

## Biodiversity Committee

**Committee members (34):** E. Abreu, J. Alston, T. Androski, C. Burgin, C. Calderón-Acevedo, J. Colella, E. Craig, T. Demos, M. Dyck, J. Esselstyn, P.-H. Fabre, A. Feijó, A. Ferguson, M. Hawkins, D. Huckaby, B. Kohli, S. Maher, J. M. Martínez Cerón, V. Mathis, M. McDonough, S. Mech, A. Mychajliw, J. Nations, R. Norris, G. Oliver, O. Ornelas, B. Patterson, N. Pradhan, D. Reeder, M. E. Rodríguez-Posada, L. A. Ruedas, B. Tanis, N. Upham (Chair), J. Widness.

Cross-committee links: Conservation (Ferguson, Upham), Informatics (Maher, Tanis), Nomenclature (McDonough, Norris, Pradhan, Reeder, Ruedas), Mammal Images Library (Huckaby, Tanis), Public Education (Mech), Publications (Ruedas), Sys Collections (Upham).

### **Mission:**

The Mammal Biodiversity Committee stewards the Mammal Diversity Database (MDD), an ASM-based, readily updatable, and online database of mammal taxonomic and biodiversity information hosted at <http://mammaldiversity.org/> [beta version; official launch in July 2017]. This database aims to serve the global mammalogy community by providing the latest information on species-level and higher taxonomic changes, thereby promoting more rigorous study of mammalian biodiversity worldwide. The initial objective for this online database is to aggregate, curate, and compile new citations on species descriptions and taxonomic revisions into regular releases in comma-delimited format. Downstream goals include expanded hosting of ecological, trait, and taxonomic data, and an online forum for discussing mammalian taxonomy and systematics. By serving as both a platform and forum, this initiative aims to stimulate interest in mammals and promote the ASM's role as a leader in high quality research on mammalian biology.

### **Information items:**

- I. Activities of the ASM Biodiversity Committee from June 2019 – April 2020 included:
  - a. In June 2019, we met in person at the ASM annual meeting in Washington, DC to discuss the future goals of this committee, what to prioritize first, and how best to reach those goals. At that meeting we hosted the following programming:
    - i. Symposium—*Mammal diversity from GenBank to the RedList: Challenges and rewards of integrating museum specimens in global databases of genetics, taxonomy, and spatial biodiversity*. N. Upham and L. Ruedas, organizers. Outcome: 7 speakers, well attended, useful follow-up discussion for long-term planning of mammal taxonomic infrastructure.

# MAMMAL DIVERSITY FROM GENBANK TO THE REDLIST

2 July 2019 | ASM 100th anniversary meetings | Washington, DC

INTEGRATING MUSEUM SPECIMENS IN GLOBAL DATABASES OF GENETICS, TAXONOMY, AND SPATIAL BIODIVERSITY



**Figure 1.** Line-up of speakers and affiliations from our committee’s 2019 symposium.

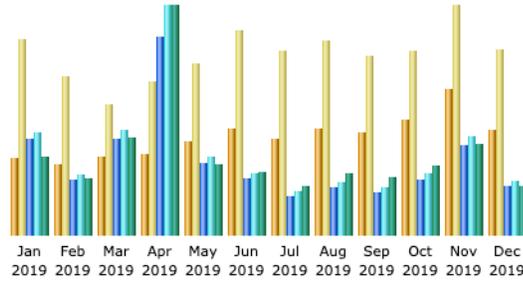
- ii. Workshop—*Hackathon for the Mammal Diversity Database*. N. Upham, J. Colella, and C. Burgin, organizers. Outcome: We had 12 participants registered from 3 countries. Much discussion of MDD goals, current infrastructure, and challenges.



