TWENTY-FIVE

Molecular signatures of Neogene biogeographical events in the Amazon fish fauna

Nathan R. Lovejoy¹, Stuart C. Willis² and James S. Albert³

¹University of Toronto Scarborough, Toronto, Ontario, Canada
²University of Nebraska-Lincoln, Lincoln, Nebraska, USA
³University of Louisiana Lafayette, Lafayette, Louisiana, USA

Abstract

Molecular genetics can contribute to biogeography by clarifying species limits (and thereby distributions) and phylogenetic relationships. Molecular data also offer the tantalizing prospect of dating the ages of lineages by adding a timescale to phylogenetic reconstructions. Thus, molecular analyses significantly enhance our ability to test models and hypotheses that address the complex relationship between biological evolution and palaeogeographic history in the Amazon drainage basin. Here, we review the use of molecular data for understanding Amazonian fish biogeography, with particular emphasis on Neogene palaeogeographic events. We provide an overview of previous work in the field, and briefly mention the possibilities and pitfalls of molecular biogeographical approaches. Challenges for molecular investigations of the Amazon fish fauna include taxon selection and sampling, molecular clock assumptions and calibration issues, and identifying clearly testable hypotheses. We provide recommendations for future investigations and methodological improvements.

Introduction

With more than 3000 species, Amazonian fishes constitute the most species-rich aquatic continental fauna on Earth (Lundberg et al. 2000; Reis et al. 2003). Although the evolutionary and ecological forces underlying the formation of this spectacular diversity are incompletely understood, advances are continually being made in the fields of historical biogeography and phylogeography. In this chapter we consider the role of molecular data sets for biogeography, at both the species and population levels. We briefly review major patterns in the geographic distributions of Amazonian freshwater fishes, summarize recent work using new molecular data sets and new methods for the analysis of biogeographical data. A central theme of this chapter is that understanding the diversification of Amazonian freshwater fishes requires detailed species-level or even population-level knowledge of geographical distributions and phylogenetic interrelationships. In turn, this biological information can only be completely understood in the context of the larger-scale climatic and geological histories of South and Middle American river basins.

Amazonia as the core of Neotropical fresh waters

In terms of taxonomic composition, the Amazonian ichthyofauna comprises the core of the Neotropical freshwater region, a vast (~17 million km²) and ecologically heterogeneous assemblage of geographical areas extending from the Isthmus of Tehuantepec in southern Mexico (16°N) to the La Plata drainage basin in northern Argentina (34°S), and including all of Central America and northern South America east of the Andes. The Neotropical ichthyofaunal region is therefore restricted to the humid tropical portions of the Neotropical realm as originally defined (Wallace 1876), excluding the cooler and/or more arid areas of the Southern Cone, the Andean Altiplano and the Pacific slopes of Chile and Peru.

The Neotropical ichthyofauna is easy to recognize. Fishes from throughout this very wide area belong to relatively few higher-level taxa, each of which is characterized by a distinct suite of morphological traits. According to current classification, the Neotropical ichthyofauna includes 17 orders of fishes; this compares with
26 orders in the Mississippi River basin and adjacent drainages. However, despite its relatively poor diversity at higher taxonomic levels, the Neotropical ichthyofauna is extremely diverse at lower taxonomic levels. Estimates of total Neotropical fish richness range between 5000 and 8000 species, compared to about 1000 species in North America (Lundberg et al. 2000). This disproportionate distribution of taxonomic categories, with many lower taxa and fewer higher taxa, has resulted from a lengthy history of geographical isolation and in situ diversification (Lundberg 1998).

As in most of the Earth's fresh waters, the Neotropical ichthyofauna is dominated by ostariophysan fishes, which include the tetrads, catfishes and electric knifefishes. Ten ostariophysan clades contain approximately 75% of all Neotropical species; the most diverse by far being the Loricarioidea (armoured catfishes and relatives, with around 1500 species currently described), and the Characoidae (tetrads and relatives, with about 1750 species currently described). Highly diverse non-ostariophysan clades include the Neotropical cichlids (500+ species) and the Cyprinodontiformes (killifishes, 600+ species). The great majority (>97%) of Neotropical freshwater fish species are members of these primary and secondary freshwater fish taxa, which have little or no tolerance for salt water and very poor capacities for dispersal over marine barriers. These fishes trace their origins to before the Early Cretaceous separation of Africa and South America (c. 110 Ma; Myers 1949, 1966). These clades can therefore be considered the ecosystem incumbents (sensu Wilson 2003). Including a handful of recent anthropogenic transplants, there are few examples of fishes from other continents that have naturally established themselves in cis-Andean South American waters (Hrbek et al. 2007; Pérez et al. 2007). Indeed, the only fish taxa that appear to have successfully joined Amazonian communities during the whole of the Cenozoic are certain groups of marine origin (Lovejoy et al. 1998, 2006; Boeger & Kritsky 2002). Most of these marine-derived clades are represented by relatively few species, with only the potamotrygonid stingrays attaining a moderate level of diversity (>25 spp.).

By the standards of biogeography in a global context, the margins of the Neotropical region are remarkably sharp (Miller 1966; Myers 1966; Lomolino et al. 2006). Only a few Neotropical lineages, including several characins, catfishes and cichlids, have dispersed as far north as central Mexico. There are only about a dozen or so Neotropical freshwater fishes known from the northern pampas of Argentina (Casciotta et al. 1989; Menni & Gomez 1995; López et al. 2002), and the Patagonian fauna is quite distinct. Similarly, very few fish taxa from other regions of the world are present in the Neotropics – a handful of North American derivatives, an African catfish clade (Lundberg et al. 2007), and a few marine-derived taxa occur in southern Mexico and nuclear Middle America (Guatemala and Honduras).

**Geological and palaeogeographic context**

The modern Amazon drainage basin drains about 7 × 10^6 km² of northern South America east of the Andes and between the Guiana and Brazilian Shields. This region is largely covered by moist tropical lowland forests, with some areas of seasonally flooded wetlands and savannas. Amazonia is the greatest interconnected freshwater fluvial system on the planet, discharging about 16% of the world’s flowing freshwater into the Atlantic (Goulding et al. 2003).

