Xenopus in Space and Time: Fossils, Node Calibrations, Tip-Dating, and Paleobiogeography

David Cannatella

Department of Integrative Biology and Biodiversity Collections, University of Texas, Austin, Tex., USA

Key Words
Divergence times · DNA phylogenetics · Fossils · Pipoidea · Xenopus

Abstract
Published data from DNA sequences, morphology of 11 extant and 15 extinct frog taxa, and stratigraphic ranges of fossils were integrated to open a window into the deep-time evolution of Xenopus. The ages and morphological characters of fossils were used as independent datasets to calibrate a chronogram. We found that DNA sequences, either alone or in combination with morphological data and fossils, tended to support a close relationship between Xenopus and Hymenochirus, although in some analyses this topology was not significantly better than the Pipa + Hymenochirus topology. Analyses that excluded DNA data found strong support for the Pipa + Hymenochirus tree. The criterion for selecting the maximum age of the calibration prior influenced the age estimates, and our age estimates of early divergences in the tree of frogs are substantially younger than those of published studies. Node-dating and tip-dating calibrations, either alone or in combination, yielded older dates for nodes than did a root calibration alone. Our estimates of divergence times indicate that overwater dispersal, rather than vicariance due to the splitting of Africa and South America, may explain the presence of Xenopus in Africa and its closest fossil relatives in South America.

Xenopus laevis, the African clawed frog, has long been a model organism for a variety of studies at the mechanistic cellular and molecular level. In recent years, its use has expanded as a model for integrating across functional, phylogenetic and comparative levels of biology across multiple species: musculoskeletal anatomy, larval biology, gene duplication, behavior, communication, functional morphology, etc. [Bisbee et al., 1977; Cannatella and de Sá, 1993; Tinsley and Kobel, 1996; Yeh, 2002; Evans et al., 2004; Evans, 2007; Tobias et al., 2011; Irisarri et al., 2011; Bewick et al., 2012]. Much has been learned by examining Xenopus in a comparative phylogenetic context.

A comparative context is particularly important for Xenopus, because the family Pipidae display an unparalleled diversity of bizarre features, many of which are associated with their highly aquatic existence. The feet are extensively webbed, and the metatarsals and phalanges are correspondingly elongate; the head is flattened, and...
the eyes are small. Vocalization consists of a series of clicks or buzzes produced in most cases without any air flow over the uniquely modified larynx. There is no tongue for securing prey; suction feeding is used. The adults retain the larval lateral line organs, in contrast to most frogs in which these disappear at metamorphosis. Overall, the skeleton is highly derived and offers useful characters for phylogenetic analysis.

Extant taxa of the family Pipidae include Pipa (7 species) in South and Central America, and Xenopus (20 species), Silurana (2 species), Hymenochirus (4 species), and Pseudhymenochirus (1 species) in Africa. The sister-taxon of Pipidae is Rhinophrynidae, with a single extant species, Rhinophrynus dorsalis, in Central America, Mexico and Texas. In contrast to pipids, Rhinophrynus is an extreme burrower, with a tiny head, eyes and mouth, and a heavily ossified skull. Its highly modified tongue is specialized for termites. Pipids, Rhinophrynus and the extinct family Palaeobatrachidae† are part of the larger clade Pipioidea.

Although the adults of pipids and Rhinophrynus are extremely divergent from each other, the tadpoles have many distinctive synapomorphies [Cannatella, 1999]. They lack keratinous mouthparts and have paired (rather than single) spiracles (slits in the skin that allow water to pass out of the gill chambers). With a couple of exceptions, they are midwater filter-feeders with flat heads, wide mouths and laterally placed eyes. Historically, many of these features were interpreted as primitive for frogs [Starrett, 1973]. However, as the phylogeny of frogs has become better known, it is now clear that these are highly derived characters [Sokol, 1975; Cannatella, 1999]. Pipoids are an early-diverging but strangely apomorphic clade [Cannatella and Trueb, 1988a, b; Blackburn et al., 2010].

Phylogenetic relationships among genera of pipids have been contentious because molecular and phenotypic (mostly musculoskeletal characters) datasets have supported different trees [e.g. de Sá and Hillis, 1990; Cannatella and Trueb, 1988a; Roelants and Bossuyt, 2005; Trueb and Báez, 2006; reviewed by Bewick et al., 2012]. The discordance can be simplified into 2 alternatives: Hymenochirus (+Pseudhymenochirus) is most closely related to Xenopus (+Silurana); this is generally supported by molecular data. Or, Hymenochirus (+Pseudhymenochirus) is most closely related to Pipa; most analyses of morphological data and fossils support this tree. We call these the Xenopus-Hymenochirus and Pipa-Hymenochirus hypotheses. There is no disagreement about the sister-group relationships of Pseudhymenochirus + Hymenochirus and Xenopus + Silurana, but see Cannatella and Trueb [1988a] and de Sá and Hillis [1990] for previous opinions about Silurana.

The conflict over the relationships among pipids has a paleobiogeographic context. Do the African taxa (Xenopus, Silurana, Hymenochirus, and Pseudhymenochirus) form a clade, or are some more closely related to South American taxa (Pipa)? Do the times of divergence between African and South American taxa predate the separation of the South American and African tectonic plates at about 100 MY? Is the current distributional pattern due to vicariance from the splitting of Gondwana, overwater dispersal occurring after the split, deep divergences predating the rifting of the continents, or extinction? Or, more generally, what is the effect of space and time on pipid evolution?

Fossils are bearers of direct temporal information, and the fossil record of pipoids is excellent compared to other frog groups. Roughly 23 genera are scattered throughout the Jurassic, Cretaceous and Tertiary [Báez, 1996; Sanchiz, 1998; Marjanović and Laurin, 2014; Martin and Sanchiz, 2015]. Previous estimates of the age of Pipidae (as a crown clade) vary from 126 to 165 MY, and its divergence from Rhinophrynidae spans 168–193 MY (table 1).

Fossils may likewise yield insights into biogeographic questions. For example, no living pipoids are known from Europe, but several fossil pipoids have been found. Although only the living taxon Pipa occurs in South America, 5 or 6 extinct genera are known from South America. In addition to the 4 extant African genera, 4 extinct genera are known. No living pipoids are known from the Arabian Plate (which was part of the larger African Plate until the Oligocene), but at least 3 fossil genera are found there.

We will use 3 sources of evidence to examine relationships of pipids: DNA, morphology of fossils and extant forms, and fossil age. This information is derived from 2 datasets: DNA sequences from Hedtke et al. [2013] and osteological data from extant and extinct taxa from Báez et al. [2007]. The fossils provide temporal information at 2 levels: as indirect information to estimate divergence times (node-dating), and as direct information about age of the fossils (tip-dating), which can be used to estimate divergence times and tree topology [Ronquist et al., 2012a; Wood et al., 2013].

We analyze 6 extant pipoid species (and several outgroups) with DNA sequences and phenotypic characters, and 15 fossil taxa (out of ~23 fossil pipoid genera), including geological age and comparable phenotypic data. We assess the conflicting topologies produced by DNA and phenotypic data as well as the role of the temporal
information encoded in fossils in resolving these conflicts. We also examine the effect of calibration priors on node ages. Finally, we analyze various paleobiogeographic scenarios to explain the current distribution of pipoids.

**Materials and Methods**

**DNA Sequences**

Rather than aggregate sequences from a variety of sources, which would result in a matrix with many missing cells, we chose the dataset from Hedtke et al. [2013], consisting of 200,090 sites, and extracted a subset of the data in which all genes were present in all taxa. The resulting dataset had 8,473 sites from 23 genes of which 1,058 sites (12.5%) were parsimony-informative. This subset yields the same topology, with similar levels of support, as the entire dataset (see below).

The taxon sample included species spanning the earliest diverging lineages of frogs: *Ascaphus truei*, *Leioptelma hochstetteri*, *Bombina variegata*, *Discoglossus cf. pictus*, *Rhinophrynus dorsalis*, *Xenopus laevis*, *Silurana tropicalis* (= *Xenopus tropicalis*), *Hymenochirus curtipes*, *Pipa pipa*, and *Scaphiopus hurteri*.

Specimen data for the DNA vouchers (Herpetology, Texas Natural History Collections, The University of Texas at Austin) are provided in Hedtke et al. [2013]. Data from *Pipa carvalhoi* (morphology) and *P. pipa* (DNA) were concatenated to form a chimaeric taxon. Similarly, data from *Ascaphus montanus* (DNA) and *A. truei* (morphology) were combined.

