

LETTERS TO THE EDITOR

Herpetological Review, 2016, 47(2), 330–334.
© 2016 by Society for the Study of Amphibians and Reptiles

The Reptile Database Turns 20

It is hard to believe, but the Reptile Database turned 20 in November 2015. In this article, I aim to provide a personal retrospective of its origin and how the history of the database ties in with developments in biodiversity research, taxonomy, and the digitization of systematics over the past 20 years or so. I will also discuss some ideas for the future of the database.

Humble beginnings.—Although I have been interested in reptiles since my early childhood, my interests drifted somewhat away from herpetology while I was still in high school (in the mid-1980s), not the least because of the advent of DNA sequencing which had been invented by 1977 (Sanger et al. 1977). The main reason why DNA sequences were so fascinating was the fact that they must somehow encode *what an organism is and does*—which obviously included reptiles, too. It was a coincidence that personal computers became available at around the same time. With a friend, Thomas Kuner, now a professor in neurobiology in Heidelberg, Germany, we developed a computer program for DNA sequence analysis for *Jugend forscht*, a science competition among high school students in Germany. To get the sequence data, we drove to the European Molecular Biology Laboratory (EMBL) in Heidelberg to pick up a stack of 20 floppy discs that had the whole (!) EMBL DNA sequence database (there was no internet at the time and CDs were not widely used for data yet). Although our program was quite primitive by today's standards, we made important contacts with EMBL which would become critical years later (and even won a national fifth prize at *Jugend forscht* in 1986).

A PhD out on a limb.—Even before I went to college, it had become obvious that developmental processes and thus morphology are determined by gene expression and transcription. I wanted to study these problems and eventually ended up as a graduate student in the lab of Rolf Zeller at EMBL in 1993. Rolf studied limb development in chicken and mouse embryos (and still does!), and his lab appeared to be the right place to study how genes determine morphology, especially limb development (or the lack thereof) which is also critical for all reptiles. As a thesis project I was given the problem to figure out what a particular gene, called *limb deformity*, was doing during limb development. In fact, I did find out that *limb deformity* proteins reorganize the actin cytoskeleton in cells (Uetz 1997). However,

we never published these results, because they did not explain the limb phenotype and a molecular explanation became clear only years later (Pavel et al. 2007). Although my thesis project had only indirectly to do with reptiles or databases, it laid the basis in several ways.

Internet databases.—While I was a graduate student during the 1990s the genomics revolution took off with the first (bacterial) genomes being published in 1995 (Fleischmann et al. 1995). Somewhat coincidentally, the internet, with the advent of the WWW and user-friendly web browsers, also took off in the early 1990s. Luckily I was in the right place at the right time as EMBL had one of the first institutional web servers that hosted the EMBL DNA sequence database (now at the European Bioinformatics Institute in Cambridge, UK). With the suddenly growing amount of DNA data, online databases became quickly critical for molecular biologists. However, while hosting DNA sequences is easy it became clear pretty quickly that this is not that simple for embryological information. So, I suggested to my thesis advisor that we should set up a database on gene expression patterns. He was *not excited*. I guess he was not even interested. In any case, I was thrilled by the idea to set up an online database. Because my supervisor did not like the idea of a gene expression database and because it seemed like a very ambitious idea anyway I decided to try a taxonomic database instead. After all, I had all those reptile books sitting around. Ironically, many books mentioned that there are about 6000 reptile species but no one seemed to have compiled a list of these species—or at least I did not know of such a list.

The EMBL Reptile Database.—The Reptile Database was born in November 1995 as a few web pages on my personal web space at EMBL that simply had lists of all reptile names that I had compiled by that time. While I was setting up that list, it became clear that this was a rather pointless exercise. What was needed was a searchable database, similar to the DNA sequence database at EMBL. So, I asked Thure Etzold, who had set up the web server for the sequence database, whether we could use the same system for a taxonomic database. To my surprise he said, “sure, no problem, just give me a floppy disc with the data,” which I did within a few days. I do not remember the exact date, but by the end of 1995 or early 1996 we had our first searchable Reptile Database online—even though there was not a whole lot of data. Technically it was the same as the DNA sequence database and a number of other molecular biology databases that EMBL was hosting at the time. We called it “The EMBL Reptile Database” and it kept that name until 2006, long after I had left EMBL in 1997. Unfortunately, the last member of