The largest geological feature of the Neotropics is the South American Platform, an ancient (Precambrian-Paleozoic >250 Ma) block of continental crust underlying all of Amazonia and adjacent block of continental crust underlying all of Amazonia and adjacent areas (De Almeida et al. 2000; see also Chapters 2 &3). The role of the South American Platform on the diversification of Amazonian fishes cannot be overemphasized. This platform lies very low in the Earth’s mantle, with more than half its total area less than 100 m elevation (P. Petry & J. Albert, unpublished observation). As a result, several areas in South America have been exposed to marine transgressions and regressions over the course of the past c. 110 million years. Documenting the exact extents of these marine transgressions is an active area of research (e.g. Hoorn 1993; Hoorn et al. 1995; Monich 1998; Roddaz et al. 2005; Rebata et al. 2006; Westaway 2006; Hovikoski et al. 2007a, 2007b; see also Chapter 8), yet regardless of the exact positions of palaeocoastlines, it is clear that episodes of marine transgression drastically affected the extent and distribution of habitat available to obligate freshwater species.

Several large-scale (~10^3–10^4 km²) geological features define the contours of the major Amazonian watershed (see Chapter 4). The two principal highland structures of the South American Platform are the Guiana Shield in the northeast and Brazilian Shield in the southeast, which together comprise about half the area of Amazonia. These shields are of Proterozoic (>600 Ma) origins and vastly pre-date the radiations of teleost fishes in the Late Cretaceous and Paleogene (120–23 Ma). The shields attain only modest altitudes (mostly up to c. 1000 m), and have long since lost their easily eroded sediments; they are drained by rivers of low-sediment clear waters (e.g. Xingu, Tocantins, Trombetas, Rio Branco). Along the entire western margin of South America, the Andes constitute another substantial influence on the distribution of Amazonian freshwater fish taxa. Rising to almost 7000 m, the Andes are of Late Cretaceous to Cenozoic age, and therefore much younger than the shields (see Chapter 4) The sediment-rich waters draining the Andes are referred to as white waters (Marañon, Napo, Madeira, Meta, etc.)

Whereas the current watersheds of the Amazon and Orinoco drainage basins are of Neogene age, the phenotypes, behaviours and ecologies of many Amazonian fishes appear to trace their origins to the Paleogene or even Late Cretaceous (Lundberg 1998; see Chapter 17). The earliest palaeo-ichthyofaunas with a modern Amazonian taxonomic composition are from the Paleocene Santa Lucia and Maiz Gordo Formations of eastern Bolivia and northern Argentina, respectively, and the Eocene Lumbrera Formation of northern Argentina (Malabarba et al. 2006). These faunas are from some of the earliest known sediments of a large river basin that drained northwards from the area of the modern Pantanal to the area of the modern western Amazon, and which is presumed to have been an area of lowland freshwater habitats following a major marine regression c. 59–55 Ma (Lundberg et al. 1998; see also Chapter 26). From that time up until the Late Miocene assembly of the modern Amazonian watershed c. 11–12 Ma, this north-flowing river basin must have been the epicentre of freshwater fish diversifications. The onset of the transcontinental Amazon River was preceded by a mega-wetland that had two successive stages
Molecular data and biogeography: considerations and opportunities

The biogeography of Amazonian fishes has previously been addressed with morphology- and taxonomy-based studies. These usually take the form of cladistic analyses of individual clades, where conclusions are drawn about the biogeographical implications of reconstructed phylogenetic trees (e.g., Vari 1988; Albert et al. 2005; Hulsen et al. 2005). Other studies include syntheses of phylogenetic (e.g., Schaefer 1997) or taxonomic data (Hubert & Renno 2006) that use historical biogeographical methods to interpret patterns derived from multiple clades of fishes. To date, these studies have made progress in discerning the effects of palaeogeographic events, but often in areas that are peripheral (albeit related) to Amazonia, such as eastern Brazil and northwestern South America (Vari & Malabarba 1998). Biogeographical studies of Amazon fishes have resulted in patterns that are ‘often contradictory or at best only partially congruent’, as a result of the highly complex geological history of the region, and the great age of many constituent clades (Vari & Malabarba 1998).

Given such difficulties, could molecular data sets and approaches enhance our ability to discern and interpret biogeographical patterns? In this section, we consider the role of molecular data in the biogeography of South American fishes. We summarize approaches and perspectives that molecular data make possible, and briefly mention potential challenges. Our discussion centers on four topics: (i) the use of molecular data for species identification; (ii) phylogeography; (iii) molecular phylogenetics, and (iv) molecular-based age estimation.

Molecular data and species identity

Species are fundamental units of evolution and ecology, and are the basic components of most biogeographical analyses. However, identifying independently evolving species lineages can be a challenging task. In the tropics, taxonomic work on fishes is complicated both by the vast diversity of the fauna and by the difficulties of achieving geographically expansive sampling across terrain that is often remote and inaccessible. The sheer size of Neotropical river systems hampers collecting projects that seek adequately to sample and assess species ranges. Taxonomic progress to date, based mostly on morphology, has been a hard-won multinational effort (Reis et al. 2003).

Molecular data offer welcome assistance for species identification and taxonomic analysis. Genetic approaches are expected to be particularly valuable when morphology-based taxonomy is obscured by phenotypic conservatism, as in the case of potentially cryptic species (e.g., Lovejoy & de Araújo 2000), or by extreme phenotypic variability or plasticity (e.g., Albert et al. 1999; Fernandes et al. 2002; Albert & Crampton 2003). In these situations, molecular data can provide valuable insights that can assist and direct morphological efforts. In several recent examples, molecular data were able to help establish species ranges, distinguish cryptic species, and elucidate confusing cases of intraspecific polymorphism in cichlids (Pérez et al. 2007; Willis et al. 2007), characins (Sivasundar et al. 2001; Dergam et al. 2002; Hubert et al. 2006), potamotrygonid freshwater stingrays (Toffoli et al. 2008) and freshwater needlefishes (Lovejoy & de Araújo 2000).