**Phenotypic (Morphological) Characters from Extant Taxa**

This data matrix of 58 phenotypic characters (osteology) scored for 11 extant taxa. The matrix is taken unchanged from Báez et al. [2007], which was based on Báez and Harrison [2005] and Trueb and Báez [2006]. It was chosen because it minimizes inter-observer bias, having been assembled largely in collaboration by Ana María Báez and Linda Trueb. Detailed character descriptions, illustrations and lists of skeletal material are given in Cannatella and Trueb [1988a, b], Báez and Trueb [1997], Báez and Pugener [2003], and Báez and Harrison [2005].

Phylogenetic studies should call attention to issues of species identifications. Some data were collected before certain species were discovered to be complexes of several species. Thus, it is possible that specimens of *A. truei*, *D. pictus* and *Hymenochirus boettgeri* may actually represent closely related species; given that divergences among these species are recent, this is unlikely to cause problems. For all 3 taxa, comparative material of more than 1 species was examined by Báez and Trueb [1997], and the character states scored were consistent across the species.

**Temporal Data: Ages of Fossils and Calibration Priors**

The ages of the fossils, which were used for node- and tip-dating, were initially taken from Marjanović and Laurin [2014], the Lisanfos database [Martin and Sanchiz, 2015], and the Fossilworks database (fossilworks.org). Due to discrepancies or lack of precision among the three, the dates of occurrence including the age of the type-locality horizon (most are known only from the type-locality) were verified from the primary literature (table 2). We used as precise an estimate of the age as possible; most of our assigned dates have a narrower range than those of Bewick et al. [2012, table 2] as well as the original descriptions. Typically, we used the minimum (younger) date of the stratigraphic range as the offset for the prior distribution, in contrast to using the mean of the minimum and maximum ages as is done in some cases.

We used an offset exponential prior for calibration of 3 internal nodes in node-dating analyses (MrBayes: calibrate Node_1 = offsetexponential [x, y]) and to parameterize the tree-age prior (treeagepr) for the root node (Anura) for all relaxed clock (IGR) analyses. The exponential distribution has 1 parameter, the scale param-

---

### Table 1. Comparison of node ages of selected taxa, rounded (see also table 4)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Anura</th>
<th>Pipoidea</th>
<th>Pipidae</th>
<th>Xenopus + Silurana</th>
</tr>
</thead>
<tbody>
<tr>
<td>This paper, DNA data, root calibration only</td>
<td>183 (166–235)</td>
<td>140 (100–189)</td>
<td>101 (62–141)</td>
<td>38 (16–65)</td>
</tr>
<tr>
<td>This paper, DNA data, node calibrations</td>
<td>194 (167–229)</td>
<td>151 (148–160)</td>
<td>107 (90–129)</td>
<td>37 (27–55)</td>
</tr>
<tr>
<td>This paper, DNA-Morph-Fossil data, node- and tip-calibrations</td>
<td>190 (167–223)</td>
<td>152 (148–161)</td>
<td>105 (92–121)</td>
<td>36 (27–51)</td>
</tr>
<tr>
<td>Evans et al., 2004</td>
<td>na</td>
<td>na</td>
<td>130 (115–146, using bootstrap resampling)</td>
<td>64 (50–81)</td>
</tr>
<tr>
<td>San Mauro et al., 2005</td>
<td>263 (223–304)</td>
<td>168 (128–206)</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>Irissari et al., 2012</td>
<td>230 (200–262)</td>
<td>168 (128–206)</td>
<td>126 (88–162)</td>
<td>52 (22–88)</td>
</tr>
<tr>
<td>Bewick et al., 2012, based on their analysis 3, except as noted</td>
<td>na</td>
<td>145 (116–176, analysis 1)</td>
<td>135 (112–156)</td>
<td>65 (57–74)</td>
</tr>
<tr>
<td>Zhang et al., 2013, MultiDivTime</td>
<td>231 (206–262)</td>
<td>180 (159–207)</td>
<td>165 (144–191)</td>
<td>106 (88–128)</td>
</tr>
<tr>
<td>Zhang et al., 2013, BEAST</td>
<td>257 (216–296)</td>
<td>178 (151–219)</td>
<td>150 (113–198)</td>
<td>83 (48–127)</td>
</tr>
</tbody>
</table>

The median/mean age (in MY) is followed by 95% HPD credible intervals, except as noted. na = Data not available.
eter, which is the inverse of the mean. We chose the exponential
distribution over the commonly used lognormal distribution to
describe the probability that the divergence occurring during a cer-
tain time interval is proportional to the length of that time interval.
 That is, the probability of sampling older values of the node age
decreases at a constant average rate.

In MrBayes, the offset parameter determines the location of the
distribution on the x-axis (time); this is typically parameterized
using the minimum age of the fossil. In addition, the mean for the
prior density of the exponential distribution must be specified (in
MrBayes the ‘mean’ parameter is the sum of the offset and the
mean of the exponential distribution). The mean and soft maxi-
mum were determined as follows. The soft maximum was set to be

---

Table 2. Fossil calibrations for clock analyses

<table>
<thead>
<tr>
<th>Species</th>
<th>Calibrated node</th>
<th>Range min</th>
<th>Range max</th>
<th>97.5% quantile</th>
<th>Mean</th>
<th>Mean (MrBayes)</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Eodiscoglossus oxoniensis</em>†</td>
<td>Anura</td>
<td>166.1</td>
<td>166.5</td>
<td>252.0</td>
<td>23.3</td>
<td>189.4</td>
<td>182.2</td>
</tr>
<tr>
<td><em>Rhadinostes parvus</em></td>
<td>Pipoidea</td>
<td>148.1</td>
<td>150.3</td>
<td>166.1</td>
<td>4.9</td>
<td>153.0</td>
<td>151.5</td>
</tr>
<tr>
<td><em>Pachycentrata taqueti</em>†</td>
<td>MRCA of <em>Xenopus</em> + <em>Hymenochirus</em> or of <em>Pipa</em> + <em>Hymenochirus</em></td>
<td>83.6</td>
<td>88.0</td>
<td>148.1</td>
<td>17.4</td>
<td>101.0</td>
<td>95.7</td>
</tr>
<tr>
<td><em>Xenopus arabiensis</em>†</td>
<td>MRCA of <em>Xenopus</em> + <em>Silurana</em></td>
<td>26.5</td>
<td>30.9</td>
<td>83.6</td>
<td>15.5</td>
<td>42.0</td>
<td>37.2</td>
</tr>
</tbody>
</table>

The Range min column was used as the x parameter for the ‘calibrate = offsetexponential (x, y)’ and ‘prset treeagepr = offsetexponential (x, y)’ commands in MrBayes. The Mean (MrBayes) column was used as the y parameter; it is the sum of the Mean and the Range min. The mean, median and 97.5% quantile refer to the density of the exponential prior. Numbers are in millions of years. MRCA = Most recent common ancestor.

---

Table 2. Fossil calibrations for clock analyses

A Fossil calibrations for node-dating analyses

<table>
<thead>
<tr>
<th>Species</th>
<th>Calibrated node</th>
<th>Range min</th>
<th>Range max</th>
<th>97.5% quantile</th>
<th>Mean</th>
<th>Mean (MrBayes)</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Eodiscoglossus oxoniensis</em>†</td>
<td>Anura</td>
<td>166.1</td>
<td>166.5</td>
<td>483.0</td>
<td>85.9</td>
<td>252.0</td>
<td>225.6</td>
</tr>
<tr>
<td><em>Rhadinostes parvus</em></td>
<td>Pipoidea</td>
<td>148.1</td>
<td>150.3</td>
<td>214.5</td>
<td>18.0</td>
<td>166.1</td>
<td>160.6</td>
</tr>
<tr>
<td><em>Pachycentrata taqueti</em>†</td>
<td>MRCA of <em>Xenopus</em> + <em>Hymenochirus</em> or of <em>Pipa</em> + <em>Hymenochirus</em></td>
<td>83.6</td>
<td>88.0</td>
<td>321.5</td>
<td>64.5</td>
<td>148.1</td>
<td></td>
</tr>
<tr>
<td><em>Xenopus arabiensis</em>†</td>
<td>MRCA of <em>Xenopus</em> + <em>Silurana</em></td>
<td>26.5</td>
<td>30.9</td>
<td>57.1</td>
<td>83.6</td>
<td>66.1</td>
<td></td>
</tr>
</tbody>
</table>

Similar parameterization as in A using the method of Ronquist et al. [2012a]. MRCA = Most recent common ancestor.