PETER UETZ

Center for the Study of Biological Complexity,
Virginia Commonwealth University, 1000 W Cary St, Richmond,
Virginia 23284, USA; e-mail: peter@uetz.us

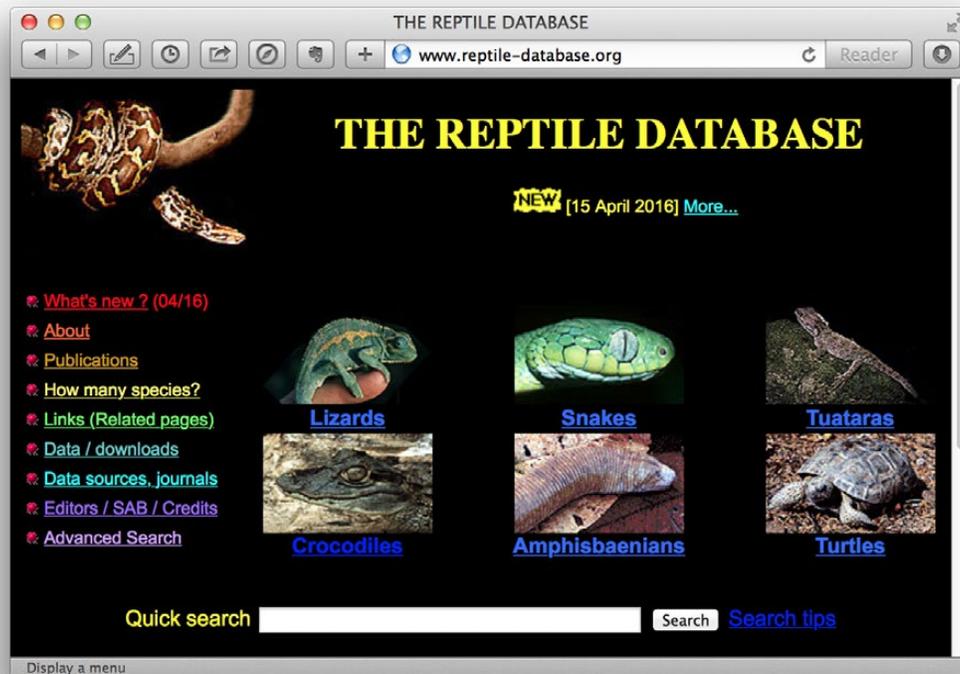


FIG. 1. The home page of the Reptile Database has barely changed in 20 years and is in dire need of an overhaul. However, there are more functions and a dramatically increased dataset.

the database team who kept the database alive, Ramu Chenna, left EMBL in 2006, and the database team asked us to move it somewhere else.

Species 2000 and the funding problem.—The Reptile Database has always been a volunteer project and none of the participants were ever compensated. However, we did get two small grants from the Species 2000 project (which later became the Catalogue of Life project, CoL). Frank Bisby, a professor of botany at the University of Reading, UK, had founded the Species 2000 project in 1997, with the goal of compiling a catalogue of all species worldwide (Cachuela-Palacio 2006). His vision was to create a meta-database that integrates individual taxon-specific databases such as the Reptile Database. Although the Reptile Database did not have much data in its first few years, Frank managed to obtain a fairly substantial grant from the European Union to fund the Species 2000 project, which included lump sums of about US \$25,000 for each contributing species database. Part of that money had to be spent on the development of wrapper software that allowed the central Species 2000 website to query individual databases. Although this idea worked reasonably well for a while, databases change and after the EU grant ran out many databases lost connectivity, including the Reptile Database. Johannes Goll, a diploma student in my lab at the time, had implemented the wrapper on a PC that acted as a local server at the Karlsruhe Institute of Technology where we worked from 2001 through 2006 (Uetz et al. 2007). So, Bisby and his team convinced the EU to fund a follow-up project to integrate more databases and to develop converters for the TDWG format (devised by the

Taxonomic Databases Working Group), so that all databases could be imported into a single database and hosted centrally (now from the Catalogue of Life web site). Thus, much of the limited funding we have received went into the wrapper and conversion software. However, I managed to use some of the CoL grant to pay students to scan key historical books, such as Boulenger's catalogues and Duméril and Bibron's *Erpetologie Générale*, whose names were indexed in the database very early on.