While molecular data represent a nearly unlimited source of information for species investigation, their use raises a number of methodological and practical issues. One concern is that most studies to date have essentially been based on a single molecular locus: mitochondrial DNA (mtDNA). Mitochondrial DNA is readily amplified from ethanol-preserved tissue, but because the mitochondrial genome is non-recombining and maternally inherited, it may not necessarily share the same genealogical history as loci from the nuclear genome. Indeed, mtDNA lineages have been observed to cross species boundaries in freshwater fishes as a result of hybridization (Bermingham & Avise 1986; Smith 1992; Bernatchez & Wilson 1998), and this phenomenon has been observed in Neotropical fishes (Willis et al. 2007). Until the prevalence of such confounding factors is assessed, reliance on a single genetic locus such as mtDNA is ill advised. This warning is relevant to the proposed ‘barcoding’ of the fauna of tropical regions (e.g., Neigel et al. 2007), an approach to species identification and discovery based on a fragment of a single mitochondrial gene. In general, more robust molecular approaches will require multiple genetic loci, and several authors have proposed methods to combine morphological and molecular data to estimate species boundaries (Wiens & Servedio 2000; Wiens & Penkrot 2002; Sites & Marshall 2003, 2004).

In the foreseeable future, it is not likely that sufficient molecular data will be on hand to allow the accurate identification of all species of Neotropical fishes. This is primarily because multi-locus samples from across the ranges of species, many of which inhabit multiple drainages, would be a requirement. This means that morphology will continue to guide much species discovery and identification, including the ongoing work of evolutionary biologists, ecologists, conservation biologists and others. However, molecular investigations can play a special role in ‘ground-truthing’ morphology-based taxonomy by allowing detailed genetic dissections of model clades. Such studies will help us to answer questions such as: How common are sympatric complexes of cryptic species?
How often are populations of widespread species genetically distinct (perhaps to the point of separate species status)? Since a clear understanding of species identity and distribution is the lynchpin for accurate investigation, such genetic investigations will surely contribute to Amazonian biogeography.

Phylogeography

The field of historical biogeography has traditionally focused on the effects of Earth history events on the distributions of clades and regional biotas (e.g. Morrone & Crisci 1995). By contrast, the field of phylogeography has emerged as a distinctive discipline that focuses on the analysis of intraspecific (population-level) data (Avise et al. 1987; Bermingham & Martin 1998). Here we maintain the distinction between these terms to emphasize the different methods and concepts employed in these two related but distinct areas of investigation. While historical biogeography and phylogeography both investigate the role of Earth history in shaping diversity, phylogeography explicitly incorporates population-level processes, such as gene flow and range expansion.

Because of its emphasis at the species and population level, phylogeography should be useful for reconstructing Amazonian fish biogeography at relatively shallow time horizons, on the order of tens of thousands to perhaps a few million years. We expect the approach to record relatively recent population events and processes associated with climate changes and hydrological evolution of the landscape (e.g. stream capture events). Phylogeography should tell us about recent gene flow along major river systems (Strange & Burr 1997), and very recent alterations of the aquatic landscape, including shifts in drainages and expansions/contractions of aquatic habitats (e.g. Burridge et al. 2007). By emphasizing the boundaries between populations and species, phylogeography should also illuminate the patterns and processes associated with speciation. The pioneering work of Bermingham et al. (e.g. Bermingham & Martin 1998; Reeves & Bermingham 2006) on Central American freshwater fishes demonstrates how the phylogeographic approach can provide important new insights. However, to date, similar analyses of Amazonian fish taxa, especially investigations of geographical phenomena, have been limited. A recent attempt to investigate Pleistocene refugia and piranha population structure (Hubert et al. 2007) is described in more detail in a later section.

Concerns regarding the phylogeographic approach, especially its traditional reliance on single mitochondrial genes, have been voiced (e.g. Degnan 1993; Hare 2001; Choat 2006). Because a single gene tree can be consistent with different scenarios of population history, it can be difficult to falsify phylogeographic hypotheses. Also, a mitochondrial gene may provide a view of population events that may or may not be mirrored in nuclear genes (Hudson & Turelli 2003). Finally, the assumption that mitochondrial DNA evolves in a neutral manner, incorporated into some population genetic models, may not be general (Meiklejohn et al. 2007) – this may call into question conclusions based on neutral models. A practical issue in phylogeographic analysis of Neotropical fishes is appropriate designation of species boundaries. Because relatively few genetic surveys have been undertaken, understanding of species limits is limited and likely to change during the course of an investigation. As in the case of other types of molecular studies, the large geographical ranges of some species, the logistical difficulties associated with making collections, and the extremely high species richness of the fauna as a whole, all conspire to make phylogeographic studies of Amazonian fishes extremely challenging.

Molecular phylogeny

Historical biogeography relies on robust phylogenetic hypotheses. Thus, the development of novel phylogenetic methods and data sets is central to the advance of biogeographical understanding. The advantages of molecular phylogenetics that are of particular importance for the study of Amazonian fish clades include:

1. The potential to resolve phylogenies for species-rich but osteologically conservative groups (e.g. some tetras and cichlids);
2. the ability to test morphological phylogenies with alternative data sources;
3. the ability to infer branch lengths for phylogenetic age estimation (see below);
4. the ability quickly and efficiently to collect vast amounts of data for large numbers of samples.

Perhaps the biggest challenge of applying molecular approaches to Amazonian fishes is obtaining appropriate taxon sampling from such a diverse and widespread fauna. Thus, the vast diversity of Amazonian fishes both hinders and beckons the application of molecular methods.

Molecular estimates of clade ages

A great potential advantage of incorporating molecular sequence data in biogeographical analysis is the estimation of clade ages. The capacity to examine phylogenies within an accurate temporal context is vital for tying cladogenetic patterns to specific palaeogeographic events. Understanding of Amazonian clades, which exhibit a confusing array of temporally overlapping biogeographical patterns, would certainly benefit from time-calibrated phylogenies.