---

Table 2. Fossil calibrations for tip-dating analyses

<table>
<thead>
<tr>
<th>Species</th>
<th>Taxon</th>
<th>Range min</th>
<th>Range max</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Chelomophrynus bayi</em>†</td>
<td>Rhinophrynidae</td>
<td>42.0</td>
<td>46.2</td>
</tr>
<tr>
<td><em>Neusibatrachus wilferi</em>†</td>
<td>Pipimorpha</td>
<td>118.0</td>
<td>118.0</td>
</tr>
<tr>
<td><em>Cordicephalus gracilis</em>†</td>
<td>Pipimorpha</td>
<td>95.0</td>
<td>99.0</td>
</tr>
<tr>
<td><em>Thoraciliacus rostriceps</em>†</td>
<td>Pipimorpha</td>
<td>118.0</td>
<td>118.0</td>
</tr>
<tr>
<td><em>Avitabatrachus ulianar</em>†</td>
<td>Pipimorpha</td>
<td>76.0</td>
<td>78.0</td>
</tr>
<tr>
<td><em>Palaeobatrachus sp.</em>†</td>
<td>Palaeobatrachidae</td>
<td>24.5</td>
<td>26.8</td>
</tr>
<tr>
<td><em>Saltenia ibanezi</em>†</td>
<td>Pipidae</td>
<td>74.0</td>
<td>74.0</td>
</tr>
<tr>
<td><em>Vulcanobatrachus mandelai</em>†</td>
<td>Pipidae</td>
<td>60.0</td>
<td>70.0</td>
</tr>
<tr>
<td><em>Exoenepoidea reuningi</em>†</td>
<td>Pipidae</td>
<td>55.5</td>
<td>58.2</td>
</tr>
<tr>
<td><em>Xenopus’ romeri</em>†</td>
<td>Pipidae</td>
<td>54.2</td>
<td>54.2</td>
</tr>
<tr>
<td><em>Llankibatrachus truberae</em>†</td>
<td>Pipidae</td>
<td>51.9</td>
<td>51.9</td>
</tr>
<tr>
<td><em>Shelania pascuali</em>†</td>
<td>Pipidae</td>
<td>46.0</td>
<td>47.0</td>
</tr>
<tr>
<td><em>Singidella latecostata</em>†</td>
<td>Pipidae</td>
<td>41.2</td>
<td>56.0</td>
</tr>
<tr>
<td><em>Shelania’ laurentii</em>†</td>
<td>Pipidae</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The minimum and maximum ages were used to calibrate a uniform distribution.
the minimum age of the fossil used to calibrate the next more inclusive relevant node; it is defined as the 97.5% quantile of the exponential distribution (table 2). We used Microsoft Excel functions to determine the soft maximum value (the upper limit of the 97.5% cumulative distribution), the mean, and the median (50% quantile). This method is motivated by that of Ronquist et al. [2012a], who used the mean of the exponential distribution of the next more inclusive node as the soft maximum rather than the 97.5% quantile. We chose the 97.5% quantile so that the maximum would not overlap the minimum age of the next most inclusive calibrated node; the details of individual calibrations are provided in table 2.

The calibration of the root node Anura was used for all clock analyses, including node-dating and tip-dating. To calibrate Anura, we used Eodoscglossus oxoniensis† from Kirtlington Quarry, Forest Marble Formation, Late Bathonian, middle Jurassic. It is universally regarded as being within crown group Anura, usually as part of the more restricted crown group Alytoida (formerly Discoglossoidea) (stated as 'Discoglossidae' by Benton et al. [2015]). In the description of E. oxoniensis†, Evans et al. [1990] described the atlas and inferred that the atlas (and therefore the entire vertebral column) was opisthocoelous owing to a concavity (cotyle) on the posterior end of the centrum. They also reported that at least 1 atlas of 3 is perforated, presumably for the notochord. A different interpretation of this morphology is possible. Frogs with notochordal ('amphicoelous') vertebrae such as the extant Ascaophys and Leiopelmia, and the Jurassic fossil Prosalirus bitit† [Shubin and Jenkins, 1995] have concavities on both ends, with a small perforation for the remnant of the notochord. Thus, contrary to Evans et al. [1990] and Benton et al. [2015], it may be that E. oxoniensis† has notochordal and not opisthocoelous vertebrae, and thus lacks a synapomorphy of crown Alytoida. Thus, we consider it minimally as part of crown group Anura rather than Alytoida.

Benton et al. [2015] used 166.1–166.5 MY for the age of the Forest Marble Formation and set 165.3 MY as the minimum, taking into account the 95% confidence interval of ±1.2 MY. There seems to be a slight mathematical error, because 166.1 − 1.2 = 164.9, not 165.3. We use 166.1–166.5 MY for the age range for E. oxoniensis† (table 2).

For the soft maximum of the root calibration, the 97.5% quantile value was set to be the minimum age of the salientian (stem-Anura) fossil Triadobatrachus massinoti† at 252.0 MY (table 2). It is not a frog, but is more closely related to Anura than to Caudata (salamanders) or Gymnophiona (caecilians). Although several Jurassic stem-salamanders and a stem-caecilian are known, Triadobatrachus† is the oldest fossil of crown Amphibia (Lissamphibia of paleontologists), and no fossils within crown Amphibia are known from the Permian (>252 MY). Therefore, we used the Permian-Triassic boundary at 252.0 MY as the soft maximum for Triadobatrachus†, although one could equally justify the top of the Induan at 251.2 MY [Benton et al., 2015; Cohen et al., 2015].

As a comparison, we explored the effect of following Ronquist et al. [2012a] in using the mean of the exponential distribution rather than the 97.5% quantile (table 2) as the soft maximum. The prior density was sampled under this scheme by running the analysis with no data (‘mcmc data = no’). We felt that the resulting CIs were too liberal; for example, the density of the prior distribution of the age of Anura between 97.5% and 50% (the median) lies between 225.6 and 483.0 MY. Given that this median is >100 MY older than the oldest anuran fossil, and that it extends 35 MY into the Permian, a period during which no crown amphibian fossils are known, we felt it was unrealistic.

We calibrated internal nodes for node-dating analyses using 3 additional fossils: Rhadinostus parvus†, Pachycentrata taqueti† and Xenopus arabiensis†. These fossils are not part of the Morph/Fossil data matrix, and are therefore independent. Because E. oxoniensis was used to calibrate the root, we will not refer to it further, and we discuss internal node calibration only in terms of the following 3 fossils.

R. parvus† was named by Henrici [1998a, b] from the Brushy Basin Member of the Morrison Formation of Utah (Late Jurassic, Kimmernidan). As the oldest fossil in the crown group Pipoidea, it is particularly crucial for calibrating the node Pipoidea, the focus of this study. We accept its uncontroversial phylogenetic position as a stem-rhynophyrid, closer to Rhinophrymus than to Pipidae [Henrici, 1998a; Báez, 2013]. However, few of the characters scored for Rhadinostus by Báez [2013] match those in our matrix, so it was not included in the tip-dating analysis.

The Brushy Basin Member was dated at 148.1–150.3 MY [Kowallis et al., 1998], which would indicate that it is from the Tithonian. Radiometric data place the Morrison Formation from 156.3 MY at its base [Trujillo et al., 2006] to 146.8 MY at the top [Bilbey, 1998], which overlaps the Kimmernidan. We use 148.1–150.3 MY for the age range as specified by Kowallis et al. [1998]; the parameters of the prior are given in table 2.

The second calibration fossil, P. taqueti†, was described (as Pachybatrachus) by Báez and Rage [1998] from the In Beceten Formation of Niger. They gave the age as late Coniacian-Santonian (Upper Cretaceous). Moody and Sutcliffe [1991] and Mateer et al. [1992] indicated that the In Beceten Formation extends from the uppermost Coniacian (86.3–89.8 MY) through almost all the Santonian (83.6–86.3 MY). We set the range as 83.6–88.0 MY, truncating the lower part of the Coniacian. Marjanović and Laurin [2014, fig. 3] reported its stratigraphic range as restricted to the Coniacian.

All our analyses of morphological data (including fossils) as well as previous analyses [Báez and Rage, 1998] placed Pachycentrata as the oldest fossil within the crown clad delimited by Xenopus and Hymenochirus (in the DNA tree) or as the oldest fossil within the crown clad delimited by Pipa + Hymenochirus (in the morphology tree). We use it to calibrate these clades as appropriate, depending on the favored topology.