**Figure 2.** Results of brain-storming for areas of short and long-term MDD planning.

- b. The 2018 *J. Mammal.* article about the Mammal Diversity Database (“[How many species of mammals are there?](#)”) has now been viewed >30,000 times and cited 150 times.
- c. From Jan – Dec 2019, the <http://mammaldiversity.org> website hosting the Mammal Diversity Database was accessed by an impressive **18,654 unique visitors**, including ~1,500 users per month (see Fig. 3). The ~4.4 million page views are likely exaggerated by API calls and bots, but the rate of 1.83 visits per visitor (34,309 total visits) indicates that most human users returned for second helpings.

**Monthly history**



Month	Unique visitors	Number of visits	Pages	Hits	Bandwidth
Jan 2019	1,222	3,090	480,819	512,857	7.46 GB
Feb 2019	1,128	2,503	277,259	303,264	5.40 GB
Mar 2019	1,248	2,062	477,444	524,116	9.28 GB
Apr 2019	1,279	2,429	989,183	1,141,810	21.70 GB
May 2019	1,482	2,710	360,336	391,440	6.63 GB
Jun 2019	1,678	3,226	283,497	310,445	5.98 GB
Jul 2019	1,523	2,907	193,293	216,230	4.64 GB
Aug 2019	1,686	3,078	235,593	263,552	5.85 GB
Sep 2019	1,613	2,830	212,696	240,765	5.46 GB
Oct 2019	1,824	2,909	278,731	309,280	6.54 GB
Nov 2019	2,302	3,628	446,527	489,673	8.57 GB
Dec 2019	1,669	2,937	243,831	269,623	4.69 GB
<b>Total</b>	<b>18,654</b>	<b>34,309</b>	<b>4,479,209</b>	<b>4,973,055</b>	<b>92.20 GB</b>

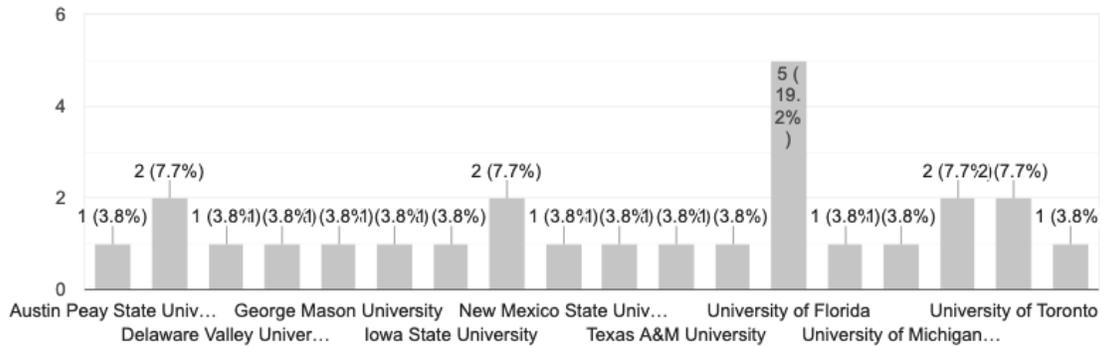
**Figure 3. Summary of usage statistics on mammaldiversity.org for 2019.** The Mammal Diversity Database was active all year in hosting the 2018 v1 taxonomy. The 18,654 unique users is likely the most reliable statistic (absent API calls and bots).

**II. Personnel updates:**

- a. In Feb 2020, Chair Upham was hired at Arizona State University as an Assistant Research Professor in the Biodiversity Knowledge Integration Center (BioKIC). As a result, his day-job responsibilities are considerably more in line with the goals of the Biodiversity Committee, and he expects to now devote more time to managing the MDD.
- b. In April 2020, the Biodiversity Committee performed a public search for Student Research Assistants to work on assembling the 2020 MDD taxonomy. The search was announced on Mammal-L and applications were accepted from April 7-13; despite the short timeline to begin work, the search yielded 27 total applicants (26 on Google Forms + 1 via email) from 18 institutions (Fig. 4).