Phylogenetic analysis of molecular data provides two distinct kinds of information: branching order (tree topology) and branch lengths. Branching order provides the history of lineage splitting or speciation events, and can be used in conjunction with palaeogeographic or fossil data to provide estimates of relative ages of clades (Lundberg 1998). For example, a clade with representatives on either side of an impermeable geographical barrier, such as the Eastern Cordillera of the Colombian Andes, may be presumed to be at least as old as the origin of that barrier. Because these types of age estimates depend only on the availability of phylogenies in conjunction with geological data, they are also derivable from morphology-based analyses. Methods for age estimation using branch lengths, however, are currently only available for molecular data sets. Molecular sequences differ from qualitative morphological data in that the constituent units (e.g. nucleotide bases in the case of DNA) are thought to evolve in a manner
that is amenable to statistical analysis (Sanderson 1997; Yoder & Yang 2000).

The logic behind molecular estimates of branch lengths and clade ages is relatively straightforward. Branch lengths may be estimated using a variety of phylogenetic approaches and optimization procedures, and are then converted to ages using appropriate calibrations of rates of molecular evolution (e.g. Near et al. 2005). In practice, however, nearly every step involved in this approach has been criticized and debated (e.g. Sanderson 1997; Yoder & Yang 2000; Pulquieri & Nichols 2006). Much of this discussion involves methodological issues that apply broadly across taxa and geographical areas outside the Neotropics, and are not discussed here. However, studies that attempt to derive molecular-based ages for Neotropical fishes face a number of additional challenges and opportunities, and these are mostly related to calibration issues (Hulsey et al. 2004).

Translating branch lengths (often measured as amount of sequence divergence) to absolute ages (in years) requires an estimate of evolutionary rate. Such rates can be assumed a priori (e.g. 2% divergence per million years is a rate that is sometimes used for mitochondrial genes of fishes), but this approach has its problems (see, e.g., Ho 2007). A preferable technique is to calculate the assumption of the molecular clock, and allow rates of evolution more calibration points. There are sophisticated methods that relax such rates can be assumed to be related to calibration issues (Hulsey et al. 2004).

Using fossils to calibrate molecular estimates of clade ages

The relative or absolute age of fossils is widely used in evolutionary and biogeographical studies to estimate minimum clade ages (e.g. Jablonski et al. 1985; Bemis et al. 1997; Arratia 1999; Albert & Fink 2007). The presence of a fossil with traits diagnostic of a particular taxon is direct evidence for its stratigraphic range (Lundberg et al. 1998; Murray 2001; Malabarba et al. 2006). Of course a lineage may be older than the age of the oldest known fossil, either because of sampling errors, or because the timing of the lineage divergence pre-dated the acquisition of morphological features by which that taxon is recognized. Thus, fossil ages are sometimes used as minimum calibration points for molecular rate estimates (Near et al. 2005).

The utility of the palaeontological record for calibration points depends on the abundance, quality and taxonomic breadth of fossils for the clade of interest. Unfortunately, the record of fossil fishes in the Amazon drainage basin is relatively sparse, especially considering the very high diversity of this region (see Chapter 17). Freshwater fishes are poorly represented as fossils worldwide, compared with near-shore marine fishes or many terrestrial vertebrate groups. This is due to unfavourable conditions for the preservation and recovery of fossils in fluvial systems. Low-energy lacustrine depositional environments, from which most freshwater fossils are known, are rare in the present-day Amazonian hydrological setting. The high current flow and low pH of many tropical rivers combined with high rates of biogenic decomposition also reduce the probability of fossil formation. Further, the discovery of sedimentary outcrops in the Amazon drainage basin is hindered by thickly vegetated landscapes and low topographic relief. To date, fossil fishes from within the watershed of the modern Amazon drainage basin are restricted to the Neogene of western Amazonia. However, fossil Amazonian fish faunas are also known from areas currently outside Amazonia, such as Andean basins to the west and the north.

Fossils of Neotropical freshwater fishes are therefore rare, and most groups are either poorly represented or entirely absent from the palaeontological record. As a result there is a dearth of fossils with phenotypes intermediate between the major groups that dominated Mesozoic and Cenozoic ichthyofaunas. In other words, most fish taxa are fully modern by the time of their first appearance in the stratigraphic record, often being ascribed to modern genera. For example, the fish fauna of the Maastrichtian (c. 71–66 Ma) El Molino Formation of Bolivia is dominated by non-teleost groups (e.g. dipnoans, pycnodonts, polypteriforms, lepisosteids) characteristic of the Cretaceous, and also some archaic teleosts (an extinct siluriform Andinchthys, an osteoglossid) (Gayet et al. 2001, 2003). By contrast, the overlying Paleocene (c. 60–58 Ma) Santa Lucia Formation is dominated by teleosts, especially characiform and siluriform taxa that characterize modern faunas (DeCellos & Horton 2003; Gayet et al. 2003). Such sudden faunal transformations indicate great gaps in the preservational sequence, extremely rapid diversification, or both.

More recent fossil formations provide some useful materials for estimating minimum ages of certain Neotropical freshwater fish clades. Perhaps the best known is the Middle Miocene (c. 12 Ma) La Venta fauna in the Villavieja Formation of what is now the Magdalena valley of Colombia (Lundberg & Chernoff 1992). Fishes of the trans-Andean (west of the eastern Andean cordilleras) La Venta fauna include many living forms that are now known only in the cis-Andean (east of the Andean cordillera) Amazon and Orinoco Basins. Many of these species are indistinguishable from living species (e.g. Arapaima gigas, the pacu Colossoma macropomum), or are closely related to Amazonian species (e.g. the catfish genera Brachyplatystoma and Hoplosternum). The geological isolation of the Magdalena from the Amazon drainage basin began with the rise of the Eastern Cordillera of Colombia about 12 Ma, suggesting a minimum age for the divergence of lineages in cis- and trans-Andean basins (Albert et al. 2006). The Late Miocene Urumaco Formation, in what is now the modern Falcon Basin of northern Venezuela, preserves fossilized remains of fishes from the estuary of the palaeo-Orinoco (Sanchez-Villagra & Aguilera 2006). Fishes of the Urumaco Formation were isolated from the rest of the Orinoco drainage basin by the rise of the Western Merida Andes, between 10 and 8 Ma (Hardman & Lundberg 2006).