The third calibration fossil, X. arabiensis†, was named by Henrici and Báez [2001], who placed the fossils in the Late Oligocene. Although they did not perform a phylogenetic analysis of this fossil, they listed synapomorphies that placed it closer to Xenopus (specifically to X. muelleri) than to Silurana. Therefore, X. arabiensis† was used to calibrate the most recent common ancestor of Xenopus + Silurana. We used 26.5–30.9 MY, roughly Mid- to Late-Oligocene following Baker et al. [1996], who determined the age of the Yemen Volcanic Group, in which the fossils were collected, with 40Ar/39Ar dating; see table 2 for parameters of the prior distribution.

The following fossils were used for tip-dating analyses (table 2). The maximum and minimum ages were used to parameterize a uniform prior distribution, or a fixed age in the case of a single age.

Chelomophrymus bayi† was described by Henrici [1991] from the Early Uintan (North American Land Mammal Age, middle Eocene) of Wyoming. More specific information is not available,
and we used 42.0–46.2 MY, the accepted range of the Uinta (fossilworks.org).

Neusibatrachus wilferti† Seiffert 1972, from Santa Maria de Meyá, Spain, was reported as late Berriasian to early Valanginian by Báez and Rage [1998] and Báez and Sanchiz [2007]. However, Soriano and Delclòs [2006] stated that the age is Barremian (Lower Cretaceous), and Báez [2013] stated early Barremian. We use the Barremian (125.0–129.4 MY) as the age range.

Cordicephalus gracilis† Nevo 1968 is from the Hatra Formation of the Lower Cretaceous, Makkhtesh Ramon, Israel [Nevo, 1968]. The Ramon Basalt, which sandwiches the layer of frog fossils, was dated at 118 MY (Aptian) by Gvirtzman et al. [1996]. We used 118 MY for Cordicephalus. Thoraciliacus rostriceps† Nevo 1968 has the same provenance as Cordicephalus†, and we calibrated this fossil at 118 MY.

Palaeobatrachus sp.† is a composite operational taxonomic unit based on several specimens of 3 species examined by Báez and Trueb [1997], all from Bechlejovice (Oligocene), near Bratislava, Czech Republic. The age of the fossil site is considered to be Late Oligocene at 24.5–26.8 MY [Bellon et al., 1998; Roček, 2003], which we used.

Avitabatrachus ulianae† was named from the Candeleros Formation, Argentina by Báez et al. [2001] who gave the age as ‘?Cenomanian-Turonian’. Marjanović and Laurin [2014, fig. 3] placed the range of Avitabatrachus at 100.5–113.0 MY, during the Albian. We followed Garrido [2010, fig. 3], who placed the Candeleros formation in the lower Cenomanian (Upper Cretaceous), at roughly 95–99 MY.

Eoxenopoides reunigi† Haughton 1931 is from a crater lake formed in Arnot Pipe (near Banke), one of the many volcanic pipes in South Africa.特斯 [1977] suggested a late Eocene or Oligocene age, but this has been discounted. A 60–70 MY date (Late Cretaceous) of the crater lake mudstone, based on 238U/206Pb dates from similar, nearby volcanic pipes, was reported by Scholtz [1985], who noted that this age is consistent with his palynological evidence. We used 60–70 MY. Marjanović and Laurin [2014, fig. 3] placed the range of E. reunigi at 72.1–83.6 MY (Campanian, Upper Cretaceous).

Singidella latecostata† was named by Báez and Harrison [2005] from an Eocene crater lake in Tanzania; they placed the age at 46–47 MY (Lutetian, middle Eocene), which we follow. Marjanović and Laurin [2014] placed Singidella in the Coniacian at 86.3–89.8 MY.

Vulcanobatrachus mandelai†, like Eoxenopoides, is from lacustrine sediments in a South African volcanic pipe. Trueb et al. [2005], who named the species, stated the age of the Stompoo Pipe as mid-Senonian (Upper Cretaceous). Although they ascribed the age estimate to Scholtz [1985], in fact his statements were based on the Arnot Pipe, which is about 300 km west of Stompoo. Following Smith et al. [2002], Roček and Van Dijk [2006] gave the age of the Stompoo volcanic pipe as 74 MY (Late Cretaceous), which is slightly older than the Arnot Pipe. Marjanović and Laurin [2014, fig. 3] placed Vulcanobatrachus in the Santonian (83.6–86.3 MY). We followed Roček and Van Dijk [2006] and used 74 MY.

Saltienia ibanezii† Reig 1959 is from the Las Curtiembres Formation (Upper Cretaceous) of Argentina. Based on pollen studies, Narváez and Sabino [2008] stated that the age of the formation is roughly 77 MY. Scanferla et al. [2011], citing Valencio et al. [1976], stated that the Las Conchas Basalt is 76–78 MY (Campanian) according to K/Ar dating. We used 76–78 MY.

Llkibatrachus truebault† was named by Báez and Pugener [2003] from Pampa de Jones, Argentina, middle Eocene-early Oligocene. Wilf et al. [2010] provided the age of Pampa de Jones as 54.2 MY, which we followed.

Shelania pascuali† Casamiquela 1960 was described from Laguna del Huncu (Eocene), Argentina. Báez and Pugener [2003] gave the age as 43–51 MY. Wilf et al. [2010] placed it precisely as 51.9 MY, which we followed.

‘Xenopus’ romeri† Estes 1975 is from the São José of Itaborai Formation of Brazil. Marshall et al. [1997] re-evaluated the age of the Itaborai fauna and dated it at 55.5–58.2 MY, late Paleocene. We set the dates at 55.5–58.2 MY.

‘Shelania laurenti† was described by Báez and Pugener [1998] from the Eocene Vaca Mahuida Formation, Argentina. They discussed evidence supporting a late Paleocene-middle Eocene age for the frog-bearing strata. The Fossilworks database reports that the Vaca Mahuida Formation is Late/Upper Eocene at 37.2–33.9 MY, based on Uliana and Camacho [1975]. We followed Báez and Pugener’s rationale and chose 41.2–56.0 MY to represent the lower to middle Eocene (Lupesrian and Ypresian stages).

Taxonomy

The clade of extant taxa Pseudhymenochirus + Hymenochirus has been named Hymenochiriini, at the rank of tribe (hymenochirines, informally). Because all analyses join Pseudhymenochirus and Hymenochirus as sister-taxa, we occasionally use the name Hymenochirus to include both genera, as in the ‘Xenopus-Hymenochirus’ hypothesis. Pipimorphia is the name for the branch (stem) that includes fossil taxa that are more closely related to living pipids than to Rhinophrynus [Ford and Cannatella, 1993].

We place ‘Shelania laurenti† in quotes because no analyses have recovered it as the sister-taxon of Shelania pascuali†, the type-species of Shelania†. Similarly, no recent analyses have placed ‘Xenopus’ romeri† as the sister-taxon of Xenopus sensu stricto or Xenopus + Silurana, so we use quotes around Xenopus. It is clear that ‘Xenopus’ romeri† requires a new generic name.

Phylogenetic Analyses and Hypothesis Testing

Phylogenetic analysis was performed using the Unix version of PAUP* 4b10-x86 [Swofford, 2002] for parsimony, the parallel mpi version of MrBayes 3.2.5 [Altekar et al., 2004; Ronquist et al., 2012b] with the BEAGLE application programming interface [Ayres et al., 2012], and BEAST 1.8.2 [Drummond et al., 2012] for Bayesian analysis. Summary analyses and visualization were done with Tree Annotator 1.8.2, Tracer 1.6, and FigTree 1.4.2 [Rambaut and Drummond, 2015; Rambaut et al., 2015]. The credibility intervals from the MrBayes analyses were given as the region of 95% highest posterior density (HPD).

The Morph and Morph-Fossil datasets were analyzed using PAUP* and MrBayes. Characters were analyzed as unordered, following the original publications; non-applicable character states were scored as missing. In the MrBayes analysis, the symmetric dirichlet prior was set to ‘fixed(infinity)’, both variable and invariant sites were used (coding = all), and the rate heterogeneity parameter was estimated (rates = gamma).

The DNA data were analyzed using parsimony and Bayesian analysis; branch support was measured by bootstrap proportions (BP; 1,000 pseudoreplicates) and posterior probabilities (PP). A BP ≥90 and PP ≥0.95 is considered strong support (decimals truncated), a BP of 70–89 and PP of 90–94 as moderate support, and val-
uses below these limits as weak support. To determine bootstrap support for alternative topologies, a PAUP\* filter was applied to the bootstrap sample using the appropriate backbone constraint. Most of the deep nodes of the phylogeny (Bombinatora, Alytoida, Pipanura, and Pipioidea) are very strongly supported in published analyses, and these were constrained to be monophyletic.