The TIGR Reptile Database.—By 2006, when Ramu left EMBL, I had accepted a job at The Institute for Genomic Research (TIGR), which was a convenient coincidence. TIGR had not only sequenced the first bacterial genome (Fleischmann et al. 1995), but one of their software engineers, Hank Wu, and an intern, Meher Preethi Boorgula, offered to set up a replacement database, which we named the “TIGR Reptile Database.” Unfortunately, the TIGR Reptile Database got into the crossfire of an internal power struggle during which TIGR was merged with the J Craig Venter Institute (JCVI) only about a year after I had arrived. As a consequence, the TIGR brand literally ceased to exist. Nevertheless, the TIGR Reptile Database was in operation for four years, from 2006 to 2010, after which I left the JCVI to accept a faculty position at Virginia Commonwealth University in 2011.

The Reptile Database and Reptarium.—Luckily, shortly after I had left the JCVI, Jiří Hošek, a Czech programmer and webmaster of the Czech Herp Society, offered to set up a replacement database and web server in the Czech Republic (at reptarium.cz). Although this was the third move of the database within a few years and despite the fact that the database was offline for a few

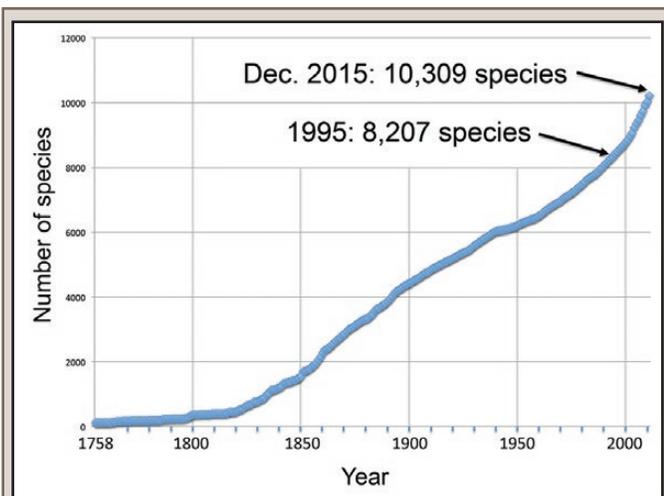


FIG. 2. The database has grown by more than 2000 species since its inception. In addition, almost 20,000 papers and notes have been published since 1995 and are cited in the database. Updated after Uetz 2010.

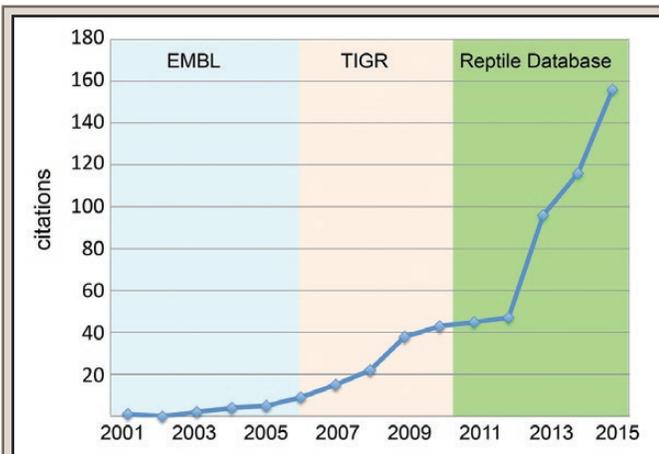


FIG. 3. It took several years until the database got taken seriously and cited in the scientific literature. The citations shown here are based on a total of 766 citations from Google Scholar, as of 9 May 2016 (with those citations left out that do not show a publication year). The three different phases of the EMBL, TIGR, and “unnamed” Reptile Database are indicated by shading. The slowdown of citations in 2011 and 2012 may be due to the move to reptarium.cz and the uncertainty of its future.

weeks, the transition was relatively smooth. First, we had used the reptile-database.org URL since 2001 and thus users did not lose access completely. Second, Jiř did an excellent job of migrating the whole database and designing a new and improved interface pretty quickly.