Aside from the general circumstances surrounding fossil preservation in Amazonia, taphonomic bias is expected and observed to impact certain clades more than others. Taphonomic biases include differential effects of body size, phenotype, ecology and behaviour.
on the environment of deposition, such as the physical and chemical properties of water and sediments that affect disarticulation, transport, interment and mineralization. Fishes with larger size, robust spines or teeth, or other morphologically diagnostic bony elements, those inhabiting shallow marine or lacustrine habitats or that occur in large schools, are all more likely to be preserved, and when discovered, correctly identified. Among Neotropical freshwater fishes, fossil records are the most abundant and readily identifiable fish fossils. Phractocephaline catfishes are the distinctly ornamented fin spines and skull bones of pimelodid and doradid catfishes, stingray caudal barbs, osteoglossid skull roofing bones, and the characteristic tooth plates of lungfishes, stingrays, and some characins, groups whose remains are common in the Villavieja and Urumaco Formations. Fossils of whole, articulated skeletons are very rare, probably due to the high-energy nature of most Amazonian fluvial environments, as well as predation, bacterial degradation and dislodgement by burrowers. Taphonomic processes therefore result in an under-representation of clades that include many small species, such as tetras (Characinae). Empirical evidence that many important clades are under-represented as fossils comes from the gymnotiform electric fishes. Fragments of articulated mid-body and posterior body sections of fossilized gymnotiform electric fishes are known from the Late Miocene Yecua Formation of Bolivia, providing a minimum date of 12 Ma for the presence of this taxon in Amazonian waters. However, biogeographical and phylogenetic data suggest that the gymnotiform lineage originated in the Lower Cretaceous (>110 Ma), leaving a c. 100 million year gap in the fossil record for this clade.

To date, few molecular biogeographical studies of Amazonian fishes have incorporated fossil-based calibrations, due in large part to the difficulties outlined above. One successful example is Hardman & Lundberg's (2006) study of phractocephaline catfish diversification. These authors used fossil representatives of the extant genus *Steindachneridium* to calibrate a molecular timescale for the rest of the clade, and compared a calibration point derived from the orogeny of the Merida Andes. Aspects of this study highlight the characteristics that will make certain taxa amenable to this type of analysis. Phractocephalines are relatively large fishes (>1 m adult total length), with robust bony elements that preserve well as fossils. They are a relatively depauperate clade, with only a few extant species in four genera, of which two (*Phractocephalus* and *Steindachneridium*) are represented in the fossil record. They are also widespread, with a distribution that includes the Maracaibo and Orinoco drainages – this allows for comparisons between palaeogeography and fossil calibration points.

In summary, fossil-based calibrations of molecular rates hold some promise for Amazonian fishes, as evidenced by Hardman & Lundberg (2006). However, the paucity of the fossil record, as well as the uneven taxonomic distribution of fossils, will necessarily limit the usefulness of the approach.

Using palaeogeographic events to calibrate molecular estimates of clade ages

This calibration method uses dates of palaeogeographic events provided by geological data as minimum ages for the divergence of sister taxa. In the case of freshwater fishes, geographical events that separate river basins and aquatic habitats, such as tectonic uplifts or marine incursions, act as important barriers to gene flow and dispersal. These barriers are expected to lead to allopatric speciation and ultimately may lead to differentiated clades isolated on either side of the barrier. Observed amounts of sequence divergence between these separated taxa, divided by time since separation, provides a rate of molecular evolution that can be used to obtain age estimates for other parts of the clade in question.

It is obvious that obtaining accurate estimates for the geological dates of palaeogeographic events is central to the biogeographical age calibration of molecular divergences. Information from dating methods like fission track analysis and radiometric decay provides the foundation for current understanding of the absolute timing of tectonic events and the ages of sedimentary formations. Biostratigraphy using indicator fossils (e.g. land mammal stages) allows relative dating of geographically disparate faunas when absolute radiometric ages are scarce, as they are in the South American record.

Ideally, a geological vicariance event for age calibration would be:

1. Rapid, separating taxa almost instantly, and also affecting all members of a regional fauna almost simultaneously;
2. Spatially extensive, affecting a broad geographical area and multiple phylogenetically independent taxa;
3. Long-lived, of sufficient geological duration that the lineages are still distributed in allopatry;
4. Impermeable to all members of the fauna (semipermeable barriers are more difficult to perceive after the fact);
5. Accompanied by a volcanism so that the date can be known with great precision by radiometric decay analysis.

The most useful geological dating events are therefore tectonic orogenies with plutonic activity, and from this perspective, the dating of events is expected to be much more reliable in the western than in the eastern Amazonia.

Although of great utility for estimating the age of taxa or whole faunas, dates obtained from palaeogeographic-based molecular calibrations may be subject to several sources of error. Some of these arise from inaccuracies in the methods for obtaining geological dates, and others from uncertainties in the effects of palaeogeological events on individual taxa. For example, the effects of the Isthmus of Panama on geminate marine lineages are more complicated than expected – with gene flow being sundered in different lineages at different times (Bermingham et al. 1997). Biases may also arise from incomplete phylogenetic resolution, incomplete sampling, or an actual history of widespread extinctions, all of which serve to overestimate the true divergence time by reducing information on sister taxon relationships. All of these sources of error in fact hinder biogeographical age calibration across the Eastern Cordillera of Colombia (Albert et al. 2006) and the Western Merida Andes (Hardman & Lundberg 2006). In both of these situations, the trans-Andean fauna is known from palaeontological sources to have been formerly much richer, and these vicariance events were accompanied by regional extinctions. These extinctions eliminate relevant taxa and interfere with accurate phylogenetic reconstruction and branch length estimation.
The timing of many key Neotropical palaeogeographic events remains the subject of active investigation. While the palaeogeography of the Neogene is much better understood than that of earlier periods, the nature and timing of many major Neogene events are still hotly debated (Hoorn et al. 1995; Hoorn 1996; Diaz de Gamero 1996; Campbell et al. 2001, 2006; Rossetti 2001; Campbell 2005; Hoorn & Vonhof 2006; see also Chapter 4). Nevertheless, the Cenozoic history of several major geographical features in Amazonia is sufficiently well established to allow their use in biogeographical age calibrations. Also, important palaeogeographic events that did not directly involve Amazonia, such as the closure of the Central American Isthmus, can also be used as calibration points for widely distributed fish clades.