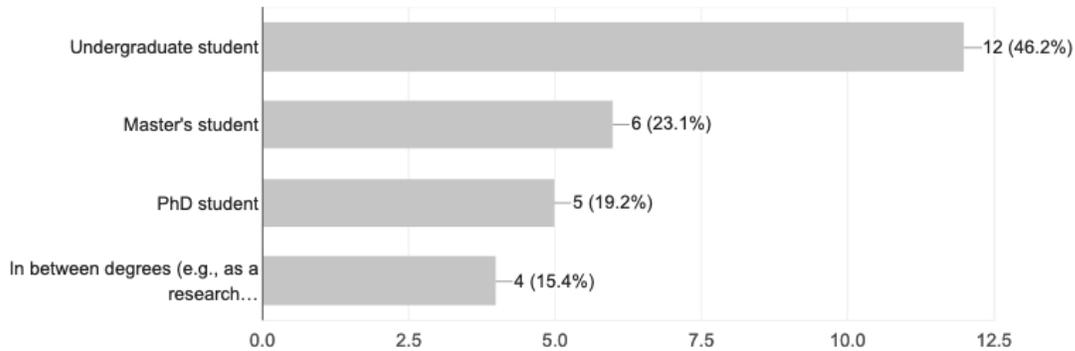
### Current University or College

26 responses



### Education level

26 responses



**Figure 4. Summary of Student Research Assistant applications for 2020.**

Undergraduate applicants were best represented, followed by M.S., Ph.D., and those students between degrees.

- c. Applications were evaluated by a subcommittee of six Biodiversity Committee members (Missy Hawkins, Jocelyn Colella, Jesse Alston, Alexis Mychajliw, Camilo Calderon, Nathan Upham) using an objective scoring rubric on Google Forms to collect responses. We came up with a ranked list of the top 10 candidates, with weighting applied for different educational levels, and prior experience with both mammal taxonomy and web databasing. We ended up hiring the following five student workers, four as Student Research Assistants and one to fill our vacant Web Programmer role:

Name	School	Education	Position
Connor Burgin	Columbus State Univ	Master's	Student Research Assistant
Jane Widness	Yale Univ	PhD	Student Research Assistant
Camilla Parker	Columbus State Univ	Undergraduate	Student Research Assistant

Madeleine Becker	George Mason Univ	BS; pre-PhD	Student Research Assistant
Schuyler Liphardt	Univ of New Mexico	PhD	Web Programmer

- d. All started work on 27 April 2020, with work plans targeted at ~5 hrs/wk (10 hrs/wk maximum) until our budget runs out in September or October 2020. We have a group Slack that is already quite active at dividing tasks and helping each other.

### III. Transitioning the MDD to a sustainable ‘Open Science’ model:

- a. Beginning in November 2019, we sought out the professional advice of Jorrit Poelen, web programmer for the Global Biotic Interactions database (GloBI), about how to transition the MDD to a data model that will be sustainable for the ASM in the long term. To fund this consultation (limited to 8 hours at \$100/hour), Chair Upham was able to apply funds from the BioKIC initiative at Arizona State University.
- b. The full results of this consultation are available on request, but the main summary recommendation includes the following points, extracted verbatim from the report:

---

#### i. Recommended Publication Workflow

##### Step 1. Curate

Curator continues to use familiar tools like Excel and Google Sheets (and perhaps git) to make updates to MDD.

##### Step 2. Review

After periodic review, a stable version of MDD is selected for publication.

##### Step 3. Publish

Curator publishes the reviewed master copy along with machine readable exports (e.g., csv/tsv/json) to a trusted data publication platform.

##### Step 4. Website Update

The curator updates the website to use a specific published MDD export. After this update, the MDD website is automatically refreshed to include the selected MDD version.