Preliminary analyses of DNA were performed using various partitioning schemes, which were evaluated using Bayes factors (BF) calculated using stepping stone sampling (see below). All analyses gave the same topology, and all nodes had PPs of 1.0 except for one, for which the value varied among analyses by <3%. We conducted DNA analyses using a 2-partition model, one for 1st and 2nd codon positions and the second for 3rd positions. Partitioning by gene or by all 3 positions yielded worse fit (2ln[BF] > 20). Partitioning by both gene and codon position yielded a very large number of partitions, many of which had very few informative sites, and thus many additional parameters to estimate (>60). The substitution matrix parameters were estimated using the GTR model, and among-site rate variation was estimated using the invariant sites parameter and alpha shape parameter of the gamma distribution.

DNA data were analyzed using standard and relaxed-clock analyses in MrBayes. The non-clock analyses were typically run with 4 Markov chains, 4 replicates, at least 10 million generations, with a sampling rate of 1/500 or 1/1,000 generations and 20–30% burn-in (determined a posteriori). These were summarized as a maximum a posteriori probability tree.

Relaxed-clock analysis was performed in MrBayes with the Independent Gamma Rates model (IGR, also called the white noise model) [Lepage et al., 2007; Ronquist et al., 2012a, b]. Under this model, substitution rates among branches are independent (i.e. not autocorrelated) and drawn from a gamma distribution. This model was chosen following Ronquist et al. [2012b], who found that the 2 autocorrelated models available in MrBayes, TK02 and CPP, tended to estimate longer basal branches in the combined fossil-DNA analyses and thus longer ghost lineages [Norell, 1992], i.e. the inferred branch between the origin of a lineage and the earliest fossil lying on that lineage, than did IGR. This is probably because the IGR allows rates among branches to vary more substantially than do the 2 autocorrelated models, reducing the duration of ghost lineages [Ronquist et al., 2012a].

For all IGR analyses (MrBayes), the root node Anura was calibrated with the tree-age prior as described in the section on fossil calibrations. All clock analyses were done using this tree-age prior. We used 3 fossils to calibrate internal nodes in some analyses as described above (table 2). The 2 groups of analyses are described as having or lacking fossil (internal) node calibrations, even though the fossil E. oxoniensis† was used to set the tree-age prior at the root node.

The clock rate prior (for the base substitution rate), measured in expected number of substitutions per site per million years, was set as \( \text{clockrate} = \log_{10}(\text{mean} \times 2.0, 0.0) \), with parameters mean and standard deviation on a natural log scale, as described in Ronquist et al. [2012a].

Preliminary analyses of the DNA data using BEAST with the lognormal relaxed clock yielded mean node ages and credible intervals that were extremely similar to those recovered from MrBayes. Because MrBayes facilitates combined data analyses with multiple Markov chains and replicate runs, as well as a simpler procedure for incorporating discrete data, we used it for all analyses.

Two non-exclusive strategies for estimating chronograms were used: node-dating and tip-dating [sensu Ronquist et al., 2012a]. In node-dating, the age of a fossil (usually the minimum age) is used to estimate the age of the ancestor node of the crown group to which the fossil belongs; the characters of the fossil are not used to estimate the tree. In tip-dating, the fossil character data are used to estimate the tree, and the age of the fossil is also used to estimate the branch age. Because both character and temporal information are used to infer a chronogram, tip-dating has been called a ‘total-evidence’ method [Ronquist et al., 2012a]. In our study, the 3 fossils that were used to calibrate nodes were not used to date the tips of the tree; we were thus able to combine node-dating and tip-dating, using different sets of fossils, in a novel approach.

Two methods were used to test hypotheses. For parsimony analyses the AU test (10,000 bootstrap replicates) [Shimodaira, 2002] and winning sites (WS) test as implemented in PAUP* were used to test the hypothesis of no difference between the likelihood (AU) or parsimony (WS) scores of the alternative DNA topologies. Full likelihood evaluation of the trees was done under the GTR+G+I model, with all parameters estimated.

Bayes factors were also used to compare rival hypotheses. A Bayes factor is the ratio of the marginal likelihoods of 2 models, which is used to measure support for one hypothesis over the other. We used 2ln(BF\text{10}) as the test statistic [Kass and Raftery, 1995; Brown and Lemmon, 2007], which is calculated as the difference of the natural logs of the 2 marginal likelihoods of model (hypothesis) 1 and model 0, times 2. A positive value indicates evidence in favor of model 1; a negative value indicates the reverse. Following Kass and Raftery [1995], we interpret the magnitude of the test statistic as: 0–2 = ‘not worth more than a bare mention’, 2–6 = positive, 6–10 = strong, and >10 = very strong. The threshold for a ‘significant’ difference was conservatively set at 6.0.

Bayes factors have advantages over other methods such as likelihood-ratio tests or the Akaike Information Criterion. Because they integrate over multiple parameters, it is not necessary to account for numbers of parameters in hypothesis testing. Furthermore, an arbitrary number of non-nested models can be compared, as opposed to likelihood-ratio tests, which are applied to nested models [Lepage et al., 2007].

In Bayesian phylogenetic analysis, marginal likelihoods are often estimated from the post-burn-in samples using the harmonic mean. However, the harmonic mean was shown to be a poor estimator of marginal likelihood, so we used the stepping stone method [Xie et al., 2011] as implemented in MrBayes 3.2.4 [Ronquist et al., 2012a]. The alpha parameter was set to 0.4, 25–50 steps were used, and the burn-in parameter was set to the number of generations in a step (‘−1’), which worked well. Four replicate runs of 10–20 million generations were used for the DNA and combined datasets and 5–10 million for Morph and Morph-Fossil datasets. Convergence of replicate runs was judged adequate if the greatest difference of the likelihoods among runs was <1.0 likelihood units.

In some analyses, partial (backbone) constraints were used on extant taxa only (MrBayes), allowing the fossils to attach to the tree on any branch, regardless of the constraint. The Pipa-Hymenochirus partial constraint forced Pipa, Hymenochirus and Pseudhymenochirus to form a clade. The Xenopus-Hymenochirus constraint used Xenopus, Silurana, Hymenochirus, and Pseudhymenochirus. Each analysis was run under both constraints.
Results

The details of the analyses are summarized in tables 3 and 4 and figures 3 and 4, and we report only the highlights here.

Effects of Non-Clock Analyses on Topology (table 3, columns 1 and 2)

The parsimony analysis of the DNA data (fig. 1; table 3, analysis A1) supported the *Xenopus-Hymenochirus* tree (p = 0.028, AU test; *Hymenochirus + Xenopus + Silurana* = 71% bootstrap support). The MrBayes non-clock analysis (table 3, analysis A2) also favored that XenHym topology. Analysis of the Morph dataset (extant taxa) with parsimony and Bayesian analyses yielded a better score for the PipHym tree under parsimony, but not significantly (B1). However, the MrBayes analysis (C2) supported the XenHym tree strongly.

Parsimony analysis of the Morph-Fossil dataset (fig. 2; table 3, D1) significantly supported the PipHym topology over the XenHym tree (p = 0.0027; WS test; BP = 0.98). Similarly, the MrBayes analysis (fig. 2; D2) strongly supported the PipHym topology with PP = 1.0. Neither parsimony nor Bayesian analyses of the DNA-Morph-Fossil dataset (E1, E2) strongly supported either tree. Not surprisingly, under parsimony the bootstrap support values were 46 and 44% for XenHym and PipHym, respectively.