To date, a relatively small number of events have been used to calibrate molecular rates for Amazonian fish clades (Fig. 25.1). These include:

- the orogeny of the Merida Andes and isolation of the Maracaibo drainage basin (c. 10–8 Ma) (Duque-Caro 1990; Diaz de Gamero 1996; Colletta et al. 1997; Lundberg et al. 1998);
- the orogeny of the Eastern Cordillera of Colombia and isolation of the Magdalena drainage basin (c. 12–11 Ma) (Hoorn et al. 1995; Guerrero 1997);
- the separation of the Orinoco and Amazon Rivers into their modern drainage basin by the rise of the Vaupes Arch (Late Miocene-Pliocene) (Hoorn 1993; Hoorn et al. 1995; Diaz de Gamero 1996; see also Chapter 4);
- the separation of the Paraná and Amazon drainage basins by the rise of the Michicola Arch (c. 42–35 Ma) and/or Chapare Buttress (c. 28–15 Ma) (Butler et al. 1995; DeCelles & Horton 2003).

Based on the criteria discussed above, orogenic dating of lineages from the Maracaibo and Magdalena drainage basins appears to represent the most reliable option for palaeogeographic calibrations. Although these vicariance events have been associated with widespread extinctions – a possible source of error – these events were spatially extensive, long-lived, relatively impermeable to fishes, and associated with well-accepted geological dates. Dating the division of Orinocuan and Amazonian faunas, in contrast, is much less straightforward. Although separation of these two drainages is thought to have occurred about 10 to 8 Ma, after the reconfiguration of an ancient Orinoco river system, they are in fact currently connected via the Casiquiare river, and exchanges of headwater tributaries between these two drainages has probably been relatively

Fig. 25.1 Map of South America showing locations of hydrogeographic barriers discussed in the text. Numbers refer to minimum divergence times in Ma. Base map generated by Paulo Petry using HydroSHEDS high-resolution elevation data.
common throughout the intervening period. Thus, distributions of fish lineages across the Amazonian and Orinocoan divide could be the result of more recent events rather than a singular 10–8 Ma vicariance (e.g. Willis et al. 2007), and calibrations using this event (e.g. Hubert et al. 2007a) should be viewed with caution. Similarly, the hypothesized ‘separation’ of the Amazon and Paraná drainage basins at about 10 Ma (used by Musilová et al. 2008), is perhaps an oversimplification of a complex historical process (Aguilera & De Aguilera 2003). Lundberg et al. (1998) describe multiple stream capture events over the past 40 million years that may have served to facilitate or disrupt gene flow between these two rivers.

Unfortunately, not all fish clades have living representatives in the Magdalena or Maracaibo drainages. The use of palaeo-geographic events to date Amazonian genetic divergences in still in its infancy, yet there are several promising areas for future work. Several palaeogeographic events have been proposed for dating fishes in the drainages of the southeastern Brazilian Shield, including the headwaters of the Upper Paraná, Tietê, São Francisco and Parnaíba drainage basins (Ribeiro 2006; Ingenito & Buckup 2007). Headwater stream capture in the western Guianas (e.g. Caura-Uaricoera, Branco-Essequibo) is currently being used to date cichlids (Lopez-Fernandez et al. 2005a, 2005b, 2006) and loricariid catfishes (Armbruster 2004; Reis et al. 2006). The Late Miocene–Pliocene (c. 9–3 Ma) rise of the Fitzcarrald Arch in southwestern Amazonia (Espurt et al. 2007; see also Chapter 6) is being used to help constrain divergence dates in some electric fishes (J. Albert, unpublished observation). The hydrological separation of the trans-Andean San Juan (Pacific) and Atrato (Caribbean) drainage basins is undated but could in principle aid in understanding the timing of vicariance events within the Choco (Albert et al. 2006).

The formation of rapids and waterfalls in large rivers partially isolates lowland faunas from those of tributary headwaters, and may therefore be used to calibrate more recent (Plio-Pleistocene) speciation events. For example, the rapids at Porto Velho and the Iguacu Falls both mark the lower limits for many fish species endemic to the Upper Madeira (Bolivian) and Upper Paraná drainage basins respectively (Chernoff et al. 2000; Castro et al. 2005; Kullander & Ferreira 2006; Ingenito & Buckup 2007). Indeed, rapids may serve as semipermeable filters for dispersal on all of the large Amazon tributaries. Several studies have attempted to correlate the locations of so-called structural arches (Rodda et al. 2005) with biogeographical distributions, hypothesized to serve as the geophysical underpinnings for rapids in western Amazonia (Da Silva & Patton 1998). However, the accuracy of geologically derived dates for the origins of rapids is clouded by some extent by uncertainties about the mechanisms of their formation, resulting from a combination of regional tectonics and erosion under the influence of eustatic sea-level changes. Further, rapids are not generally a fixed landscape feature as they continually erode their basement sediments and therefore move upstream on geological timescales. Continental islands, such as Trinidad, offer another potential source of more recent palaeogeographic dating events. However, precise dating of lineage-splitting can be hampered by a complex history of iterative isolations and connections caused by long-term fluctuations in sea level. In the case of Trinidad, transoceanic dispersal by means of the Orinoco freshwater plume may be an ongoing phenomenon in some taxa (e.g. Pecilia reticulata) (D. Phillips, personal communication).

