AmphiNom: an amphibian systematics tool

H. Christoph Liedtke

➕ View supplementary material ⇗

📅 Published online: 22 Oct 2018.

✏️ Submit your article to this journal ⇗

📊 Article views: 1

View Crossmark data ⇗

Perspective
AmphiNom: an amphibian systematics tool

H. CHRISTOPH LIEDTKE ⓘD

Ecology, Evolution and Development Group, Department of Wetland Ecology, Estación Biológica de Doñana (CSIC), 41092 Sevilla, Spain

Large-scale comparative and systematic studies rely on the seamless merging of multiple datasets. However, taxonomic nomenclature is constantly being revised making it problematic to combine data from different resources or different years of publication, which use different synonyms. This is certainly true for amphibians, which have experienced a spike in taxonomic revisions in part as the result of the widespread use of DNA barcoding to resolve cryptic species delimitation issues and large-scale collaborative efforts to revise the entire amphibian tree. The 'Amphibian Species of the World Online Reference' (ASW) is one of the most widely used and most regularly updated databases for amphibian taxonomy, but existing R tools for querying synonyms such as 'taxize' do not include this resource. 'AmphiNom' is a tool suite written in the R programming language designed to facilitate batch-querying amphibian species names against the ASW database. This facilitates the merging of datasets that use different nomenclature and its functionality is easily integrated into customizable R workflows. Moreover, it allows direct querying of the ASW website using R and straightforward reporting of summary information on current amphibian systematics.

Key words: Amphibia, comparative analysis, phylogenetics, R, software, systematics

Introduction

What's in a name? – With classic taxonomy becoming an imperilled discipline, it is pertinent to remember that the use of taxonomic names and the inherent hierarchical structuring of Linnaean classification is fundamental to most biological studies. Species are used as the most common delimiting units of groups of organisms and are important for conservation initiatives (Mace, 2004), estimates of biodiversity and evolutionary diversification (Bickford et al., 2007; Liedtke et al., 2016) and for at least some phylogenetic methods (e.g., Heled & Drummond, 2010). Despite efforts to standardize naming (e.g., International Commission on Zoological Nomenclature; http://iczn.org/), not all names are universally accepted, many delimitations of species are problematic resulting in regular name changes, and there is often no adherence to a single taxonomy across publications or online resources. Comparative studies are often heavily reliant on combining data from such resources, a task that requires inspecting and unifying species synonyms so that datasets can be merged. This can be time-consuming and error-prone when done

manually and there is thus a need to facilitate this process, ideally in a programmatic fashion that retains a log of changes made, and allows batch-querying and integration into existing analysis workflows (e.g., Chamberlain & Szöcs, 2013).

Amphibians are a widely used system for comparative studies and their systematics is particularly turbulent, with taxonomy constantly being updated through the discovery of new species and through reassessments of current species status resulting in division of described units into multiple species ('splitting') or combining multiple species into one ('lumping'). New insights gained from molecular barcoding has contributed to this in a major way by facilitating the detection of cryptic species (Vieites et al., 2009) as well as the recognition of polymorphism (Liedtke et al., 2014). Moreover, global taxonomic endeavours such as The Amphibian Tree of Life and associated Amphibian Species of the World Online Reference (AWS; Frost, 2018) and AmphibiaWeb (2018), regional projects such as Anfibios de Ecuador (Ron, Yanez-Muñoz, Merino-Viteri, & Ortiz, 2018), AmphibiaChina (2018) and class-wide molecular phylogenies (e.g., Pyron & Wiens, 2011) have resulted in contested renaming at almost all taxonomic levels (Frost et al., 2008; Wiens et al., 2007).

Correspondence to: H. Christoph Liedtke. Email: christoph.liedtke@ebd.csic.es

As a result, primary literature and different resources that are frequently queried for amphibian studies such as the IUCN (2017), GenBank (Benson et al., 2013), Global Biodiversity Information Facility (GBIF.org, 2018), or AnAge (Tacutu et al., 2017) use differing synonyms and combining data from these based on species names used can be problematic. Here, I introduce a tool written in the R programming language for streamlining the merging of such datasets, based on the record of synonyms curated by ASW, arguably the most widely used, and most regularly updated amphibian taxonomic list.