Note that, depending on the desires of the MDD curator, steps 1-4 can be made for small, frequent, and incremental changes, but can also involve large infrequent batch updates. Also notice that the website update can be seen as an integral part of the publication process. Rather than seeing a website as a dynamic and every changing entity, I’d like to introduce the idea of a website as an interactive version of a static, well-defined version of MDD. More like a web-based data paper than a dynamic editable website that changes uncontrollably.”

#### ii. Recommended Migration Strategy

Rather than starting from scratch, I imagine a gradual transition from the current website toward a more fitting solution. This can be done as follows.

- First, the existing repository containing the website code, data and documentation is forked and archived.
- Second, the current website is scraped and page templates are extracted, with a focus on the species pages and associated permalinks.

- Third, a version of the MDD master copy is exported to csv/tsv and published as a data paper.
- Fourth, a prototype website is built using a static site generator (e.g., Jekyll) to automatically populate static html pages for each species in the selected csv/tsv version of MDD (for example, see <https://github.com/open-traits-network/open-traits-network.github.io> ). Static pages (e.g., about page) are included as is.
- Fifth, when step four produces a minimal viable website (e.g., a static website with a big table with all names, along with individual species pages for each), the DNS settings at Tucows (the domain name registrar) are updated to point to the new website.
- Sixth, the old site continues to for another couple of weeks/ month as a backup.”

### iii. Recommended Feature Modifications

F1. Download Taxonomy:

Keep, but change to a static download link to published MDD data paper version.

F2. Hierarchical Name Search:

Keep, but rather than relying on a (complex) graph engine to do name searches, include a simple searchable table with text filters. On clicking a row, a species page opens.

F3. Species Pages:

Keep as is.

F4. Permalinks:

Keep as is.

F5. Citation generation:

Keep, and include a DOI to the data publication of the specific MDD version.

F6. Statistics pages:

Replace with a version generated from data included in the MDD data publication version.

F7. Taxonomy page:

Replace with a version generated from the data included in the MDD data publication version.

F8. Forum:

Drop feature. If a forum is needed, suggest to re-use existing platforms (e.g., Google groups, Github issues etc.)

F9. Random page:

Drop feature or include as “nice to have” - the random page feature can be implemented using some javascript if needed.

F10. Web API:

Drop feature. The MDD data products included in the MDD data paper are sufficiently small (6 MB, ~7k lines) that commonly used compute platforms like R or python to work with the entire directly dataset. The cost of maintaining a web server to host custom API request outweigh the benefits in my opinion.”

### iv. Recommended Technology Modifications

To accommodate the skills of the MDD community and to reduce maintenance, I recommend to no longer use a mysql database, graph database, php programming language, sql queries, Cypher queries etc to run a public facing MDD website.

Instead, as hinted above, I would recommend using a commonly used static website generator like Jekyll (<https://jekyllrb.com/>) or Hugo (<https://gohugo.io/>). These site generators use html, css, javascript in combination with data-driven page templates to automatically generate websites. Dynamic functionality can be introduced using javascript. The generated website can then [be] served using Github Pages, or by existing, self-hosted, nginx, apache or caddy web servers. Whether MDD chooses to hire one of the many companies to (relatively) cheaply develop such a website, or whether MDD contributors chose to build their own, I would expect the cost and skills required will be far less compared to the custom architecture that is currently in place.

If more advanced user interface (taxonomic) features are needed to satisfy MDD users, I would recommend to use an existing solution like TaxonWorks or similar. In my experience, building custom solutions requires professional software engineers, is expensive, has high risk of failure and often distracts from the core activities of a project. Similarly, if more advanced data integration methods are needed (e.g., automatically linking MDD to a related cyberinfrastructure or external ontology), adoption of readily available tools and existing infrastructures are recommended over developing custom data integration tools.