Progress to date: molecular signatures of Neogene palaeogeographic events

Because fishes are strictly confined to aquatic habitats, their biogeographical histories are expected to record a detectable signature of river history, including drainage modification and capture events, as well as effects of marine incursions. However, detection of these events depends on the scale of the analysis (taxonomic and temporal) and the response of individual clades (mediated by ecology) to particular palaeogeographic alterations. In the case of Neogene Amazonia, large-scale riverine events of particular importance include:

1. The break-up of the northward flowing palaeo-Orinoco into separate Maracaibo/Magdalena, Orinoco and Amazon drainage basins;
2. The establishment of the east-west transcontinental axis of the modern Amazon River;
3. Connections between the Amazon/Madeira and upper Paraná Rivers;
4. The existence of the vast Pebas mega-wetland system in what is now the area of the western Amazonia;
5. Periodic marine incursions into the lowland basins of various river systems.

These events are of such significant scale (geographically and temporally) that they could reasonably be expected to leave detectable signatures on the biogeographical patterns of Amazonian taxa. At the level of populations, fishes might be expected to record the putative effects of climatic oscillations, recent drainage capture and vicariance events, as well as significant habitat alterations (caused, e.g., by marine intrgressions).

Results from species- and clade-level analyses

To date, molecular analyses have been applied to Amazonian fishes representing a broad range of sizes, life-history strategies and habitat preferences, including: migratory detritivorous Prochilodus Characiformes (Sivasundar et al. 2001; Turner et al. 2004; Moyer et al. 2005), large-bodied riverine predatory phractocephaline catfishes (Pimelodidae) (Hardman & Lundberg 2006), small-bodied, stream-dwelling algivorous catfishes (Hypostomus) (Montoya-Burgos 2003), benthic potamotrygonid stingrays (Lovejoy et al. 1998; Toffoli et al. 2008), as well as cichlids (Willis et al. 2007; Musilová et al. 2008), needlefishes (Lovejoy & de Araújo 2000), piranhas (Hubert et al. 2007a; Freeman et al. 2007; Orti et al. 2008) and killifishes (Hrbek & Larson 1999). All of these studies are based on molecular phylogenetic analyses; some include age estimates based on molecular data (e.g. Hardman & Lundberg 2006) and some incorporate additional types of biogeographic analyses (e.g. Dispersal Vicariance Analysis in Hubert et al. 2007a). The taxonomic scale ranges from comprehensive analyses of
small genera (e.g. three species of *Potamorrhaphis* needlefishes; Lovejoy & de Araújo 2000) to investigations of larger clades (e.g. 60+ species of the killifish family Rivulidae; Hrbek & Larson 1999).

Because of the ecological diversity of the taxa considered above, and the variety of proposed hypotheses and analyses, a simple summary of biogeographical findings is elusive. A general conclusion, with significant ramifications for the testability of geological events, is that biogeographical patterns are often very complex. Even in relatively young groups (<10 Ma) with relatively few (<10) species, observed patterns suggest that dispersal, extinction and hybridization have taken place (e.g. Moyer et al. 2005; Hubert et al. 2007a; Willis et al. 2007) – these processes can obscure and complicate biogeographical interpretation. A related issue is that many historical or current barriers affecting Amazonian fishes are neither permanent nor impassable, as evidenced by multiple reconstructions of dispersal across them. For example, although the fracture of the palaeo-Orinoco resulted in separate Amazon and Orinoco drainages approximately 8 Ma, since that time multiple lineages have moved between these basins using the Casiquiare or the Rupununi ports, or other dispersal pathways (e.g. Lovejoy & de Araújo 2000; Hubert et al. 2007a; Willis et al. 2007). Dispersal is problematic because it can erode the signature of earlier palaeogeographic events, and complicate attempts to use palaeogeographic barriers to calibrate age estimates.

Despite these impediments, there are clear signals of some palaeogeographic events. The Neogene orogeny of the northern Andes is well recorded by endemic Magdalena/Maracaibo lineages in several clades, and as expected for a permanent barrier, observed biogeographical patterns are consistent with a single vicariant event, and no post-barrier dispersal (Lovejoy et al. 1998; Sivasundar et al. 2001; Montoya-Burgos 2003; Turner et al. 2004; Albert et al. 2006; Hardman & Lundberg 2006). Several authors have hypothesized that observed sister clades in the Orinoco and Amazon Rivers can be explained by the conversion of the palaeo-Orinoco into separate Amazon and Orinoco drainage basins. For example, Hubert et al. (2007a), assumed that Orinoco and Amazon sister lineage patterns were established by this palaeogeographic event. Relationships between Paraná and Amazon clades have been attributed to headwater capture or boundary displacement events in several studies (Sivasundar et al. 2001; Montoya-Burgos 2003; Musilová et al. 2008). Montoya-Burgos (2003) and Hubert et al. (2007a) present additional hypotheses associating biogeographical patterns with river changes involving the Madeira, Tocantins, Negro and Ucayali.

Although debate continues about the extent of Neogene marine transgressions in the Amazon region, there seems little doubt that such events strongly affected the distribution of freshwater fishes. A simple prediction is that marine inundated lowland regions would have been uninhabitable by freshwater fishes; hence, we expect such regions to be occupied by lineages that are recently diverged compared to more upland relatives. Hrbek & Larson (1999) documented such a pattern in rivulid killifishes, where taxa on upland shield formations were determined to have diverged earlier than taxa in lowland regions, including the Orinoco llanos, Amazonian savanna, and coastal regions.

Freshwater *Potamorrhaphis* needlefishes show a similar pattern, where widespread haplotype clades in the lower Orinoco and Amazon Rivers are recently diverged relative to lineages in the upper Orinoco (Lovejoy & de Araújo 2000). Hubert et al. (2007a) suggest that their data from *Serrasalmus* and *Pygocentrus* piranhas provide additional evidence of marine effects on lowland taxa. Finally, Lovejoy et al. (1998, 2006) and Lovejoy & Collette (2001) discussed the possibility that marine transgressions also played a role in the evolutionary transitions of several marine fish lineages to freshwater habitats (including stingrays, needlefishes, anchovies and others).