Web tools useful for resolving naming issues resulting from different synonyms already exist at a broader taxonomic scale, such as Global Names (http://resolver.globalnames.org/) and the Integrated Taxonomy Information System (https://www.itis.gov/), but what is needed is a system that facilitates working with multiple and large datasets in a script-based fashion. This has the advantages of allowing batch processing of many names at once, being repeatable, updateable, and leaving a record of changes made, as well as the possibility of integrating code into an existing R workflow for downstream analysis. The R package 'taxize' (Chamberlain & Szöcs, 2013) is a web Application Programming Interface (API) dependent tool that allows querying of the above-listed taxonomic resources in such a script-based way, but unfortunately it does not search the ASW database and as such is limited when it comes to amphibian studies. 'AmphiNom', the application described herein, aims to fill this gap by allowing the harvesting of the ASW website to reconstruct the most current taxonomic hierarchy as well as a list of all documented synonyms per species. 'AmphiNom' is not a tool designed for automatizing the meticulous process of taxonomic revisions; instead its functions facilitate the standardization of naming systems across datasets as well as allowing for direct querying of the ASW species pages and the compilation of statistics on the systematics of desired branches of the amphibian tree of life.

Brief software description

The application 'AmphiNom' is written in the R programming language (R Core Team, 2017) and consists of two components (Figure 1); the first comprising functions that harvest the ASW website and construct datasets on amphibian systematics and synonyms as well as accompanying summary and species-specific search functions, and the second comprising functions that allow querying lists of species names against the constructed ASW datasets.

ASW is a database-driven website based on the PHP programming language and is structured so that the URL consists of a stem (http://research.amnh.org/vz/herpetology/amphibia/) that is then extended as one descends the rungs of the taxonomic ladder, so that the final extension may include $<$order$>$/$<$superfamily$>$/$<$family$>$/$<$subfamily$>$/$<$genus$>$/$<$species$>$. At each level, the webpage returns a list of all nested units at the next taxonomic rung. The function *getTaxonomy()* takes advantage of this conserved structure by starting with the common URL stem and then reading the list of nested units to extend the URLs accordingly, one taxonomic rung at a time. This process is then iterated as many times as needed until the URLs are fully extended for each species (theoretically a maximum of five times, but less if taxonomic rungs such as superfamily are absent). The species-specific webpages list all synonyms for that particular species and the underlying HTML code is constructed so that all synonyms are contained within a single element of $<$div$>$ class 'synonymy'. The *getSynonyms()* function extracts all synonyms listed within this element.

The current version of 'AmphiNom' (v1.0.0) contains four additional functions as well as three datasets (Table 1). The core of these additional functions, *aswSync()*, takes a query as an input (a list of species), matches these names to the list of synonyms generated with *getSynonyms()* and either confirms whether the query matches the most current name according to ASW or suggests the appropriate update, where this is unambiguously possible. If a query is listed as a synonym of more than one species, all possible matches are returned with a warning. The *searchASW()* function allows for direct information retrieval for any species listed on ASW and the remaining two functions *aswStats()* and *synonymReport()* facilitate summarizing the outputs of 'AmphiNom' (Table 1).

The latest releases of 'AmphiNom' can be downloaded from GitHub (https://github.com/hcliedtke/AmphiNom).