---

#### IV. New bare-bones website to host the MDD taxonomy:

- a. In response to the recommendations of Jorrit Poelen, we have created a new prototype for an MDD website that only contains the bare essential information, which has the critical benefit of being *updatable directly by students and volunteers of the ASM Biodiversity Committee*. This new site is temporarily hosted on GitHub at <https://mammaldiversity.github.io/> with the simple backend visible at <https://github.com/mammaldiversity/mammaldiversity.github.io>. Switching to this framework helps to free us of the previous obstacles of closed-door encryption posed by the PHP backend (as here <https://github.com/tigerhawkvok/asm-mammal-database>). We will transition this bare-bones site to mammaldiversity.org when it is more advanced.
- b. Our plan is to have our student Web Programmer (currently Schuyler Liphardt, UNM) build features using the Jekyll, CSS, and Javascript, in this open source framework. We will be able to add per-species pages, direct searching, subset download, and other features using common tools of the Jekyll framework.
- c. The highly simplified web interface recognizes that the primary users of our mammalian taxonomy are (i) computer literate biology researchers (professional or amateur) and (ii) biodiversity aggregation services like iNaturalist that seek up-to-date taxonomic information for global mammals. Thus, *our primary goal is to produce tabular CSV outputs of the MDD taxonomy at regular intervals*, rather than any user-friendly visualizations on the front end (although those add-ons could come at a later date).

## V. Publishing v2.0 of the MDD taxonomy:

- a. Following the recommendations of Jorrit Poelen, we will soon be submitting version 2.0 of the MDD taxonomy for publication on the open-source platform Zenodo. This version of the taxonomy is dated 23 March 2019, and is now being cleaned for consistency across fields. By publishing on Zenodo, this snapshot of the taxonomy will have a DOI and thus be citeable/reproducible for use by others.
- b. We plan to release additional snapshot taxonomies at tighter time intervals as well, basically to have a Zenodo version of the live MDD taxonomy on [mammaldiversity.org](http://mammaldiversity.org).

## VI. Work on the 2020 MDD taxonomy (v3.0):

- a. Work with the four new Student Research Assistants will focus on unifying the cumulative taxonomy from the Handbook of the Mammals of the World series (Vols 1-9) with our independent understanding from the MDD literature surveys. Some aspects of the HMW taxonomy, like the ~300 additional species of ungulates recognized by Groves & Grubb (2011), differ from our current MDD taxonomy, so will require careful reconciliation.

**Table 1.** Taxonomies compared among previous compendia: Mammal Species of the World (MSW), International Union for the Conservation of Nature (IUCN), and the Mammal Diversity Database (MDD). The MDD v3 is being generated relative to Handbook of the Mammals of the World (HMW). Note that the apparent decrease in MDD v3 species total is an artifact of the fact that we have not yet vetted all the HMW species relative to our existing database.

Taxa	MSW3 2005	IUCN 2008	MDD v1 2018	MDD v2 2019	MDD v3 (in progress; merge w/ HMW)
Species					
<i>Total</i>	5,416	5,513	6,495	6,526	6,341
<i>Extinct</i>	1	77	96	100	100
<i>Living</i>	5,415	5,436	6,399	6,426	6,241
<i>Domestic</i>	0	0	16	17	6,224
<i>Flagged</i>	0	0	212	212	?
<i>Wild &amp; valid</i>	5,415	5,436	6,171	6,197	?
Genera	1,230	1,226	1,314	1,322	1,322
Families	153	149	167	167	167
Orders	29	24	27	27	27

## VII. Continued goals for work Mammal Diversity Database (2020 and beyond):

- a. **Synonyms.** Our initial matching yielded synonym data for 3,113 species (1 or more junior synonym equivalency with authority & year). We also have one or more listed subspecific epithet for 1,299 species. Merging and reconciling the MSW3 data, along with data from other biodiversity aggregators (e.g., ITIS) is a priority for our group.

