values obtained by computing 800 pseudo-replicates and searches replicated 10 times with random addition of areas and tree-bisection reconnection branch-swapping. (b) The 11 areas of endemism identified by the maximum parsimony analyses (Paraná–Paraguay, I; São Francisco, II; Upper Amazon, IIIA; Parnaiba, IIIBa; Guyana, IIIBb; Maranhão, IIIBc; Tocantins–Xingu, IIIBd; Lower Amazon, IIIBe; Orinoco–Upper Negro, IIIBf; Atrato–Maracaibo, IV; San Juan, V).
Mro + Ama/Jur; BP = 60), and the clade Ama/Caq + Ama/Put (BP = 92). The Guyana (area of endemism IIIBb) included two main lineages with the western drainages grouped together (Ess + Cou + Sur; BP = 98) against the eastern ones (Mar + Man + Sin + App + Oya; BP = 51). As we aimed to identify major breaks in species distribution and because these subsets of area were strongly related to others, these subclades were not retained as areas of endemism.

Analysis 2: Detection of congruent geographical distributions

The likelihood approach, applied to the matrix of the 200 non-endemic species across the 11 areas of endemism, unambiguously identified nine CGDs (LRT; d.f. = 11; \( p_k = 9 \rightarrow 9 = 0.0009; p_k = 9 \rightarrow 10 = 0.996 \)) corresponding to distinct distribution patterns (Fig. 5; Table 2). CGD 1 was restricted to the coastal drainages of the areas II, IIIBa and IIIBc, and most of its species were observed in the Maranhão (IIIBc). CGD 2 was almost exclusive to the Amazon (IIIA, IIIBd, IIIBe), and only a few species were also observed in the Guyana shield (IIIBb) and the Maranhão (IIIBc). CGD 3 was restricted to the Pacific drainages of the San Juan area (V), the periphery of Maracaibo Lake and Atrato (IV), the Orinoco and partially to the Amazon as it occurred in the Negro and Trombetas rivers (IIIBb). CGD 4 was widely distributed in the eastern drainages from the Guyana shield (IIIBb) and Orinoco (IIIBf), where almost all of the species were observed, up to the Brazilian shield (IIIBd) and lower course of the Amazon (IIIBe). CGD 5 was largely distributed throughout the continent but was absent from the Brazilian shield and Maranhão, while most of the species occurred in the Paraná–Paraguay. CGD 6 and CGD 9 were very similar but most of the species of CGD 6 were observed throughout the Amazon (IIIBd, IIIA, IIIBf, IIIBe), the Guyana shield (IIIBb) and the Paraná–Paraguay (I), while the species of CGD 9 were lacking in the Guyana shield and some of them were not observed in the Brazilian shield (IIIBd). CGD 7 was largely distributed but absent from the Paraná–Paraguay and the coastal drainages of the Maranhão and São Francisco (II), while most of the species were observed in the Orinoco (IIIBf) and the Guyana shield (IIIBb). Finally, CGD 8 was observed in the 11 area of endemism.

The number of endemic species varied greatly among the 11 areas of endemism (Table 2) and was higher for the Paraná–Paraguay (I), Upper Amazon (IIIA), Guyana (IIIBb), Tocantins–Xingu (IIIBd) and Orinoco–Upper Negro (IIIBf). The Lower Amazon (IIIBe) and Maranhão (IIIBc) harboured a high ratio of non-endemic/endemic species, of 5.5 and 15.5, respectively (Table 2), indicating that the majority of the species occurring in the Lower Amazon and Maranhão were shared with other areas. This was in contrast to the Upper Amazon (IIIA), Parnaiba (IIIBa), Guyana (IIIBb), Tocantins–Xingu (IIIBd) and Orinoco–Upper Negro (IIIBf), where the numbers of non-endemic species were on average twice the
Table 2  Numbers of non-endemic and endemic species and ratio of non-endemic to endemic species for the nine congruent geographic distributions among the 11 areas of endemism identified in MP analyses (Fig. 1). I, Parana; II, Sao Francisco; IIIA, Upper Amazon; IIIBa, Parnaiba; IIIBb, Guyana; IIIBc, Maranhao; IIIBd, Tocantins–Xingu; IIIBe, Lower Amazon; IIIBf, Orinoco–Upper Negro; IV, Atrato–Maracaibo; V, San Juan