Examples

Both a reference manual as well as a tutorial showing the full features of the 'AmphiNom' are provided as package vignettes and on GitHub. Here I will detail a short example of the application's intended use, with the relevant code in Figure 2 and a more extensive example as Appendix 1 (see online supplemental material, which is available from the article's Taylor & Francis Online
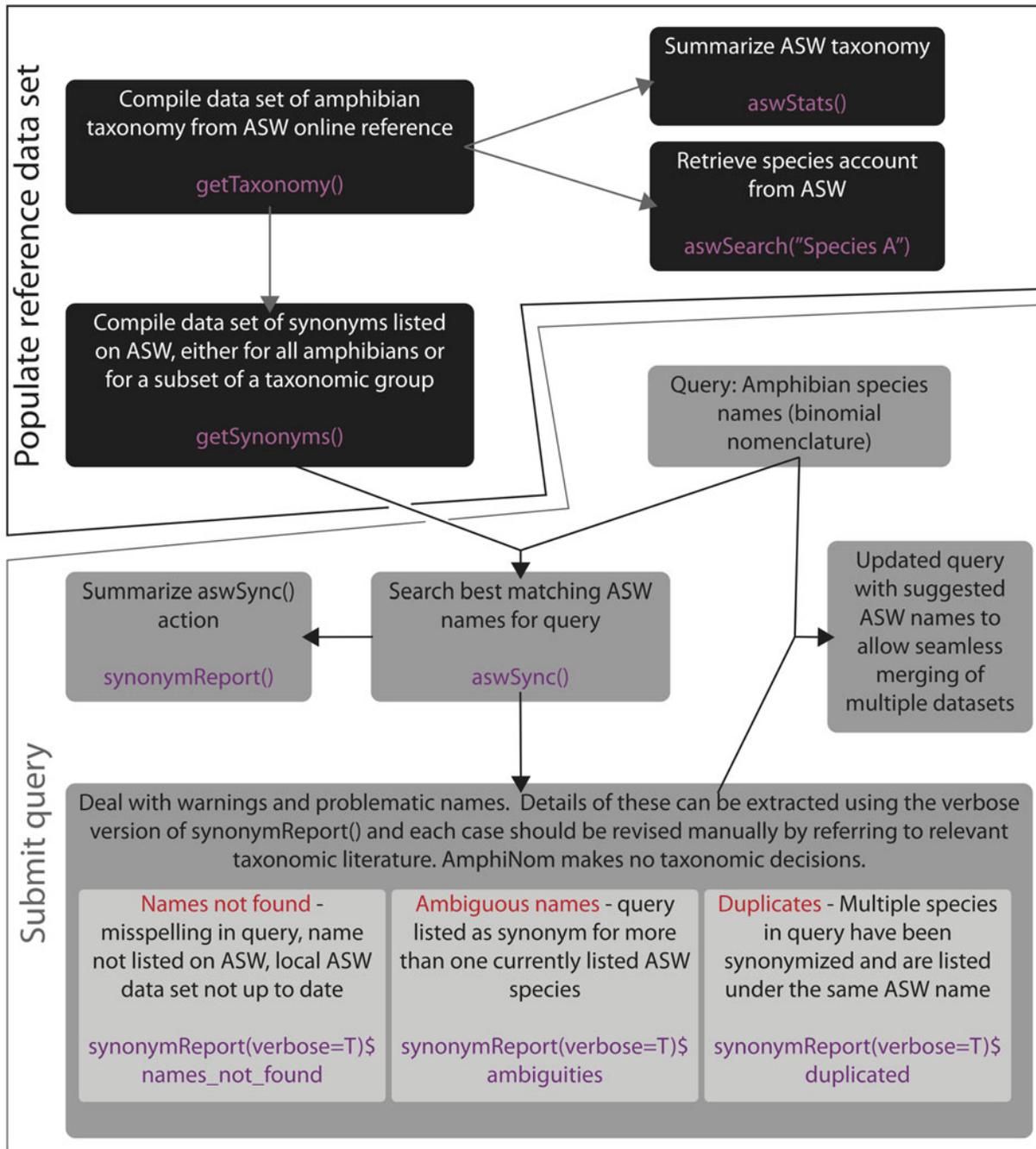
Fig. 1. Pipeline for the intended use of 'AmphiNom' for retrieving information from the Amphibian Species of the World (ASW) online database (black) and for querying a list of species names to conform to ASW taxonomy (grey). R functions are given in purple and require the resulting data object from the previous step as input.

Harvesting ASW

The first step is to compile datasets detailing the taxonomy and synonyms as archived on ASW (Figure 1). This is done sequentially, first by running *getTaxonomy()* with no arguments and then *getSynonyms()* on the dataset produced by *getTaxonomy()*. Copies of the resulting outputs are also

Table 1. List of functions and datasets currently available in 'AmphiNom' v1.0.0 and their descriptions.