Effects of Clock Analyses on Topology: Node-Dating

The results of the relaxed-clock analyses are presented in columns 3–6 of table 3 and figure 4. Under a relaxed-clock model without fossil calibrations, the DNA dataset strongly favored the XenHym tree (PP = 0.98) over the PipHym tree (A3). Addition of 3 fossil calibrations very strongly supported the same result (A4). In contrast, analysis of the Morph dataset with or without these calibrations provided either positive but non-significant sup-
### Table 4. Estimated ages for selected nodes

<table>
<thead>
<tr>
<th>Node</th>
<th>Dataset</th>
<th>IGR clock with tree-age prior (MrBayes), MY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no node calibrations</td>
<td>3 node calibrations</td>
</tr>
<tr>
<td></td>
<td>(see table 2, column 3)</td>
<td>(see table 2, column 4)</td>
</tr>
<tr>
<td></td>
<td>median</td>
<td>lower</td>
</tr>
<tr>
<td>Anura DNA</td>
<td>182.8</td>
<td>166.1</td>
</tr>
<tr>
<td>Morph</td>
<td>182.6</td>
<td>166.1</td>
</tr>
<tr>
<td>DNA-Morph</td>
<td>181.2</td>
<td>166.1</td>
</tr>
<tr>
<td>Morph-Fossil</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DNA-Morph-Fossil</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Bombinunura DNA</td>
<td>173.9</td>
<td>144.1</td>
</tr>
<tr>
<td>Morph</td>
<td>166.6</td>
<td>123.9</td>
</tr>
<tr>
<td>DNA-Morph</td>
<td>169.9</td>
<td>139.9</td>
</tr>
<tr>
<td>Morph-Fossil</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DNA-Morph-Fossil</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pipanura DNA</td>
<td>159.6</td>
<td>125.9</td>
</tr>
<tr>
<td>Morph</td>
<td>149.1</td>
<td>101.5</td>
</tr>
<tr>
<td>DNA-Morph</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>Morph-Fossil</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DNA-Morph-Fossil</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pipoidea DNA</td>
<td>139.8</td>
<td>99.8</td>
</tr>
<tr>
<td>Morph</td>
<td>129.9</td>
<td>75.8</td>
</tr>
<tr>
<td>DNA-Morph</td>
<td>144.7</td>
<td>102.9</td>
</tr>
<tr>
<td>Morph-Fossil</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DNA-Morph-Fossil</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pipidae DNA</td>
<td>101.1</td>
<td>61.8</td>
</tr>
<tr>
<td>Morph</td>
<td>90.6</td>
<td>43.6</td>
</tr>
<tr>
<td>DNA-Morph</td>
<td>105.6</td>
<td>67.9</td>
</tr>
<tr>
<td>Morph-Fossil</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DNA-Morph-Fossil</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>XenHym DNA</td>
<td>87.5</td>
<td>51.7</td>
</tr>
<tr>
<td>Morph</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DNA-Morph</td>
<td>90.6</td>
<td>54.5</td>
</tr>
<tr>
<td>Morph-Fossil</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DNA-Morph-Fossil</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PipHym DNA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Morph</td>
<td>54.7</td>
<td>17.3</td>
</tr>
<tr>
<td>DNA-Morph</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Morph-Fossil</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DNA-Morph-Fossil</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Median is the median age of the node. Lower and upper refer to the limits of the 95% credible interval (HPD). Tip-dating analyses were performed only with datasets that included fossils. A subset of these data is presented graphically in figure 4; see related analyses in table 2. No results are presented for Pipanura for the DNA-Morph dataset because the node is the same as Pipoidea due to the exclusion of *Scaphiopus* and *Pelobates* from the analysis. na = Not available.
port (B3) for the PipHym tree or no substantial support for either hypothesis (B4).

The clock analysis of the DNA-Morph data without internal calibrations (C3) did not significantly favor either tree, although there is positive support for the XenHym tree. The incorporation of 3 fossil calibrations, however, yielded strong support for the XenHym tree (C4). In general, the use of fossil calibrations strengthened the evidence for the XenHym model in analyses in which DNA data were included.
Effects of Clock Analyses on Topology: Tip-Dating and Combined Tip- and Node-Dating

Tip-dating was performed on the Morph-Fossil and DNA-Morph-Fossil datasets only (fig. 3). Analysis of the Morph-Fossil dataset favored the PipHym tree strongly (D5). Addition of node calibrations improved the likelihood under both the XenHym and PipHym topologies (compare D5 to D6).

Tip-dating of the DNA-Morph-Fossil dataset (E5) strongly supported the XenHym tree over the PipHym tree. Addition of node calibrations increased the support for the XenHym tree (E6). Under both topologies the node calibrations greatly increased the evidence for the more complex model.

Relationships among Fossil Taxa in Non-Clock Analyses

In the PAUP* and MrBayes analyses of the Morph-Fossil dataset, the position of fossils in the favored PipHym topology was identical. However, in both analyses support for most nodes is weak (fig. 2). The strongly supported clade Pipidae includes a weakly supported subclade consisting of *Xenopus* + *Silurana* and closely related fossils. Within this subclade the relationships among fossils under the 2 analyses are similar, but weakly supported. Only the clade composed of *Shelania pascualii* + its sister-clade has moderate support. The clade *Xenopus* + *Silurana* is strongly supported in both analyses.

The second subclade, which is weakly to moderately supported, includes *Pipa*, Hymenochirini and their fossil relatives. The clade *Pachycentrata†* + *Singidella†* is strongly supported as the sister-group of Hymenochirini. We name this clade (*Hymenochirini + Pachycentrata† + Singidella†*) the Hymenochirinimorpha, following the pattern set in Trueb and Báez [2006]. In turn, *Pipa* is the sister-group of this clade.

Moving toward the root, *Avitabatrachus†* is strongly supported as the sister-group of Pipidae in both analyses. The position of *Palaeobatrachus†* (Palaeobatrachidae†) and its relationship to *Cordicephalus†* is weakly supported. *Neusibatrachus†* is placed as the most basal branch within the stem Pipidae (Pipimorpha). The position of *Chelomorphynus†* as the sister-taxon of *Rhinophrynus* is strongly supported. Overall, these results echo those of previous parsimony analyses [e.g. Báez et al., 2007].

Effects of Clock Analyses on Node Ages

Figure 4 and table 4 summarize the various analyses across a nested subset of selected nodes. For each node, the results of analyses of the 5 datasets are presented in pairs identified by dark and light shades of the same color; the first of each pair (dark colors) does not include node-dating. The second of the pair (light colors) incorporates node calibrations from 3 fossils (*Rhadinosteus†, Pachycentrata†*, and *X. arabiensis†*).

Three analyses (DNA, Morph, and DNA-Morph; red, blue and purple bars) show the effects of inclusion of fossil calibrations (node-dating). These do not include fossils and so were not analyzed by tip-dating (table 4). The fourth and fifth analyses contrast tip-dating with combined tip- and node-dating for the Morph-Fossil and DNA-Morph-Fossil datasets.

The variation in the length of the 95% credible intervals (HPD) across nodes and analyses is systematic (fig. 4; table 4), and the major sources of variation seem to be the dataset and calibration type (node-, tip- or node-tip-dating). In almost all cases, the credible intervals are larger in analyses in which the node calibrations are not used (compare each darker bar with its lighter partner). In the first 3 datasets (red, blue and purple), the median age is older in the analysis lacking internal calibrations. The 2 datasets (green and orange) in which tip-dating was used present a different picture. The upper limits of the 95% HPD and the median ages estimated by the tip-dating analyses are older than those from the combined tip- and node-dating analyses. In all 5 datasets, inclusion of the internal node calibrations increases the precision of the 95% HPD.

For any node, the DNA-Morph-Fossil 95% credible intervals are the most precise (shortest) across all datasets. One striking feature (fig. 4; table 4) of the analyses of Pipioida using 3 calibrations is the narrowness of the 95% CIs. This results from the narrowness of the prior distribution (148.1 MY minimum and 166.1 MY soft maximum) for which the soft upper bound was set so as not to exceed the minimum age of *E. oxoniensis†*.

Palaeobiogeography

As has been found in previous parsimony analyses, the clades of Pipidae do not sort cleanly between South America and Africa (fig. 3). No formal quantitative biogeographic analysis was performed here, but distributional discordance is apparent. Regardless of whether the XenHym or PipHym topology is favored, certain relationships are consistently recovered. The clade Xenopodinomorpha [Trueb and Báez, 2006] occupies 2 continents, with *X. laevis, S. tropicalis* in Africa, and *Shelania†* (2 species), *X.’ romerit, Saltien†*, and *Llankibatrachus†* in South America. The South American Pipa groups with the African *Eoxenopoides†* under the XenHym topology.
**Fig. 3.** Phylogenies estimated from Morph-Fossil and DNA-Morph-Fossil datasets using tip- and node-dating. This corresponds to analyses D6 and E6 in table 3 and the light green and light orange bars in figure 4. The Xen-Hym topology is very strongly supported. The gray bars on the tree are 95% credible intervals; the numbers below the branches are posterior probabilities; the white circles indicate nodes that were constrained; the daggers indicate the positions of the fossils calibrated to the tree-age prior for Anura and the 3 internal calibrations. The shape and limits of the exponential calibration prior are shown for Anura. AF = Africa; AR = Arabia; EU = Europe; NA = North America; SA = South America.
Under the PipHym topology, Pipa and Eoxenopoides† group with Singidella†, Hymenochirus and Pseudohymenochirus. In either case, the closest relative of Pipa is an African taxon. Outside of Pipidae, Avitabatrachus† is the only, and oldest, South American fossil.