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>IIIA</th>
<th>IIIBa</th>
<th>IIIBa</th>
<th>IIIBb</th>
<th>IIIBc</th>
<th>IIIBd</th>
<th>IIIBe</th>
<th>IIIBf</th>
<th>IV</th>
<th>V</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-endemic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGD 1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>CGD 2</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>7</td>
<td>6</td>
<td>15</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>19</td>
</tr>
<tr>
<td>CGD 3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>7</td>
<td>4</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>CGD 4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>13</td>
<td>21</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>CGD 5</td>
<td>17</td>
<td>14</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>CGD 6</td>
<td>3</td>
<td>0</td>
<td>25</td>
<td>0</td>
<td>9</td>
<td>1</td>
<td>25</td>
<td>24</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>CGD 7</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>34</td>
<td>2</td>
<td>2</td>
<td>19</td>
<td>33</td>
<td>0</td>
<td>0</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>CGD 8</td>
<td>9</td>
<td>5</td>
<td>21</td>
<td>10</td>
<td>22</td>
<td>19</td>
<td>22</td>
<td>21</td>
<td>22</td>
<td>2</td>
<td>1</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>CGD 9</td>
<td>2</td>
<td>0</td>
<td>43</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>33</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>Total (N)</td>
<td>31</td>
<td>22</td>
<td>115</td>
<td>16</td>
<td>76</td>
<td>31</td>
<td>79</td>
<td>137</td>
<td>129</td>
<td>9</td>
<td>5</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>Endemic (E)</td>
<td>57</td>
<td>39</td>
<td>79</td>
<td>8</td>
<td>41</td>
<td>2</td>
<td>41</td>
<td>25</td>
<td>68</td>
<td>28</td>
<td>14</td>
<td>401</td>
<td></td>
</tr>
<tr>
<td>Ratio N/E</td>
<td>0.5</td>
<td>0.6</td>
<td>1.5</td>
<td>2.0</td>
<td>1.9</td>
<td>15.5</td>
<td>1.9</td>
<td>5.5</td>
<td>1.9</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>88</td>
<td>61</td>
<td>194</td>
<td>24</td>
<td>117</td>
<td>33</td>
<td>120</td>
<td>162</td>
<td>197</td>
<td>37</td>
<td>19</td>
<td>601</td>
<td></td>
</tr>
</tbody>
</table>

numbers of endemic species, the ratio ranging from 1.5 to 1.9, or the Parana–Paraguay (I), Sao Francisco (II), Atrato–Maracaibo (IV) and San Juan (V), where the numbers of non-endemic species were half the numbers of non-endemic species, the ratio ranging from 0.3 to 0.6 (Table 2), indicating that the endemic species constituted the major components of their biotas. This finding should be related to the basal and unresolved position of these four areas in the MP cladogram of the 49 hydrological units (Fig. 4a).

Analysis 3: Conflicting support

Relationships among Sao Francisco, Maranhao and Parana–Paraguay areas of endemism

CGD 1 and CGD 5 suggested that the Sao Francisco received influences from both the Maranhao and the Parana–Paraguay system. Hence, we analysed the occurrence of contradictory distributions of the species distributed across areas I (Paraná), II (Sao Francisco), IIIBa (Parnaiba) and IIIBc (Maranhao). The MP analysis yielded a single most-parsimonious and highly supported tree (L = 103; CI = 0.583) for the 10 hydrological units included (Fig. 6a). The MP tree indicated that the Sao Francisco area of endemism shared more species with the northern coastal drainages (Pla, Prn and Mrn; Fig. 6a) than with the Parana–Paraguay. The assignment procedure identified six species clusters for this subset (LRT; d.f. = 9; \( p_k = 5 - 6 = 0.0009; p_k = 6 - 7 = 0.663 \)). The PBS analyses identified conflicting support for CGD 1 and 5, mostly arising at the nodes related to the position of Sf and Pab for cluster 1 and Pla for cluster 5 (Fig. 6a). The MP analysis of these conflicting clusters yielded 45 equiparsimonious trees (L = 28; CI = 0.6), and the majority-rule consensus was fairly resolved, except for the position of Parg and Uru (Fig. 6b). This consensus mainly differed from the MP tree obtained with the whole data set by the position of the hydrological unit Sao Francisco (Sf) nested with Parana (Parn), indicating that Sf from the Sao Francisco area of endemism shared some species almost exclusively with Parn from the Parana–Paraguay area of endemism. This alternative branching involving two hydrological units from distinct areas of endemism may be related to the influence of a dispersal route (Fig. 6; arrow 1).

Relationships among Upper Amazon, Tocantins–Xingu and Parana–Paraguay areas of endemism

Despite the great differences in species composition of the biotas from the Parana–Paraguay (I), Upper Amazon (IIIA) and Tocantins–Xingu (IIIBd) areas of endemism (Fig. 4), CGDs 5 and 6 suggested species sharing among them (Fig. 5). The MP analysis yielded a single most-parsimonious and moderately supported tree (L = 317; CI = 0.419) for the 20 hydrological units included (Fig. 7a). Among the 133 parsimony-informative species, eight species clusters were identified (LRT; d.f. = 20; \( p_k = 7 - 8 = 0.021; p_k = 8 - 9 = 0.358 \)), and PBS analyses showed that all the clusters, excepting clusters 1, 4 and 5, yielded conflicting support (Fig. 7a). As complex PBS patterns were observed, we computed a dendrogram among the eight clusters depending on their PBS scores across the tree in order to detect groups of clusters sharing similar conflicts. The dendrogram identified four groups of clusters according to their support in the tree (Fig. 7b). The group including clusters 4 and 5 showed high BSI values and no conflicting support like cluster 1 did too. Clusters 2, 6 and 7 were grouped together since these clusters were in conflict for the area branching within the Upper Amazon area of endemism (Fig. 7a). Cluster 3 provided a distinct PBS pattern since it supported the area branching among the Tocantins–Xingu area of endemism (IIIBd) while providing conflicting support for the area branching among the Upper Amazon area of endemism.
endemism (Fig. 7a). Cluster 8 supported most of the area branching among the Paraná–Paraguay (I), Tocantins-Xingu (IIIBd) and Upper Amazon (IIIA) areas of endemism like cluster 1, but provided conflicting support among the Upper Amazon like clusters 2, 6 and 7.