Type	Function	Description
Web scraping	getTaxonomy	Iterates through the ASW website to compile a dataset of all listed amphibian species and their higher-level taxonomy. Includes a URL string for each species webpage.
	getSynonyms	Uses URLs generated from *getTaxonomy()* to compile a dataset of all listed synonyms per species.
Data processing	aswSearch	Retrieves species account information from the ASW website. Includes current scientific name and author, full taxonomy, detailed list of synonyms, common names, distribution note and other taxonomic comments.
	aswSync	Matches query (list of species names) against list of synonyms generated with *getSynonyms()* and returns suggested updates to names. Notifies users of problematic names either because of taxonomic ambiguities, duplications, or names not currently listed on ASW.
Summaries	aswStats	Returns statistics on the number of downstream taxonomic units contained within a given taxonomic unit.
	synonymReport	Returns summary statistic on the object created by the *aswSync()* function.
Datasets	asw_taxonomy	Internally stored version of the output generated by the *getTaxonomy()* function.
	asw_synonyms	Internally stored version of the output generated by the *getSynonyms()* function.
	amphweb	Internally stored version of the www.amphibiaweb.org taxonomy.

stored in the package internally for exemplary use (*asw_taxonomy* and *asw_synonyms* respectively), but both these datasets are only updated with package updates and so it is recommended that users compile their own taxonomy and synonym tables to ensure these are up to date. At the time of writing (2.1.2018), the *aswStats()* run on the internally stored *asw_taxonomy* dataset from the same date counts 7776 species listed on the ASW website classified into 549 genera and 75 families. For these, a total of 30,014 synonyms are documented in the AmphiNom-generated *asw_synonyms* table.

Synchronizing names

The molecular, time-calibrated phylogeny of Pyron (2014) is the most complete phylogeny for amphibians to date in terms of species numbers and the anuran subtree is available on dryad (http://dx.doi.org/10.5061/dryad.jm453). Since this phylogeny was published, Duellman, Marion, and Hedges (2016) revised the taxonomy of treefrogs and raised the subgenus *Dryophytes* to genus level, affecting the nomenclature of North American and some Asian species of *Hyla*. The ASW has adopted the changes by Duellman et al., (2016) as has the IUCN Red List database, a commonly used resource for the natural history and conservation status of amphibians. To be able to use both the divergence time information from the phylogeny and, for example, distribution range information provided by the IUCN Red List, the names in the *Hyla* clade must be standardized across the two datasets. The most straightforward way to do this is to 'update' tip labels on the phylogeny to the ASW names, which can then be used as search

terms to extract and plot the distribution maps provided by the IUCN database, as shown in Figure 2.

The true utility of this package is for working with large datasets (exemplified in Appendix 1, see supplemental material online). The amphibian chronogram presented in Pyron (2014) contains 3310 species and if we were to extend the *Hyla* example to the whole tree, inspecting and updating all names manually is not practical and error-prone. Running the *aswSync()* function on all the tip labels followed by the *synonymReport()* function on the resulting data frame indicates that only 2551 names are direct matches to species names on the ASW and 721 more were unambiguous synonyms for ASW-listed species. 'AmphiNom' has done most of the heavy lifting and a researcher is required to more carefully check only the remaining 38 names that could not unambiguously be traced to a single ASW species name. Specifically, 'AmphiNom' tells us that 29 of these require a closer look because the names are synonymous with more than one species, nine species names are not listed in the ASW database (usually spelling mistakes in query) and additionally, we are given a warning that the 721 suggested name updates would result in 64 duplicates, most frequently because the query contains multiple names that are treated as sub-species (and have therefore been 'lumped') by ASW.