Users then and now.—Not surprisingly, the herpetological community was initially reluctant to use the database. In the early days of the database I wrote letters (yes, printed letters, not emails!) to experts to ask them for lists of species or other information. One of my first targets was Miguel Vences, who responded something along the lines of “there have been other such projects and they all failed—why should we trust you?” Hence I was rather surprised that some of my early articles in *Herpetological Review* (Uetz and Etzold 1996; Uetz 2000) were reviewed positively, which certainly helped the database to give it scientific credibility. Sure enough, over time the database got better and

by now has reached the same reliability as its sister databases of other taxonomic groups. This has also been recognized by the scientific community, which has cited it a thousand times in the herpetological literature (>600 in Google Scholar at the time of this writing). A Google search for <http://www.reptile-database.org> returns about 450,000 hits. Google Analytics tells us that the database has more than 32,000 users per month that produce more than 200,000 page views in more than 50,000 sessions (Oct 2015). This number has roughly stabilized over the past few years but we may be able to increase that with improved features. But what features do taxonomic databases need in the future?

The future.—First, we have no plans to give up—there is simply so much to be done! Although I have put in an estimated 15,000 hours of work (that’s only an hour or two per species) there are an infinite number of things that can be improved. However, what *will* be done depends both on contributions and help from volunteers, but also on funding. Below are some ideas and plans.

What information should be in the database?—Although the Reptile Database is primarily a taxonomic database, many other biological characters can be of taxonomic relevance, e.g., geography and ecology, behavior, genetics, etc. Due to our limitations with manpower and funding, however, not everything can be collected even if it were relevant. We try to alleviate this problem by linking to other databases and resources.

Distribution and biogeography.—Clearly, species ranges have been neglected in the database, not only because they are time-consuming to compile, but primarily because other projects have focused on them such as the GARD Initiative (Meiri 2015), VertNet (Constable et al. 2010), and the Map of Life project (Jetz and Guralnick 2015). We have been working with these groups to link or even import their data and maps into the Reptile Database and expect to have at least a subset online by 2016.

Photos.—Currently, as of November 2015, we have photos of about 4300 species, including 3200 on our own server and another 1100 or so species represented by linked photos. Although it might be difficult to obtain photos of all species, we have alleviated this problem somewhat by providing links to Google images and other sources (the latter at the bottom of each species page). Independently, we have been working on a database of published photos, so that we can at least provide a reference where a photo can be found, even though this might be a rather obsolete method to find photos in our digital age.

One important aspect we would like to stress in the future is to provide photos of both species-specific details and **variation**. Many books and journal articles, even new species descriptions, often provide only a single photo. This is clearly not sufficient to document the diversity within a species. Given the increasing trend to split species into smaller units, variation needs to be documented better, ideally together with meta-data such as sex, age, location, etc. Similarly, we need photos of details including close-ups of **characteristic features**. Some species are so variable that it is nearly impossible to enable positive identification from a single photo. Think of Mediterranean lizards such as *Podarcis siculus* for which dozens of subspecies have been described. Even though most of them are now synonymized into only three subspecies (Podnar et al. 2005), it remains an important goal to document their bewildering diversity.

Morphology.—Many species cannot be identified from external characters. We prefer to outsource anatomical photos to other sites such as Morphbank (O’Leary and Kaufman 2015) and DigiMorph (Rowe 2015). The latter represents an effort to scan skeletons and other anatomical features by micro-CT-scanning

and this will eventually revolutionize our understanding and documentation of anatomical features. However, these efforts will also require huge data repositories of 3D images (or even 4D images if developmental stages are included). For many reptiles, especially snakes, it would also be important to database 3D images of their hemipenes but there appear to be no larger efforts yet to do so.

DNA and genomes.—More than 5000 species in the Reptile Database already have links to the NCBI Taxonomy database and thus GenBank. It will be only a matter of time until genome sequences will become available for many, eventually most, reptile species. This will be the basis for detailed molecular studies of their biology and evolution, and therefore to any other aspect of their life.

Genome data will become increasingly important with the merger of molecular biology, morphology, and evolution into “*evo-devo*” research—the attempt to understand form and function from genomes and their evolution. It is not clear yet how to database this information though. However, I am convinced that the current typological and phylogenetic species concepts will be replaced by a **genomic species concept** that will define species by their adaptive and reproductive characters as predicted from genome sequences.

Phylogenetics.—Although DNA sequences are the basis for most phylogenetic studies today, there is still no satisfactory system in place to keep track of all published phylogenies despite some serious recent efforts (Hinchliff et al. 2015). Maybe this is unnecessary as more DNA sequences will become available for reptiles, so that definitive phylogenies can be automatically created from sequences in GenBank.