Clusters 2, 6, 7 and 8 provided conflicting support in the same area branching, and separate MP analyses were conducted including these four clusters to understand the ichthyological relationships in this region (Fig. 8a). The MP analyses yielded four trees ($L = 91; CI = 0.319$) and the strict consensus was fairly resolved (Fig. 8a). The major alternative area branching found in the consensus was related to the position of the hydrological units Gua, MdD, Mam and Sol from the Upper Amazon area of endemism, nested with the units Parn and Parg from the Paraná–Paraguay area of endemism by contrast with the MP tree observed when all species were included (Fig. 7a). This suggested that, despite the great number of endemic species restricted to the Upper Amazon,
some species were shared exclusively between the headwaters of the Paraná–Paraguay and Upper Amazon (southern tributaries) areas of endemism, thereby suggesting the influence of a dispersal route (Fig. 8; arrow 2).

The dendrogram of the species clusters (Fig. 7b) indicated that cluster 3 harboured a distinct PBS pattern. Hence, a separate MP analysis was conducted for this cluster (Fig. 8b). MP analyses of the 16 species of cluster 3 provided 46 most-parsimonious trees ($L = 37; CI = 0.432$). The major alternative area branching found in the majority-rule consensus was related to the position of the hydrological unit Tap from the Tocantins–Xingu (IIIBd) area of endemism nested with Parg from the Paraná–Paraguay area of endemism by contrast with the MP tree from the overall data set (Fig. 7a). Similarly, the presence of shared species restricted to the headwater of the Parg and Tap suggested the influence of a dispersal route (Fig. 8; arrow 3).

**Figure 8** Parsimony analysis of endemicity of the species clusters with conflicting signal among the areas of endemism I, IIIBd and IIIA (analysis 3). (a) Strict consensus of the four trees ($L = 91; CI = 0.319$) obtained with the conflicting species clusters 2, 6, 7 and 8 (29 parsimony-informative species). Values in the consensus are bootstrap proportion (BP) values. (b) Majority-rule consensus of the 46 trees ($L = 37; CI = 0.432$) obtained for the conflicting species cluster 3 (16 parsimony-informative species). Values above branches are majority-rule scores/BP values (*node with a BP value < 50). Bold lines identify area of endemism limits, and dashed lines correspond to unsupported ambiguous branching. Area branching involving distinct drainages belonging to different areas of endemism are in grey, and the arrows represent putative dispersal routes 2 and 3.

Relationships among Guyana, Maranhão, Lower Amazon, and Orinoco–Upper Negro areas of endemism

CGD 2 suggested that the Guyana area of endemism (IIIb) was influenced by the lower Amazon (IIIBe) area of endemism, while CGDs 4 and 7 indicated that it was also influenced by the fauna from the Orinoco–Upper Negro area of endemism (IIIb). Hence, we searched for alternative area-branching support considering endemic and non-endemic species from the Guyana (IIIb), Maranhão (IIIb), lower Amazon (IIIBe), and Orinoco–Upper Negro (IIIb) areas of endemism. The MP analyses of the 181 species yielded a single most-parsimonious and fairly supported tree ($L = 397; CI = 0.456$; Fig. 9a), and the likelihood analyses identified eight species clusters ($L = 397; CI = 0.456$; Fig. 9a), and the likelihood analyses identified eight species clusters ($L = 397; CI = 0.456$; Fig. 9a). The PBS analysis indicated that the eight clusters harboured conflicts (Fig. 9a). As a complex PBS pattern was observed, a dendrogram of the eight clusters according to their PBS scores across the tree was constructed, allowing the identification of three patterns of PBS (Fig. 9b). The clusters 1, 2 and 8 were grouped together according to their support of the area branching within the Orinoco–Upper Negro (Tro, Neg, Llan, Mat). However, the conflicting support for the area branching among the Guyana area of endemism provided by cluster 1 was mostly a result of the absence of this cluster from several drainages of Guyana (data not shown). By contrast, clusters 2 and 8 were present in almost every hydrological unit and provided conflicting support for several nodes (Fig. 9a). Clusters 3 and 4 provided conflicting support for area branching in all the areas of endemism included, and harboured high global BSI values (Fig. 9b). Clusters 5, 6 and 7 were grouped together, but cluster 7 provided only minor conflict in the area branching within the Orinoco–Upper Negro area of endemism while providing positive support to the Guyana area of endemism (Fig. 9a).