Acknowledgements

```r
# load required libraries
library(ape)
library(AmphiNom)
library(maptools)
library(maps)

# set working directory
setwd("~/amphinom_example") #path may differ

# general stats on state of ASW database
aswStats()
```

```
##      number_of_units
##Oder              3
##Families         75
##Genera          549
##Species        7776
```

```r
# Download phylogeny from dryad
pyron_anura<-read.tree("http://datadryad.org/bitstream/handle/10255/dryad.63062/amph_shl_dates
_frogs.tre?sequence=1")

# Keep only the Hyla clade for this example
hyla_tree<-drop.tip(pyron_anura, pyron_anura$tip.label[!pyron_anura$tip.label %in%
grep(pyron_anura$tip.label, pattern="Hyla_", value = T)])
par(mfrow=c(1,2))
plot(hyla_tree, no.margin=T)

# aswSync() the tip labels of the phylogeny to inspect what names should be updated
asw_hyla<-aswSync(hyla_tree$tip.label)
synonymReport(asw_hyla)
tiplabels(asw_hyla$ASW_names, frame="none", adj=-1,
col=ifelse(grepl(x=asw_hyla$ASW_names,
pattern="Hyla_"),"#2ECC71", "#1C75BC"))
```

```
##                          number_of_units
##queries                             26
##names_up_to_date                    10
##names_successfully_updated          16
##names_not_found                      0
##ambiguities                          0
##duplicates_produced                  0
```

```r
# download IUCN distribution shape files for all
anurans from the IUCN read list website and load
these into R. Spatial data available at:
http://www.iucnredlist.org/technical-documents/spatial-data
anura_iucn<-readShapePoly("./ANURA/ANURA.shp") #path may differ

# plot world map
par(mar=c(0,0,0,0))
map("world", interior=F, border = NA, fill=T, col="grey90")

# plot only range maps of Hyla species included in the phylogeny and colour the ranges of Hyla
and Dryophytes species differently
for(i in 1:length(asw_hyla$ASW_names)){
     plot(anura_iucn[anura_iucn$binomial==gsub(asw_hyla$ASW_names[i], pattern="_",
replacement = " "),], add=T, col=ifelse(grepl(asw_hyla$ASW_names[i],
     pattern="Hyla"),"#2ECC7190", "#1C75BC90"), border=NA)
}
```
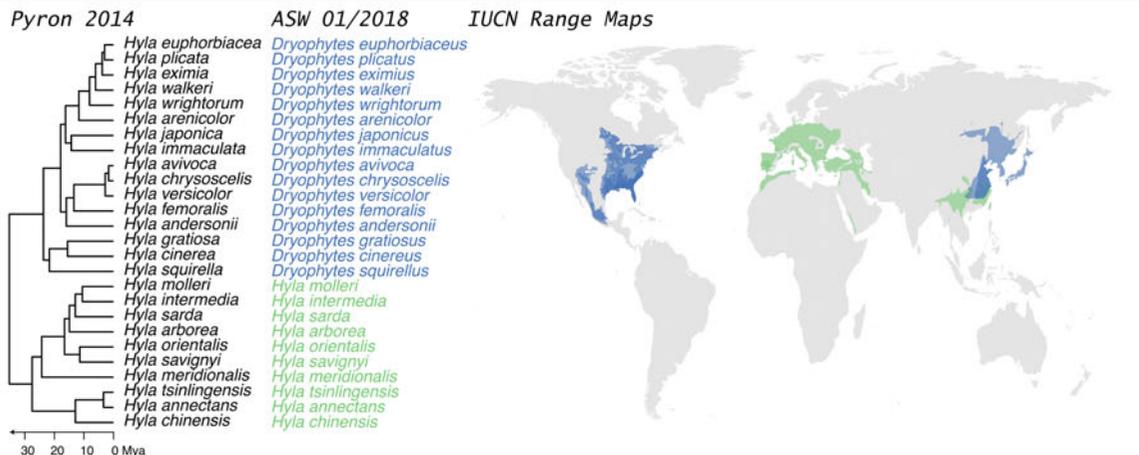

Figure: Hyla *clade of the* Pyron (2014) *phylogeny with original and updated names. World map shows the distribution of the remaining members of the genus* Hyla *(green) and the newly reclassified* Dryophytes *(blue).*

Fig. 2. R code for printing statistics on current Amphibian Species of the World (ASW) taxonomy and a short example of standardizing names to facilitate merging datasets.