Results from intraspecific (phylogeographic) analyses

A wide variety of intraspecific genetic patterns have been reported for Amazonian fish groups. In studies of large-bodied fish species inhabiting Amazonian floodplains, including *Brachyplatystoma* catfishes (Batista & Alves-Grömes 2006), *Arapaima gigas* (Hrbek et al. 2005) and *Colossoma macropomum* (Santos et al. 2007), little population genetic structure was observed, suggesting extensive gene flow across hundreds of kilometres. In contrast, the needlefish *Potamorrhaphis guianensis*, a smaller-bodied species (30 cm) also found on Amazonian floodplains, shows substantial genetic differences among geographically separated populations (Lovejoy & de Araújo 2000). Cooke & Beheregaray (2007) found extremely high levels of genetic variability in a nuclear intron for cardinal tetras (*Paracheirodon axelrodi*) sampled across the Negro drainage basin. In contrast, investigations of some prochilodontid species from a number of river drainages have reported comparably low amounts of intraspecific genetic variation (Sivasundar et al. 2001; Turner et al. 2004; Moyer et al. 2005). However, none of these studies explicitly tested hypotheses concerning the possible effects of Neogene palaeogeographic events on observed intraspecific patterns.

In a first attempt to link intraspecific patterns to palaeogeographic events, Hubert et al. (2007b) proposed to test the effects of Pleistocene climatic fluctuations by examining population genetic variation within the piranha *Serrasalmus rhombus* in the Madeira River. These authors hypothesized a role for forest refuges in preserving genetic diversity during reductions in the species range, and presented results that they interpreted as evidence for a population expansion during the last 800,000 years. However, the refugia theory in general has been widely criticized on both theoretical and empirical grounds (Endler 1982; Weitzman & Weitzman 1982; Lundberg 1998; Colínvaux & De Oliveira 2000; Colínvaux et al., 2000; see also Chapter 20), and Hubert et al. (2007b) do not present palaeobotanical evidence for the existence of forest habitat, particularly in the regions hypothesized to function as refugia. Indeed, a recent palaeo-vegetation study of Amazonia during the Pleistocene (Anhuf et al. 2007) does not indicate forest refugia in the areas hypothesized by Hubert et al. (2007b) (i.e. Aripuana and Beni regions). Further, *S. rhombus* inhabits rapids and large river channels, and the distribution of this species is not tightly linked to forested habitats. The role of forest refugia in limiting population sizes and dispersal therefore remains unclear. Thus, although the study is a valuable initial attempt, the conclusions of Hubert et al. (2007b) are best viewed with caution until more detailed investigations can be carried out.
Conclusions and future challenges

The use of molecular genetics in biogeography is currently in a pioneering stage of development. Many studies continue to suffer from a reliance on analyses of single clades, the use of single genes (usually mitochondrial), and the assumption that gene trees accurately represent species trees. Studies on the biogeography of Amazonian fishes offer additional challenges, including the extraordinary taxonomic complexity of the fauna and the associated logistical requirements of sample collection and experimental design. Yet even a cursory review of the list of references cited for this chapter demonstrates the advances now being made in all these areas.

The fields of phylogenetics and biogeography face an even greater challenge: to generate and test useful alternative hypotheses. As in the other historical sciences (e.g. geology, astronomy), progress is made by new and more accurate methods of measurement, new sources of information, new hypotheses with increasingly restrictive predictions, and new models for matching observations to predictions. The use of molecular data in historical biogeography and phylogeography offers great potential in all these regards.

Interactions between biology, palaeogeography and geology are increasingly proving to be mutually beneficial. Each of these fields can and should be used to test hypotheses proposed by the others. For biologists, the delineation of testable biogeographical hypotheses is often difficult, in part because Neotropical palaeogeography is such a dynamic field, and also because the relevant geological literature can be difficult to parse. In this regard, it is important for Neotropical biogeographers to set up explicit hypotheses with care, and to consider multiple alternatives.

Acknowledgements

We thank G. Arratia, J. Armbruster, J. Baskin, E. Bermingham, D. Bloom, P. Buckup, P. Chakrabarty, W. Crampton, W. Fink, M. Goulding, S. Kullander, G. Ortí, H. López-Fernandez, J. Lundberg, L. Malabarba, T. Near, J. Neigel, P. Petry, R. Reis, S. Schaefer, G. Smith and K. Winemiller for thoughtful discussions of the ideas presented in this chapter. We thank two anonymous reviewers for their comments on an earlier version of the manuscript. N.R.L. acknowledges support from an NSERC Discovery Grant. J.S.A. acknowledges support from US National Science Foundation DEB 0138633, 0215388 and 0614334.

References


Butler, R.F., Richards, D.R., Sempere, T., Marshall, L.G. (1995) Interactions between biology, palaeogeography and geology are increasingly proving to be mutually beneficial. Each of these fields can and should be used to test hypotheses proposed by the others. For biologists, the delineation of testable biogeographical hypotheses is often difficult, in part because Neotropical palaeogeography is such a dynamic field, and also because the relevant geological literature can be difficult to parse. In this regard, it is important for Neotropical biogeographers to set up explicit hypotheses with care, and to consider multiple alternatives.

Acknowledgements

We thank G. Arratia, J. Armbruster, J. Baskin, E. Bermingham, D. Bloom, P. Buckup, P. Chakrabarty, W. Crampton, W. Fink, M. Goulding, S. Kullander, G. Ortí, H. López-Fernandez, J. Lundberg, L. Malabarba, T. Near, J. Neigel, P. Petry, R. Reis, S. Schaefer, G. Smith and K. Winemiller for thoughtful discussions of the ideas presented in this chapter. We thank two anonymous reviewers for their comments on an earlier version of the manuscript. N.R.L. acknowledges support from an NSERC Discovery Grant. J.S.A. acknowledges support from US National Science Foundation DEB 0138633, 0215388 and 0614334.

References


Molecular signatures in Amazon fish fauna