- b. **Type specimens.** We recognize a key priority of reconciling these recent taxonomic changes in mammals with the *museum location of type specimens* that pertain to these recognized mammalian species. To date, we have manually added holotype museum numbers to 247 species. We also began parsing VertNet data and found 2,313 extant holotypes and 6,257 extant other types (e.g., para- and topotypes) in VertNet collections. Our aim is to focus on primary types (holotypes, lectotypes, syntypes, cotypes, and neotypes), and exclude secondary types like paratypes and topotypes.
- c. **Authority full citation & year.** This content priority has become clear and will be especially useful for mammalogists without strong library resources. Our goal here is to provide the DOI and links to digitized open-source materials when available.
- d. **Improved organization of volunteers.** In 2019, we assembled an initial list of per-clade ASM members and non-member specialists to help vet and provide editorial curation as the MDD taxonomy continues to evolve. See Table 2 in our 2019 Annual Report for that list of people spread across 25 taxon-specific subcommittees. However, we still lack the organizational infrastructure needed to properly use the efforts of those volunteers, so we have yet to formally reach out to them as a group. We still maintain the goals of these subcommittees as follows:
- i. Establish *group* efforts such that no single individual has full authority over the advising decisions for a particular taxon.
  - ii. In keeping with the objectivity goals of the Mammal Diversity Database, only published taxonomic opinions are allowed to be considered. The taxonomic subcommittees will contain multiple voices spanning diverse perspectives toward the goal of judging which published works will be accepted into the MDD active listing.
  - iii. Thus, these subcommittees will be primarily *objective*, but also have limited *one-directional* subjective control over the accepted species in their group. While species additions must be derived from the peer-reviewed literature (including compendia), species can be excluded or ‘flagged’ for additional study under the judgement of supporting evidence by group members.
  - iv. Setting a 6-month schedule for returning an updated listing of revisions per clade;
  - v. Incorporating these per-clade updates into a full MDD version release on a 10-month basis (e.g., July to April would be useful given the June ASM meetings); and,
  - vi. Giving *citation credit* to these taxonomic subcommittees, with author order determined by the individual subcommittees.
  - vii. We emphasize that “top down” data integration will still be needed to make use of these “bottom up” volunteers, which is why we continue to propose funding for student assistants (see below).
- e. **Harmonization of the MDD web interface with existing ASM initiatives—namely Mammal Images Library and *Mammalian Species*.** Discussions with David Huckaby (Chair of the MIL Committee) have continued, particularly in the form of his help with curating the species list. For example, Huckaby has helped us flag a series of inconsistencies with the database, which we’ve organized into 67 separate issues (see [Google Sheet here](#)). However, no additional steps were taken this year regarding a formal union of the MDD taxonomy with the taxonomy underlying the MIL. Informal

discussions were also had with Meredith Hamilton (*Mammalian Species* EIC) about future efforts to link and index the MS accounts via the MDD taxonomy, presenting opportunities for modernizing old species accounts.

- f. **Other content-based priorities:** These goals remain for longer-term implementation:
  - a. Hyperlinks to new species citations (e.g., in *J. Mammalogy*, *Mammal. Biol.*, *Zootaxa*)
  - b. Expansion of per-species content to include detailed natural history, ecological, and geographic information (e.g., collaboration with Animal Diversity Web).
  - c. Downstream: per-species info on phylogenetic position (VertLife tree of mammals), geo-distributions by region (continent, country, state in USA), and NCBI/IUCN links.

### VIII. Plans for non-ASM funding support:

- a. One of the main recommendations derived from our 2019 ASM Workshop on “Hacking the MDD” was the need to pursue a “burst” of outside funding from NSF or similar agencies in order to push the content of the MDD to a higher level. The ASM could then continue to support MDD with low-level sustainable funding. As a result, Chair Upham has been in regular conversations with DeeAnn Reeder of *Mammal Species of the World* (MSW) and Nancy Simmons of *Batnames.org* with respect to applying for major grant funds. These discussions have been productive, and joined with taxonomic name-to-concept alignment ideas that Upham and his ASU colleagues (Nico Franz, Beckett Sterner) have been independently developing. A grant proposal is now in progress that if successful would unify the MDD with *Batnames.org* and MSW (volume 3 was the last edition of MSW, but the delayed vol. 4 will now likely be joined with our group ASM-based effort). Thus, our plan is to present mammals as a test case for the potential benefits of enhanced taxonomic concept mapping, while also centralizing an updated global taxonomy for use in biodiversity portals like GBIF, CoL+, and iNaturalist.