As the clusters 2 and 8 together provided conflicting support for area branching in several of the areas of endemism, a separate MP analysis was performed for this group (Fig. 10a). MP searches including the 28 parsimony-informative species of clusters 2 and 8 yielded two equiparsimonious trees.
(L = 64; CI = 0.437), and the strict consensus was fairly resolved (Fig. 10a). This consensus mainly differed from the tree obtained with the overall data set by the position of the hydrological units Mat, Bar and Llan from the Orinoco–Upper Negro, nested within the Guyana area of endemism and more closely related with Ess (Figs 9a and 10a). This result suggested the influence of a dispersal route between the Orinoco–Upper Negro and Guyana areas of endemism (Fig. 10a; arrow 4). It is worth noting that a significant correlation was found between the number of changes inferred in MP among the hydrological units Mat, Ess, Cou, Sur, Mar, Man, Sin, App, Oya and Mrn (coastal drainages), and the geographical distances of their estuaries along the Atlantic coast (analysis of covariance; r = 0.547; P = 0.0026), suggesting a coastal dispersal route (Fig. 10a; arrow 5).

Similarly, clusters 3 and 4 provided conflicting support among several hydrological units from the Guyana and Orinoco–Upper Negro areas of endemism. Hence, separate MP analyses were conducted for this group, and MP searches, including the 54 parsimony-informative species, yielded six trees (L = 95; CI = 0.568). The majority-rule consensus consisted of a basal polytomy and differed from the tree obtained with the overall data set by the position of the hydrological units Lneg and Bra from the lower Amazon area of endemism and the area of endemism Maranhão (Mrn) together nested with the hydrological units Mat, Bar and Llan from the Orinoco–Upper Negro.

The position of the Ess nested with the Orinoco–Upper Negro units confirmed that species sharing occurred between the Essequibo and the Orinoco and seems to corroborate dispersal route 4 (Fig. 10a), while the positions of the Bra and Tro seem to confirm dispersal routes 6 and 7 (Fig. 10b,c). By contrast, the position of the Mrn, owing to species sharing with the Tro, Ess and the hydrological units Mat, Bar and Llan from the Orinoco suggested the influence of another dispersal route occurring in the lower part of the Amazon (Fig. 10c; arrow 8).

DISCUSSION

Congruence and discrepancies between aquatic and terrestrial biotas

Thus far, Amazonian biogeography has been intensively addressed for terrestrial animals (see Hall & Harvey, 2002 for a review). However, the recognition of similar biogeographical patterns for both terrestrial and freshwater biotas is of major importance for elucidating to what extent the palaeogeographical events that have occurred in South America have caused the extant diversity. We highlight here that the tributaries of the Amazon valley have a complex history and include several areas of endemism, as stated by Vari & Weitzman (1990). Although previous studies relying on PAE focussed exclusively on interfluvial area relationships, congruent patterns between terrestrial and aquatic biotas were detected for the delimitation of the areas of endemism.
The two closely related areas of endemism observed here in the clade IIIA are traditionally recognized in the Upper Amazon (Napo and Inambari clades; Ron, 2000; Fig. 2) based on distributional data among birds (Cracraft, 1985; Prum, 1988), butterflies (Hall & Harvey, 2002; Racheli & Racheli, 2004), anurans and lizards (Ron, 2000), and were further confirmed by several molecular phylogenetic studies (Aleixo, 2004, Hoffman & Baker, 2003). The distribution of characiform species yielded similar results, as our inferences based on PAE identified a well-supported clade in the area (Upper Amazon, IIIA) that matched the Napo + Inambari clade (Fig. 4).

The distribution of terrestrial vertebrates also supported another Amazonian area of endemism located in the Brazilian shield (Pará; Cracraft, 1985; Racheli & Racheli, 2004), anurans and lizards (Ron, 2000), and were further confirmed by several molecular phylogenetic studies (Aleixo, 2004, Hoffman & Baker, 2003). The distribution of characiform species yielded similar results, as our inferences based on PAE identified a well-supported clade in the area (Upper Amazon, IIIA) that matched the Napo + Inambari clade (Fig. 4).

Although we demonstrated a great overlap in the locations of the areas of endemism between terrestrial and freshwater biotas, relationships among them based on characiform fishes were somewhat incongruent with terrestrial biotas. This was expected, since previous biogeographical studies focussed on interfluvial area relationships, considering rivers as barriers when they actually provide dispersal opportunities among areas for fishes and aquatic vertebrates. Most of the discrepancies detected from this data set concerned the area relationships within the Amazon. The Guyana area of endemism was recognized as including both coastal drainages and inland areas for terrestrial biotas and was often closely related to the northern areas of the Amazon Valley (Hall & Harvey, 2005).