Ecology.—Although we do have sporadic information about habitats, we have not collected such information in a systematic way. However, especially in conjunction with biogeographic data this would be highly useful, e.g., for macro-ecological questions. Only recently such data became available on a larger scale (e.g., Myhrvold et al. 2015). Similarly, there is not much databased information about interspecies interactions, such as diets or parasites. Such information is still highly dispersed and difficult to find.

Web technology and databasing.—Eventually, the database needs to move from a central **curator** model to a distributed system so that multiple curators can edit and update the database in parallel. In addition, we plan to add more search and display choices, e.g., that search results can display more than one field (currently only the species name).

Citizen Science.—Collaborations with citizen science projects will be increasingly important to document species distributions, their change over time, and geographic variation. These aspects cannot be covered by a limited number of professional herpetologists. We are working with **iNaturalist** and other citizen science projects and need to improve in-time exchange of information with these projects. Related to that are links to **Wikipedia** and **Wikispecies**; although Wikipedia is a great place for community-driven data annotation it does not (yet) offer the advantages of a structured database such as complex queries. Currently, only the French Wikipedia appears to have a more or less complete list of reptile species (compiled in collaboration with us) but others will most likely follow. However, the problem will be to keep all those Wikis up to date.

ID tool.—Identification of species in the field is an important goal of the Database for both experts and laymen alike. However, the problem is non-trivial. As I am writing this manuscript, a

British birder sent me a photo of a Mexican snake from Oaxaca which I forwarded to four experts to have it identified. Sure enough, I received four different IDs!

We are working to improve the ID capabilities of the database and will add searchable characters so that species can be identified more reliably. A combination of features—e.g., location, color or pattern, size, and certain morphological characters (such as keeled vs. smooth scales)—should allow a user to identify many, if not most, species. However, this will also require that we obtain many more photos with key characters including photos documenting variation.

Long-term storage.—A few months ago I submitted a grant to the US National Science Foundation (NSF) and several reviewers pointed out that the Reptile Database had no clear plan for future maintenance and availability. Yes, even one of the longest-lived biological databases such as the Reptile Database cannot make any promises about its future without funding. But given that I may formally retire only in another 20 years, there is a good chance that we will be able to keep it running for a while. On a more objective note, we have deposited past versions of the database as flat text files at **figshare.org**. That way, if the Reptile Database runs out of stamina, someone else can pick up and continue where we left off.

Epilogue and your personal wish list.—Instead of the Reptile Database I should have started a web business like Amazon (which was founded in 1994) or eBay (founded in 1995). Then I would have the funds to run a much better Reptile Database, too. Even the major biomedical databases funded by the US National Institutes of Health (NIH) are threatened by loss of funding even though they have been quite generously funded for years (Kaiser 2016). For instance, FlyBase and the Saccharomyces Genome Database have received US \$4.2 and 2.7 million in 2015, respectively. I mention these two because their number of records (in terms of genes) is similar to the number of records in the Reptile Database. In fact, they also have similar usage levels with about 51,000 and 65,000 unique users per month (the Reptile Database has about 30,000 unique users per month).

That said, let me know what you would like to see! Even without funding, we try to be as useful as possible. If you are interested in contributing some time to this project, please get in touch. There is plenty of work to do.

Acknowledgments.—Foremost, I thank the people who have implemented the database to make it useful to the world, especially Thure Etzold, Ramu Chenna, Hank Wu, Johannes Goll, Meher Preethi, and currently Jiří Hošek, webmaster of reptarium.cz. In addition to these computer experts, I want to thank the countless volunteers and sponsors who submitted data, photos, papers, or books. They are too many to list, but there is a list on the website (including more than 250 photographers). The Scientific Advisory Board and many professional herpetologists have provided guidance and advice with taxonomic decisions. Robert Hansen and Kraig Adler made valuable comments on an earlier version of this manuscript. Last but not least, the Species2000 and 4D4Life programs of the European Union have provided funding for short periods.