### IX. COVID-19 involvement of the Biodiversity Committee:

- a. Beginning in March 2020, Chair Upham began representing the MDD in the [CETAF-DiSSCo Taskforce](#) formed in response to the global COVID-19 epidemic. To date, ~60 scientists from Europe, the United States, and Brazil are involved in this taskforce aiming to centralize the most relevant and up-to-date museum-based biodiversity information for the biomedical research community. This has involved multiple weekly Zoom meeting and a flurry of activity on Slack, and already led to several spin-off grants and synthesis efforts to help scientists respond to this zoonotic outbreak.

---

## RECAP OF BUDGET 2020

---

### 1. Website construction and maintenance:

- Website stability, updates to content and interface, feature modifications
  - o 2020 approved (~60 hrs @ \$20/hr) ..... \$1,200.00
  - o Spent to date (April 2020, set to soon start paying Liphardt) ..... \$0.00

### 2. Hiring of graduate student research assistant(s):

- We added four student assistants employed at ~5/hrs per week to facilitate integration of updated taxonomic data in this ASM initiative.

- 2020 approved (~5 hrs/w @ \$15/hr, 4 graduate students, 18 weeks) ..... \$5,400.00
- Spent to date (April 2020, set to soon start paying Burgin, Widness, Parker, and Becker) ..... \$0.00

TOTAL BUDGET APPROVED (2020) ..... \$6,600.00  
 Spent to date (April 2020) ..... \$0.00  
 Projected spending (by October 2020) ..... \$6,600.00

## **PROPOSED BUDGET 2021**

### **1. Website construction and maintenance:**

- Website stability, updates to content and interface, feature modifications
  - 2021 proposed (~60 hrs @ \$20/hr) ..... \$1,200.00

### **2. Hiring of graduate student research assistant(s):**

- We plan to continue needing 2-4 student assistants employed at ~5/hrs per week. These student assistants will facilitate integration of updated taxonomic data in this ASM initiative. Specifically, student assistants will perform functions of:
  - (i) secondary processing of per-clade data sheets that are returned;
  - (ii) emailing and coordination of per-clade volunteer tasks (“bottom-up” data);
  - (iii) “top-down” data gathering efforts in terms of parsing, matching, and curating global mammal databases (e.g., VertNet) relative the data already in the MDD.

- ⇒ Therefore, we request funds for continued efforts in curating the taxonomy database:
- 2021 proposed (~5 hrs/w @ \$15/hr, 2-4 students, 18-36 weeks) ..... \$5,400.00

**TOTAL BUDGET REQUESTED (2021) ..... \$6,600.00**

## **SUMMARY**

Overall, we highlight that 2020 will continue to be a year of change for the Mammal Diversity Database, as it is showing itself to be for much of society with the COVID-19 outbreak. However, this pandemic interestingly also highlights the overarching need for updated taxonomies about the mammalian hosts of zoonotic viruses such as SARS-CoV-2. The ASM Biodiversity Committee is well positioned to contribute to the robustness of this information, and should continue to do so. Mammals are now in closer parity to other tetrapod groups where taxonomic databases have existed for over a decade (e.g., <http://amphibiaweb.org>, <http://reptile-database.org/>); however, keeping track of “How many species of mammals are there?” is only growing in importance in our changing world.

**Respectfully submitted,**  
 Nathan S. Upham, Chair  
 ([nathan.upham@asu.edu](mailto:nathan.upham@asu.edu))