![Figure 10](Image) Parsimony analysis of endemicity of the species clusters with conflicting signal among the areas of endemism IIIBf, IIIBe, IIIBb and IIIBc (Mrm). (a) Strict consensus of the two trees \( L = 64; CI = 0.438 \) obtained with the conflicting species clusters 2 and 8 (28 parsimony-informative species). Values in the consensus are bootstrap proportion (BP) values. Bold lines identify cross-drainage limits. (b) Majority-rule consensus of the six trees \( L = 37; CI = 0.432 \) obtained with the conflicting species clusters 3 and 4 (54 parsimony-informative species). Values above branches are majority-rule scores and those below branches are BP values > 50. (c) Majority-rule consensus of the five trees \( L = 30; CI = 0.667 \) obtained with the conflicting species clusters 5 and 6 (54 parsimony-informative species). Bold lines identify cross-drainage limits and dashed lines correspond to unsupported ambiguous branching. Area branching involving distinct drainages belonging to different areas of endemism are in grey and the arrows represent putative dispersal routes 4, 5, 6, 7 and 8.
Fish communities showed that coastal drainages from Guyana were differentiated from the nearby inland tributaries and more closely related to the eastern drainages of the Amazon Valley and Orinoco.

The Upper Amazon harboured poor species sharing with the other areas from the Amazon basin, which in turn suggested that the aquatic biotas were shaped more by an eastern-western differentiation than a southern-northern one in terrestrial vertebrates (Hall & Harvey, 2002). This was related to the fact that the establishment of the lower course of the Amazon did not affect terrestrial and freshwater biotas in the same way, providing limited dispersal abilities for the former while enhancing dispersal for the latter.

**Marine incursions and freshwater refuges**

The levels of endemism found in the 11 areas of endemism described here suggested that the 100-m marine highstand that occurred c. 5 Ma deeply influenced the distribution of fish species. Previous work has emphasized such relationships between endemism and emerged land during the late Tertiary for terrestrial biotas (Nores, 1999, 2004; Hall & Harvey, 2002). Fjeldså (1994) and Roy et al. (1997) postulated that the evolution of tropical ecosystems might be driven by a dynamic process of local differentiation in the emerged lands during marine incursions and later accumulation in the lowlands during low sea-level stages. Following this hypothesis, the Tropical lowlands act as ‘museums’ where large numbers of species accumulate (museum hypothesis; Nores, 1999). This hypothesis identifies marine incursions as major vicariant events promoting divergence, and two predictions might be expected for obligate freshwater fishes following this hypothesis: (1) higher levels of endemism in the areas of endemism located in the Miocene emerged land, and (2) higher number of species in the lowlands contrasting with a low level of endemism. We observed here that the areas of endemism hosting emerged land during the late Miocene incursion (Paraná–Paraguay, São Francisco, Upper Amazon, Parnaíba, Guyana, Tocantins–Xingu, Orinoco–Upper Negro, Atrato–Maracaibo and San Juan) harboured higher level of endemism than the areas located in the lowlands (Lower Amazon, Maranhão). This result contrasted with the high number of species observed in the lower Amazon. Although the geological evidence of the Miocene incursion is scarce for South America at this time (see Nores, 1999 for a review), these results were in agreement with previous observations supporting the museum hypothesis.

The museum hypothesis predicts that freshwater refuges occurred during the Miocene marine incursions, and the distribution of the 11 areas of endemism allows us to postulate the existence of at least eight freshwater refuges related in part to the Miocene marine incursion (Fig. 11). The first is located in the north-western Andes and is associated with the Atrato–Maracaibo (IV) and San Juan (V) areas of endemism (Fig. 11; refuge I). These areas might be associated with two distinct refuges, as Vari (1988) proposed a vicariance hypothesis for the diversification of these two areas preceding the Miocene marine highstand. However, the analysis of non-endemic species distribution showed that Atrato–Maracaibo and San Juan areas share some species, as illustrated by CGD 3, and the lack of resolution of the position of San Juan in the PAE cladogram remains inconclusive concerning the biogeographical affinities of the area. Thus, until more evidence is available, the most conservative hypothesis of a single refuge is retained here. Another freshwater refuge is hypothesized in the Upper Amazon (Fig. 11, refuge V), which was already isolated.
from the Orinoco and Paraná rivers at 8 Ma and further isolated from the other parts of the Amazon at 5 Ma. Although we cannot reject the hypothesis that the differentiation of the Upper Amazon is linked to the differentiation of Lago Pebas during the Miocene, the distribution of terrestrial animals in the Amazon also matches a 100-m marine highstand at 5 Ma, which is hardly explained by the establishment of Lago Pebas (Nores, 1999). Two other refuges are postulated in the Guyana shield (Fig. 11, refuges VI and VIII), since the Orinoco and Guyana constituted two distinct areas of endemism supported by a large number of endemic species and were physically isolated during the Miocene marine incursions. Finally, four freshwater refuges are postulated for the Brazilian shield, according to the unambiguous identification of four areas of endemism in the emerged part of the shield between 5 and 4 Ma (Fig. 11, refuges II, III, IV, VII). The Paraná–Paraguay area of endemism is hypothesized here to host a distinct refuge (Fig. 11, refuge II), since geological evidence argues for a diversification of the Paraná–Paraguay River, preceding the Miocene marine highstand (Lundberg et al., 1998). However, the timing of the diversification of the Paraná and the Paraguay remains unclear. According to the large number of endemic species supporting the Paraná–Paraguayan clade, we suggested that a single refuge occurred in the area, but further evidence from other sources is needed to confirm if a single or two distinct refuges occurred. In addition, previous phylogenetic studies have argued that the São Francisco area of endemism was already isolated from the Tocantins–Xingu and Parnaiba areas of endemism at 5 Ma (Costa, 2003; Montoya-Burgos, 2003), thereby suggesting the occurrence of distinct refuges in each area (Fig. 11; refuges III, IV, VII).