The age of the final separation of South America and Africa during the breakup of Gondwana is typically set at ∼100 MY [Veevers, 2004]. The divergence of Pipidae from Avitabatrachus† is ∼120–130 MY (fig. 3), well before the separation. However, the divergences of Xenopus + Silurana from the South American xenopodinomorphs are much younger than 100 MY (∼68–75 MY; fig. 3); in fact, the 95% CI of the divergence does not include 100 MY. The divergences of Pipa from African pipids are slightly older, 80–88 MY, and the CI barely overlaps 100 MY.

**Discussion**

**Topology Differences under Non-Clock Models**

In general, non-clock analysis of DNA and DNA-Morph datasets favored the XenHym topology. The Morph dataset alone did not support either tree significantly (although it favored the PipHym tree), but inclusion of fossils in the Morph matrix tipped support to sig-
nificantly favor the PipHym tree. This suggests that more data help resolve the tree, although the result cannot be attributed to adding extinct taxa per se. The combined 3-dataset analysis supported neither topology. This is likely a canceling effect of combining fossils (generally favoring PipHym) and DNA (favoring XenHym). In this case, the non-clock analysis of all data did not provide a clear answer, and the DNA data did not overwhelm the morphological-fossil data.

Most analyses of pipid relationships using DNA data only have recovered the XenHym topology. However, Bewick et al. [2012] recovered the Pipa-Hymenochirus relationship, which is favored over the Xenopus-Hymenochirus tree by a parsimony length difference of 100 steps. This result is particularly notable because their matrix, which is remarkably complete, contains the largest number of characters relevant to this question: 114 loci of a total length of 35,673 bp, of which 3,166 (8.9%) were parsimony-informative; 5 taxa are represented. Given that the Hedtke et al. [2013] dataset, which also consists of many nuclear genes but more taxa than that of Bewick et al. [2012] (n = 13 vs. 5), supports the XenHym tree, the best explanation would seem to be that a very large number of genes is needed to converge on the correct result. Or, the paucity of taxa and long branches are causing the tree to root in the wrong place.

**Topology Differences from Node- and Tip-Calibrations under Clock Models**

Incorporation of 3 calibrations into the simple clock analyses of the DNA (A3, A4) and DNA-Morph datasets (C3, C4) improved the support for the XenHym tree over the PipHym tree. However, no significant improvement was seen by incorporating node calibrations into the analysis of the Morph data (B3, B4) (table 3).

The analysis of the Morph-Fossil matrix, which included tip-dating, strongly favored the PipHym tree (D5). However, the parsimony and non-clock Bayesian analyses also strongly favored the PipHym tree (D1, D2), so the addition of tip-dating seems to have not added to the discriminatory power of the Bayes factor tests for this dataset. Interestingly, the addition of node calibrations reduced the level of discrimination between the 2 topologies (D5, D6). We have no explanation for this, although it is possible that missing data may play a role.

Matrices used for tip-dating will systematically lack data because no DNA sequences exist for the fossils [Heath and Moore, 2014]. There may be a trade-off between lack of resolution due to the large amounts of missing data caused by a large number of fossils and the increased resolution from reducing the number of fossils to those that maximize information. Our data matrix for fossil taxa is relatively complete compared to many. Nonetheless, many of the nodes are poorly supported in both parsimony and MrBayes analyses. The poor support does not appear to be primarily the result of particular taxa with missing data, based on our limited analyses (not shown).

In contrast to the non-clock analyses, tip-dating of the DNA-Morph-Fossil analysis provided strong support for the XenHym tree, and even more support when the node calibrations were added (E5, E6) (table 3).

Overall, the clock analyses of the 3 datasets including DNA provided strong or very strong support for the XenHym tree. This is most apparent from a comparison of analyses E1 and E2 with E5 and E6. The addition of calibrations of the Morph and Morph-Fossil datasets, either in the form of tip-dating, node-dating, or tip- and node-dating, did not alter the degree of positive or strong support for the PipHym over the XenHym tree.

Without simulations, it is difficult to generalize about the reasons for the various results. It seems clear that adding taxa with morphological data to the Morph dataset improves support. However, calibrated clock analyses of the Morph or Morph-Fossil using either type of fossil calibration do not alter the preference for the PipHym tree. However, this contrasts with analyses of the DNA-Morph-Fossil dataset. Under parsimony and the standard MrBayes analyses of the DNA-Morph-Fossil data, support for PipHym and XenHym topologies was not statistically different. However, under the tip- and tip- and node-dating analyses, XenHym was strongly and very strongly favored over PipHym.

Lastly, it is notable that, even though the addition of node or tip calibrations might not change the support for a topology, in almost all cases the model itself was improved (as measured by marginal likelihoods) when node calibrations were added either to the simple clock model or to the tip-calibrated model (table 3, compare A3 to A4, D5 to D6, etc.).

**Node-Age Differences Based on Tip-Dating and Node-Dating**

We contrasted the effects of dating methods on age estimates by comparing analyses in which no internal calibrations were used to those in which 3 internal calibrations (node-dating) were employed (table 4). Our comparisons are qualitative.

Node calibrations gave older ages than those estimated using only a root calibration for the DNA, Morph, and

---

296 Cytogenet Genome Res 2015;145:283–301
DOI: 10.1159/000438910

Cannatella
DNA-Morph datasets. The *Xenopus* + *Silurana* node is an exception in this comparison and in the others below.

Tip-dating calibrations estimated older ages than those found using only the root calibration for the DNA-Morph and Morph datasets. To minimize missing data, we used tip-dating only for matrices that included morphological data for both extant and extinct species.

Comparison of age estimates for the Morph data showed that the tip-calibrated dates were younger for the 2 oldest nodes (Anura and Bombinatora) and the 2 youngest nodes (most recent common ancestor of *Pipa* + *Hymenochirus*) and *Xenopus* + *Silurana*. The dates were older for the 3 nodes of intermediate age. A similar comparison of ages derived from the DNA-Morph dataset found that the tip-calibrated dates were consistently older than the node-calibrated dates.

The combination of node and tip calibrations gave older ages than those found using only a root calibration for DNA-Morph and Morph datasets.

Importantly, for all 5 datasets and calibration methods, our age estimates are substantially younger than those from several studies based on DNA data (Table 1). It is difficult to infer the causes for this because those analyses were conducted using different classes of prior distributions (e.g. normal, lognormal), calibration points, method of calibration, taxon samples, gene sampling, etc. However, the most obvious difference is that our upper soft maximum for the prior distribution was much younger in this analysis than that of the other analyses.

Using MrBayes, Ronquist et al. [2012a] introduced tip-dating of Hymenoptera under the rubric of ‘total-evidence’. They showed that analyzing fossils as tips improved the precision of age estimates (95% HPD interval was smaller) compared to node-dating. Furthermore, age estimates based on tip-dating were less sensitive to assumptions of the priors than were ages estimated by node-dating.

Wood et al. [2013] noted that when 5 fossil spider taxa were analyzed as tips, the node ages estimated were in general older than when node-dating was used. They also did not find the increase in precision noted by Ronquist et al. [2012a]. Arcila et al. [2015] examined the diversification of tetraodontiform fishes using node- and tip-dating calibrations. In a set of extensive comparisons using BEAST and MrBayes, they also found that tip-dating approaches yielded older age estimates.

These trends are similar to those we found, with the important proviso that we did not use the same set of fossils for tip-dating and node-dating analyses. Our method, however, has the advantage that we were able to assess the complementary effects of node and tip calibrations.

**Effects of Priors**

Our criterion for parameterization of the prior distribution is different from that of Ronquist et al. [2012a; their table 3, our table 2], who set the mean of the exponential prior to the same value as the minimum age of the calibrating fossil at the next deepest node of the tree. This means that a large percentage of values sampled from the prior distribution overlapped the minimum age of the next oldest fossil. Our use of the 97.5% quantile limit avoided this overlap, and no doubt accounts in part for the younger age estimates that we found compared to other studies. The effect of overlap of prior distributions is largely unappreciated [Warnock et al., 2014] and bears investigation.