LITERATURE CITED

- CONSTABLE, H., R. GURALNICK, J. WIECZOREK, C. SPENCER, A. T. PETERSON, AND THE VERTNET STEERING COMMITTEE. 2010. VertNet: a new model for biodiversity data sharing. *PLoS Biol* 8(2): e1000309.
- FLEISCHMANN, R. D., M. D. ADAMS, O. WHITE, R. A. CLAYTON, E. F. KIRKNESS, A. R. KERLAVAGE, C. J. BULT, J. F. TOMB, B. A. DOUGHERTY, J. M. MERRICK,

- AND ET AL. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269:496–512.
- HINCHLIFF, C. E., S. A. SMITH, J. F. ALLMAN, J. G. BURLEIGH, R. CHAUDHARY, L. M. COGHILL, K. A. CRANDALL, J. DENG, B. T. DREW, R. GAZIS, K. GUDE, D. S. HIBBETT, L. A. KATZ, H. D. T. LAUGHINGHOUSE, E. J. MCTAVISH, P. E. MIDFORD, C. L. OWEN, R. H. REE, J. A. REES, D. E. SOLTIS, T. WILLIAMS, AND K. A. CRANSTON. 2015. Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proc. Natl. Acad. Sci. USA* 112:12764–12769.
- JETZ, W., AND R. GURALNICK. 2015. Map of Life. Available online at: <http://mol.org>.
- KAISER, J. 2016. Funding for key data resources in jeopardy. *Science* 351(6268):14.
- MEIRI, S. 2015. Global Assessment of Reptile Distributions. Available online at: <http://www.gardinitiative.org/index.html>.
- MYHRVOLD, N. P., E. BALDRIDGE, B. CHAN, D. SVAM, D. L. FREEMAN, AND S. K. M. ERNEST. 2015. An amniote life-history database to perform comparative analyses with birds, mammals, and reptiles. *Ecology* 96:3109.
- O'LEARY, M. A., AND S. G. KAUFMAN. 2015. MorphoBank 3.0: Web application for morphological phylogenetics and taxonomy. Available online at: <http://www.morphobank.org>.
- PAVEL, E., W. ZHAO, K. A. POWELL, M. WEINSTEIN, AND L. S. KIRSCHNER. 2007. Analysis of a new allele of limb deformity (*ld*) reveals tissue- and age-specific transcriptional effects of the *Ld* Global Control Region. *Int. J. Dev. Biol.* 51:273–281.
- PODNAR, M., W. MAYER, AND N. TVRTKOVIAE. 2005. Phylogeography of the Italian wall lizard, *Podarcis sicula*, as revealed by mitochondrial DNA sequences. *Mol. Ecol.* 14:575–588.
- ROWE, T., ET AL. 2015. Digimorph. Available online at: <http://digimorph.org>.
- SANGER, E., S. NICKLEN, AND A. R. COULSON. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463–5467.
- UETZ, P. 1997. Biochemical studies on limb deformity proteins in vertebrates. University of Heidelberg, Germany.
- . 2000. How many reptiles species? *Herpetol. Rev.* 31:13–15.
- . 2010. The original descriptions of reptiles. *Zootaxa* 2334:59–68.
- , AND T. ETZOLD. 1996. The EMBL/EBI Reptile Database. *Herpetol. Rev.* 27:174–175.
- , J. GOLL AND J. HALLERMANN. 2007. Die TIGR-Reptiliendatenbank. *Elaphe* 15(3):16–19.

Art in Herpetology

As space and available material permit, we will occasionally publish examples of historical as well as contemporary herpetological art. We welcome the submission of historical material under the following conditions:

- The work should be copyright-free and be provided as a high-resolution digitized file (consult Herpetological Art and Illustration Coordinator or Editor for specifications).
- Accompanying text should provide information on the subject matter, artist/illustrator, and publication source. Include references as needed to appropriately attribute information from other sources.

Contemporary artists are invited to submit their own work for publication:

- Artist will retain copyright of all submitted material.
- Accompanying text should provide information about technique, media, or other information likely to be of interest, especially to other artists. Other details of interest include the source of illustration subject (e.g., scientific name, geographic origin, etc.) or the work (e.g., book publication, etc.) for which the illustration was produced.
- Artists are welcome to include contact information (e.g., email address, website location), as well as indicating whether original art or prints are available for sale or if the artist is available for commissioned work.

Material may be submitted for review to Jackson Shedd, Herpetological Art and Illustration Coordinator at jackson_shedd@sbcglobal.net.