The museum hypothesis argues that the colonization of the lowland is recent and after the Miocene marine incursion at 5 Ma. However, the subsequent range expansion of each species depends on its dispersal ability (Cracraft & Prum, 1988; Haffer, 1997; Nores, 1999), and several studies have emphasized that large river channels further enhanced allopatric divergences among terrestrial taxa (river barrier hypothesis; Wallace, 1852; Capparella, 1988; Ayres & Clutton-Brock, 1992; Patton et al., 1994; Peres et al., 1996; Bates et al., 1998). Although the PAE cladogram from characiform taxa does not support the southern–northern pattern of differentiation of the areas of endemism as observed for terrestrial biotas, several CGDs detected here hint that the lower Amazon acted as a partial boundary and provided limited dispersal abilities for some species of freshwater fishes between the southern and northern Amazonian drainages (CGDs 2 and 3, Fig. 5). The CGD observations supported previous works focussing on distribution or phylogenetic patterns, which state that a southern–northern shift in species distribution or genetic diversity can be attributed to the main channel of the Amazon river acting as a physical barrier to dispersal (river barrier hypothesis; Wallace, 1852; Capparella, 1991; Ayres & Clutton-Brock, 1992; Peres et al., 1996; Ron, 2000; Hall & Harvey, 2002; Hayes & Sewlal, 2004).

**Palaeoarches and flooded plains**

Andean tectonics during the late Tertiary promoted the uplift of several arches, and geological data have demonstrated their control on fluvial perturbations in the western Amazon (palaeogeography hypothesis; Räsänen et al., 1987, 1990, 1992; Hoorn et al., 1995; Irion et al., 1995). These palaeoarches were hypothesized to be responsible for the Orinoco–Amazon divide (Vaupés arch, Fig. 1), Paraná–Amazon divide (Michicola arch, Fig. 1) and the final establishment of the Western Amazon (reviewed by Lundberg et al., 1998). As mentioned by Haffer (1997), the uplift of the palaeoarches not only was responsible for forest or river fragmentation but also produced several sedimentary basins hosting disconnected and temporarily flooded plains (e.g. Madre de Dios–Beni basin, Acre basin, Ucayali basin, Pastaza–Marañón basin; Fig. 1). First evaluated in several phylogeographical studies among terrestrial mammals, high levels of genetic differentiation were found between populations across these arches (Patton et al., 1994; Patton & Da Silva, 1998). The results presented herein suggest that the palaeoarches contributed to shaping Amazonian fish communities (Fig. 11). The delimitation of the Upper Amazon area of endemism (IIIA) suggested that the Vaupes and Michicola arches enhanced allopatric differentiation in western South America, which was further influenced by marine incursions (Fig. 11, arches c and f). Furthermore, several arches in the Upper Amazon further promoted allopatric divergences, as illustrated by the differentiation of the well-supported Putumayo + Caquetá clade from the Ucayali + Marañón + Solimoes + Jurua clade related to the Iquitos–Florence arch (Fig. 11, arch d) and the Guaporé + Mamoré + Madre de Dios + Beni clade separated from the Ucayali by the Fitzcaraldo arch (Fig. 11, arch e) and from the lower Madera by the Purus and Caravari arches (Fig. 11, arches a and b). Although, the Fitzcaraldo, Purus and Caravari arches were related to the great differentiation of the Purus River, the influence of the Purus arch in differentiating the Upper Amazon from the Lower Amazon area of endemism is more ambiguous, since marine incursions concomitantly enhanced the differentiation of the Upper Amazon. However, despite a continuous connection between the Solimoes and the lower Amazon, the great differentiation of characiform fauna from each side of the arch as revealed by PAE seems to confirm its role in shaping the present biogeographical structure.