**Homoplasy**

Why is it that morphological/fossil evidence support one set of relationships and DNA supports the other? The incongruence of trees based on different classes of data is a long-standing conundrum [e.g. Losos et al., 2012]. Resolution is often attempted by combining datasets against each other, in the expectation that one will prevail. In this study, the DNA data have many more informative characters (1,058, a subset of the complete matrix) than do the fossil/morphological data (58), so it is not unexpected that in analyses of all data, the XenHym topology wins out. However, the Bewick et al. [2012] dataset of DNA sequences (3,166 informative sites) supports the PipHym topology, as does the morphology and fossils, in stark contrast to the Hedtke et al. [2013] dataset.

It is worth asking why one dataset is misleading us into accepting the wrong tree, even if we do know which one. This might result from a variety of mechanisms that are both common to and unique to different datasets: positive selection; long-branch attraction due to inappropriate models of molecular evolution or phenotype change; biased taxon sampling; developmental canalization; heterochrony, etc.

Investigation of these mechanisms in the current study is far beyond the scope of this paper. But as an example, it is worth examining the strong support from osteological characters for the clade of extant taxa *Pipa*, *Hymenochirus* and *Pseudhymenochirus*, which is also supported by the fossil characters. *Pipa* (all species) shares many apomorphic features with the hymenochirines that are not present in *Xenopus* or *Silurana* [Cannatella and Trueb, 1988a, b]. Two of these (not part of the Báez et al. [2007] matrix) are the fusion of ulnare to the postaxial centrale (bones of the wrist) and the ossification of intermuscular septa, in the form of bony flanges on the meta-
tarsals, tibiale and fibulare (bones of the ankle and foot). These characters are extremely unusual among frogs. Under the XenHym hypothesis these characters are best interpreted as independently evolved, but why might this be so? Although the skeletons of pipids are in general well-ossified, those of *Pipa* and the hymenochirines are clearly extreme in this regard [Cannatella and Trueb, 1988a, b]. Thus the fusion of bones and the elaboration of bony flanges may result from general hyperossification. This ad hoc hypothesis can be tested by analysis of development mechanisms, for example, gene expression.

**How Do Fossils Affect the Root of Pipidae?**

Bewick et al. [2012] pointed out that the discordance between the rival hypotheses could be interpreted as ambiguity in the placement of the root of the pipid tree. Fossils greatly expand the possibilities for placement of the root, because several placements may result in the same rooting of the extant taxa. This in turn may affect the age of the root.

We have largely confirmed what was found by previous analyses; the DNA data support *Xenopus + Hymenochirus* and the morphological/fossil data support *Pipa + Hymenochirus*. However, what was not evident previously is that the support for the rival topologies is not clear-cut in all cases. The Morph dataset alone (table 3, row B) is indecisive as to which topology is most favored, even under the node-dating analysis; in other words, the position of the root is not clear. However, addition of the Fossil data to the Morph data (table 3, row D), using either a non-clock or clock analysis (with either tip- or node-dating) strongly favors the PipHym tree. Similarly, the parsimony and MrBayes analyses of the combined DNA-Morph-Fossil data, in the absence of a clock (table 3, row E), do not support one hypothesis over the other. However, the addition of tip- or node- and tip-dating to the analysis, tips the ambiguity in the direction of the Xen-Hym topology. Thus, addition of character and/or temporal information from fossils informs the position of the root, although it does not alter the position of the root if that position is already strongly supported.

**Paleobiogeography: Xenopus in Space and Time**

Living pipids are the only archaeobatrachians (excluding *Ascaphus*) with a significant distribution in Gondwana. Fossil pipids are also restricted to Gondwana. However, several of the earliest branching pipoid taxa are Laurasian: *Rhinophrynus, Cheiromorphynus†, Rhadinosteus†, Neusibatrachus†, Gracilibatrachus†* (not analyzed). Thus, the earliest diverging pipoids were associated with Laurasia, and the restriction of Pipidae to Gondwana is more recent.

The sister-group of Pipidae, *Avitabatrachus†*, is also Gondwanan. Given the age of *Avitabatrachus†* (95–99 MY) and the estimated divergence of *Avitabatrachus† + Pipidae* at 120 MY (fig. 3, lower tree, which is the framework for this discussion), we can infer that the ancestor of Pipidae was likely present in Gondwana during the splitting of South America and Africa at ~100 MY [Vevers, 2004], and that this represents a vicariant event separating pipid lineages. If the rifting of South America and Africa formed an impassable barrier to early pipid lineages, then we would expect 2 clades with each restricted to South America and Africa. The DNA phylogeny of extant pipids seems to support this. *Pipa* is restricted to South America, and the remaining taxa form an African clade.

However, integration of fossil data demonstrates the deceptiveness of this pattern. The clade Xenopodomorpha occupies 2 continents, as does the clade *Pipa + Eoxenopoideust*. However, within Xenopodomorpha, the divergence of *Xenopus + Silurana* from the South American taxa is only ~68 MY, and its 95% CI does not overlap 100 MY. Similarly, the divergence of *Pipa from Eoxenopoideust* occurred at about 81 MY, and its 95% CI barely overlaps 100 MY. It is challenging to reconcile these ages with geological history, under the assumption that over-water dispersal of pipids after the rift would not be possible.

Given that *Xenopus + Silurana* are embedded within a group of South American taxa suggests strongly that the presence of this clade in Africa represents an eastward dispersal at ~76–86 MY, depending on the topology. Variance does not seem a feasible explanation. Likewise, the presence of relationships of *Pipa to African fossils*, under either topology in figure 3, suggests westward dispersal from Africa. We view these statements as hypotheses in need of rigorous testing by formal analysis that incorporates both biogeographic and tip-dating analysis [e.g. Wood et al., 2013].

Bewick et al. [2012] used *BEAST* multilocus coalescent analyses, which they calibrated using the rifting of Africa from South America at ~100 MY, and the separation of West Africa from North America at ~190 MY. In general, they found that the divergence of *Pipa from Xenopus* occurred at 109.4, 134.7, or 143.4 MY, depending on which combination of calibrations was used. These dates suggest that *Xenopus* diverged from *Pipa* before or at the separation of South America and Africa at ~100 MY. The first date is consistent with our estimates, but the latter 2 are much older (table 1).
However, the interpretation of the biogeographic discordance among fossil pipids does not fit the vicariance model. Although Estes [1975], Canntella and de Sá [1993] and Báez and Pugener [1998] concluded that the several lineages of fossil pipids must have been present before the continental breakup, this study does not support this model. Rather, overwater dispersal seems to be necessary to explain the phylogenetic and biogeographic patterns of fossil pipids, specifically, the finding that Xenopus and Silurana are embedded in a clade of South American fossils from which they diverged only ~68–75 MYA. Thus, 2 different interpretations of historical biogeography of Xenopus result from differences in the choice of a soft maximum, the incorporation of temporal and phenotypic data from fossils, and the use of calibrations from geological information.

If our inference of the timing of divergences, and thus the necessity of dispersal, is correct, how then does one reconcile this with geological history? Pipids are not the only taxa that display this Gondwanan discordance. For example, Poux et al. [2006] investigated the timing of divergences of New World plathyrrhine monkeys from Old World catarrhines and of New World caviomorph rodents from their Old World relatives. These patterns have long existed as biogeographical enigmas, in part because both groups appear in the late Oligocene fossil record in South America. Poux et al. [2006] concluded that the both New World groups originated by trans-Atlantic migration in the Eocene.

De Oliveira et al. [2009] reviewed the models of dispersal that might account for these distributions from a geological perspective. Their reconstructions indicate the presence of large islands in the South Atlantic at ~40–50 MY. These were not land bridges, but rather emergent land areas that would have reduced the distances for overwater dispersal by hundreds of kilometers. Ezzurra and Agnolin [2012] built upon this evidence, positing a new palaeobiogeographical model to explain overlooked but unexpected patterns in several groups, primarily late Mesozoic archosaurs.

Although the support for the phylogeny is less than optimal, pipids may add one more challenge to the paradigm of Gondwanan vicariance. Xenopus (and Silurana) may be a relative recent immigrant to Africa, having entered Africa ~40–75 MYA, rather than part of an ancient lineage present in Africa before it split from South America.

Acknowledgement

I thank David Blackburn for discussions about fossils and pipoid frogs.

References


Irissari I, San Mauro D, Abascal F, Oehler A, Vence- ces M, Zardoya R: The origin of modern frogs (Neobatrachia) was accompanied by ac- celeration in mitochondrial and nuclear sub- stitution rates, BMC Genomics 13:626 (2012).


Xenopus in Space and Time

Cytogenet Genome Res 2015;145:283–301
DOI: 10.1159/000438910