ORCID

H. Christoph Liedtke (iD) http://orcid.org/0000-0002-6221-8043

References

AmphibiaChina. (2018). *The database of Chinese amphibians*. Kunming, Yunnan, China: Kunming Institute of Zoology (CAS). Retrieved from http://www.amphibiachina.org/ (accessed 6 October 2017).

AmphibiaWeb. (2018). *AmphibiaWeb*. Berkeley, CA, USA: University of California. Retrieved from https://amphibiaweb.org (accessed 01 January 2018).

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research*, *41*, 36–42.

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., … Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, *22*, 148–155.

Chamberlain, S. A., & Szöcs, E. (2013). taxize: taxonomic search and retrieval in R. *F1000Research*, *2*, 1–28.

Duellman, W. E., Marion, A. B., & Hedges, S. B. (2016). Phylogenetics, classification, and biogeography of the treefrogs (Amphibia: Anura: Arboranae). *Zootaxa*, *4104*, 1–109.

Frost, D. R. (2018). *Amphibian Species of the World: an Online Reference*. Version 6.0 New York, USA: American Museum of Natural History. Retrieved from http://researchamnhorg/herpetology/amphibia/indexhtml (date of (accessed 01 January 2018).

Frost, D. R., Grant, T., Faivovich, J., Bain, R. H., Haas, A., Haddad, C. F. B., … Wheeler, W. C. (2008). Is The Amphibian Tree of Life really fatally flawed? *Cladistics*, *24*, 385–395.

GBIF.org. (2018). *GBIF Home Page*. Retrieved from: https://www.gbif.org (accessed 02 January 2018).

Heled, J., & Drummond, A. J. (2010). Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, *27*, 570–580.

IUCN. (2017). *The IUCN Red List of Threatened Species*. Version 2017-3. Retrieved from http://www.iucnredlist.org (accessed 01 February 2018).

Liedtke, H. C., Hügli, D., Dehling, J. M., Pupin, F., Menegon, M., Plumptre, A. J., … Loader, S. P. (2014). One or two species? On the case of *Hyperolius discodactylus* Ahl, 1931 and *H. alticola* Ahl, 1931 (Anura: Hyperoliidae). *Zootaxa*, *3768*, 253.

Liedtke, H. C., Müller, H., Rödel, M.-O., Menegon, M., Gonwouo, N. L., Barej, M. F., … Loader, S. P. (2016). No ecological opportunity signal on a continental scale? Diversification and life-history evolution of African true toads (Anura: Bufonidae). *Evolution*, *70*, 1717–1733.

Mace, G. M. (2004). The role of taxonomy in species conservation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*, 711–719.

Pyron, R. A. (2014). Biogeographic analysis reveals ancient continental vicariance and recent oceanic dispersal in amphibians. *Systematic Biology*, *63*, 779–797.

Pyron, R. A., & Wiens, J. J. (2011). A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Molecular Phylogenetics and Evolution*, *61*, 543–583.

R Core Team. (2017). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, Retrieved from http://www.R-project.org/ (accessed 01 January 2018).

Ron, S. R., Yanez-Muñoz, M. H., Merino-Viteri, A., & Ortiz, D. A. (2018). *Anfibios del Ecuador*. Version 2018.0. Museo de Zoología, Pontificia Universidad Católica del Ecuador. Retrieved from https://bioweb.bio/faunaweb/amphibiaweb (accessed 20th March 2018).

Tacutu, R., Thornton, D., Johnson, E., Budovsky, A., Barardo, D., Craig, T., … de Magalhães, J. P. (2017). Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Research*, *46*, D1083–D1090.

Vieites, D. R., Wollenberg, K. C., Andreone, F., Köhler, J., Glaw, F., Vences, M., & Wake, D. B. (2009). Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 8267–8272.

Wiens, J. J. (2007). Review of "The amphibian tree of life" by Frost et al. *Quarterly Review of Biology*, *82*, 55–56.

Associate Editor: Mark Wilkinson