**Multiple historical connections among areas**

Palaeontological and biogeographical work suggests that fish diversification up to the level of modern species predated the Pleistocene (Weitzman & Weitzman, 1982; Vari, 1988, 1989a,b; Vari & Weitzman, 1990) and highlights the fact that the South American fish fauna was essentially modern across a wide taxonomic and ecological range during the late Miocene (25–5.2 Ma) (Gayet & Meunier, 1998; Lundberg et al., 1998; Malabarba, 1998). According to Lundberg et al. (1998), much
of the diversity of freshwater fishes might have resulted from palaeohydrological changes that promoted dispersal followed by allopatric divergence, and this theory was further formalized as the hydrogeological hypothesis (Montoya-Burgos, 2003). Following this hypothesis, several dispersal routes promoted great faunistic exchanges between the areas of endemism defined. In this study, statistically well-supported alternative schemes of area branching were found by the combined use of PAE, the likelihood approach, and PBS analysis. The hypothesis that some of these branchings were the result of a former widespread distribution followed by incomplete speciation cannot be discarded. However, this hypothesis hardly explains the occurrence of alternative clades involving only a restricted set of hydrological units from distinct areas of endemism and supported by numerous species with limited geographical distribution. To evaluate the generality of the dispersal routes inferred, comparisons with other sources of evidence are needed. The analyses of conflicting patterns among the Paraná-Paraguay, São Francisco, Parnaíba and Maranhão showed that some species were shared between the Paraguay and São Francisco rivers, supporting a new area relationship. This finding suggests the influence of a dispersal route (Fig. 11; dispersal route 1), corroborating the faunistic exchanges through the Upper Paraguay and São Francisco headwater previously depicted for some Hypostomus species c. 6 Ma during a headwater capture event (Montoya-Burgos, 2003).

The analyses of the Upper Amazon, Paraná and Tocantins-Xingu provided alternative area branching, with sister area relationships between the Paraná–Paraguay rivers and Upper Madera (Guaporé, Mamoré, Madre de Dios at least) and between the Paraguay and Tapajos rivers. Lundberg et al. (1998) reported headwater-capture events between the Upper Madera and the Paraguay, and our results were consistent with this hypothesis. The occurrence of headwater capture provided opportunities for a dispersal route between the Upper Amazon and Paraná–Paraguay areas of endemism (Fig. 10; dispersal route 2). It is worth noting that Montoya-Burgos (2003) dated a dispersal event between the headwaters of the Amazon and the Paraná–Paraguay at c. 10–12 Ma for some Hypostomus species, and Sivasundar et al. (2001) detected discrepancies between molecular and geological dating for Prochilodus, consistent with a faunistic exchange by dispersal between the Amazon and the Paraná–Paraguay.

The relationship between the Paraguay and Tapajos rivers was supported by species with restricted distributions, suggesting that faunistic exchanges occurred between the headwaters of these two rivers through the influence of a putative dispersal route (Fig. 10; arrow 3). This historical connection has not previously been documented, but might be related to the occurrence of headwater-capture events that have been reported in the Paraná (Lundberg et al., 1998) and the Paraguay (Montoya-Burgos, 2003).

Finally, in the analyses among the Orinoco–Upper Negro, Guyana, Lower Amazon and Maranhão, the Essequibo was grouped with the Trombetas and Branco in at least two MP analyses (Fig. 10b,c). Indeed, previous phylogeographical studies supported faunistic exchanges between the Essequibo and Branco. Lovejoy & De Araújo (2000) reported dispersal in the genus Potamorrhaphis between the Branco and Essequibo rivers and hypothesized that dispersal occurred through the current inundated savannah of the Rupununi, and between the Amazon and Orinoco through the Guyana drainages and the current Casiquiare River. The species sharing detected between the Essequibo and Branco is consistent with the hypothesis that the Rupununi provided extended dispersal abilities and constituted a current dispersal route (Fig. 11, dispersal route 5).

The Trombetas was frequently nested with the Essequibo in the analysis, providing evidence of species sharing between these rivers (Fig. 10b,c). Although species sharing between the Branco and Trombetas was expected owing to their close location, dispersal through the Rupununi hardly accounts for the presence of species from the Essequibo in the Trombetas. Another influence of a dispersal route may be hypothesized between the headwater of the Essequibo and Trombetas (Fig. 11, dispersal route 6) to account for faunistic exchange between the Essequibo and the Amazon. However, dispersal events through this route have not previously been reported, and the mechanism that provided dispersal abilities between the Essequibo and Trombetas remains unclear. Faunistic exchanges between the Essequibo and Trombetas through the Branco cannot be discarded, and the influence of an ancient or current dispersal route in the area will need further evidence from other sources to be confirmed.

The high statistical support for an Upper Negro nested with the Orinoco argued for high species sharing between these rivers. This result is consistent with great faunistic exchanges between the Orinoco and Negro through the Casiquiare (Fig. 11, dispersal route 7), as previously reported by Lovejoy & De Araújo (2000). However, the strong ichthyological difference between the Upper Negro and the Amazon suggested a recent headwater capture event between the two rivers.

Finally, high species sharing was detected between the Essequibo and the Orinoco (Fig. 10a–c). The significant correlation found between the number of species shared between the Orinoco and the coastal drainages from Guyana including the Maranhão and the geographical distance of the estuaries supported the hypothesis that dispersal occurred along the coast, either through coastal swamps or a freshwater layer floating along the coast. Following this hypothesis, the species sharing between the Maranhão and the Trombetas (Fig. 10c) was explained by the occurrence of coastal dispersal arising from the location of the Trombetas close to the mouth of the Amazon. Montoya-Burgos (2003), Renno (1989) and Renno et al. (1990, 1991) previously detected dispersal along the coastal of the Guyana shield from putative aquatic refuges. Following these new observations, a large dispersal route was hypothesized along the Atlantic coast of the Guyana Shield, and species sharing between the Essequibo and the Orinoco was hypothesized to occur through this route (Fig. 11, dispersal route 4).
CONCLUSION

As suggested by the majority of the biogeographers involved with the origin of South American ichthyofauna, the diversity of Amazonian fishes arose from a complex history of river systems in relation to the palaeogeographical evolution of the continent (Weitzman & Weitzman, 1982; Vari & Weitzman, 1990; Lundberg et al., 1998). Where phylogenetic information is scarce, the combined use of PAE and likelihood analysis of species distribution proved to provide new insights into the biogeography of South American fishes and to some extent identified the processes that generate their diversity. The most striking implication from the present study is the deep interaction of several palaeogeographical events that concomitantly or successively shaped the freshwater diversity of South America, highlighting the fact that it is a combination of several hypotheses, rather than a single one, that should be favoured (Bush, 1994). As mentioned, numerous hypotheses have been proposed to explain the origin of the Neotropical diversity (review in Bush, 1994; Haffer, 1997), and the present study highlights that at least four of them contributed predominantly in shaping the Amazonian primary freshwater ichthyofauna: the museum hypothesis (allopatric differentiation in the stable highlands and latter accumulation in the lowlands), the palaeogeography hypothesis (allopatric differentiation across palaeoarches prompted by the Andean foreland dynamic), the river hypothesis (fragmentation of the biotas subsequent to the final establishment of the main channel of the South American rivers), and the hydrogeological hypothesis (post-dispersal allopatric speciation resulting from multiple headwater-capture events or temporal connections among drainages).

Distribution ranges of characiform species were consistent with the hypothesis that the Vaupes and Michicola arches acted as major events promoting allopatric divergence among the Amazon, Orinoco and Paraná rivers, further enhanced by marine incursions that fragmented the river systems in at least eight freshwater refuges. The origin of the 11 areas of endemism may be attributed mainly to the rise of the palaeoarches and to the persistence of freshwater in the emerged land during the Miocene–Pliocene marine highstand. The analyses of conflicting patterns also identified six continental dispersal routes across South America and a dispersal corridor along the Atlantic coast. These dispersal routes are consistent with several molecular phylogenetic studies, demonstrating that some of these routes promoted post-dispersal speciation (Lovejoy & De Araújo, 2000; Sivasundar et al., 2001; Montoya-Burgos, 2003).

One of the most complex and intriguing patterns was found in the Lower Amazon. Assessment of endemism levels matched the prediction of the museum hypothesis, highlighting the composite origin of its ichthyofauna that probably occurred by accumulation of species originating in the northern freshwater refuge of the Guyana and the southern freshwater refuge of the Brazilian shields. By contrast, some distribution patterns also provided evidence of a barrier to dispersal across the main channel of the Amazon, which possibly limited the dispersal abilities of some fish species.

ACKNOWLEDGEMENTS

We thank Michel Jégu and Thierry Oberdorff for their longstanding interest in our work on biogeographical patterns among South American fishes. We also thank the two anonymous referees for constructive comments on the paper.

REFERENCES


**SUPPLEMENTARY MATERIAL**

The following supplementary material is available for this article online from http://www.Blackwell-Synergy.com

**Appendix S1.** Matrix of distributional data of the South American characiform species.

---

**BIOSKETCHES**

Nicolas Hubert obtained his PhD from the Institut de Recherche pour le Développement on the evolution of piranhas. His research interests include biogeography, phylogenetics and the evolution of the South American ichthyofauna. Currently he is the coordinator of the barcoding project of the North American freshwater fishes in the Louis Bernatchez Laboratory.

Jean-François Renno is a researcher at the Unité de Recherche 175 at the Institut de Recherche pour le Développement. Specializing in population genetics and evolution, he received his PhD at Montpellier University for his work on the population genetics of *Leporinus* in the Guyana shield. After spending 8 years in Africa working on the genetics of the domestication of pearl-millet, he took up his current post as the coordinator of the Amazonian part of the IRD project ‘Caractérisation et Valorisation de la Diversité Ichtyologique pour une Aquaculture Raisonnée’ (CAVIAR).

Editor: Robert McDowall