

1 **The Vertebrate Taxonomy Ontology: A framework for reasoning across**
2 **model organism and species phenotypes**

3
4 Peter E. Midford*^{1,2,^} (peter.midford@gmail.com)

5 T. Alex Dececchi*³ (alex.dececchi@usd.edu)

6 James P. Balhoff^{2,4} (balhoff@nescent.org)

7 Wasila M. Dahdul^{2,3} (wasila.dahdul@usd.edu)

8 Nizar Ibrahim⁵ (nibrahim@uchicago.edu)

9 Hilmar Lapp² (hlapp@nescent.org)

10 John G. Lundberg⁶ (lundberg@ansp.org)

11 Paula M. Mabee³ (pmabee@usd.edu)

12 Paul C. Sereno⁵ (dinosaur@uchicago.edu)

13 Monte Westerfield⁷ (monte@uoneuro.uoregon.edu)

14 Todd J. Vision^{2,4} (tjv@bio.unc.edu)

15 David C. Blackburn^{8,9} (dblackburn@calacademy.org)

16

17 *co-first authors

18 ¹Department of Ecology and Evolutionary Biology, University of Kansas

19 ²National Evolutionary Synthesis Center, Durham, North Carolina

20 ³Department of Biology, University of South Dakota

21 ⁴Department of Biology, University of North Carolina

22 ⁵Department of Organismal Biology and Anatomy, University of Chicago

23 ⁶Academy of Natural Sciences, Philadelphia

24 ⁷Institute of Neuroscience, University of Oregon

25 ⁸Department of Vertebrate Zoology and Anthropology, California Academy of

26 Sciences

27 ⁹Corresponding author

28 [^]current address: 6003 C Willow Oaks Drive, Richmond VA 23225

29 **Abstract**

30

31 **Background:** A hierarchical taxonomy of organisms is a prerequisite for
32 semantic integration of biodiversity data. Ideally, there would be a single,
33 expansive, authoritative taxonomy that includes extinct and extant taxa,
34 information on synonyms and common names, and monophyletic supraspecific
35 taxa that reflect our current understanding of phylogenetic relationships.

36 **Description:** As a step towards development of such a resource, and to enable
37 large-scale integration of phenotypic data across the vertebrates, we created the
38 Vertebrate Taxonomy Ontology (VTO), a semantically defined taxonomic
39 resource derived from the integration of existing taxonomic compilations, and
40 freely distributed under a Creative Commons Zero (CC0) public domain waiver.
41 The VTO includes both extant and extinct vertebrates and currently contains
42 106,927 taxonomic terms, 23 taxonomic ranks, 104,506 synonyms, and 162,132
43 taxonomic cross-references. Key challenges in constructing the VTO included (1)
44 extracting and merging names, synonyms, and identifiers from heterogeneous
45 sources; (2) replacing subgroups with more authoritative local taxonomies; and
46 (3) automating this process as much as possible to accommodate updates in
47 source taxonomies.

48 **Conclusions:** The VTO is the primary source of taxonomic information used by
49 the Phenoscope Knowledgebase (<http://phenoscope.org/>), which integrates genetic
50 and evolutionary phenotype data across both model and nonmodel vertebrates.

51 The VTO is useful for crudely inferring phenotypic changes on the vertebrate tree

52 of life, which enables queries for candidate genes for different episodes in
53 vertebrate evolution.

54 **Keywords:** data integration, evolutionary biology, paleontology, taxonomic rank

55 **Background**

56 Integration of data about organisms almost always requires a taxonomic
57 framework. The Phenoscope project aims to integrate morphological and genetic
58 data, incorporating information from both model organism databases and data
59 from the literature on non-model organisms, including extinct taxa. Phenoscope
60 requires a semantically defined taxonomic resource that includes extant and
61 extinct species, can be used to recognize both valid names and synonyms as they
62 are used by different authors, and that is constructed, to the greatest extent
63 possible, in line with current phylogenetic understanding.

64 Phenoscope's initial focus is on data from vertebrates. While vertebrates
65 comprise only a small fraction of all biodiversity, the group is sufficiently large
66 that the relevant taxonomic information is distributed among several different
67 resources. We combined information from multiple sources to build the
68 Vertebrate Taxonomy Ontology (VTO; <http://purl.obolibrary.org/obo/vto.owl>).
69 As of May 2013, the VTO contained 106,927 terms annotated with 104,506
70 synonyms, 162,132 cross-references, and 23 ranks.

71 Here we discuss the three main challenges encountered while building the
72 VTO and our approaches to solving them: (1) extracting and merging names,
73 synonyms, and identifiers from different and highly heterogeneous sources; (2)
74 grafting together phylogenetically-based hierarchies and higher-level taxonomic
75 node-based names from sources with conflicting information; and (3) automating
76 this process so that it is repeatable and can accommodate updates in the source
77 taxonomies.

78

79 **Content and construction**

80

81 **Selection of sources**

82 The VTO does not seek to publish new taxonomic names, and thus we only
83 import names from existing source taxonomies. Sources have been selected based
84 on their coverage, authority, and availability. All the current sources are
85 electronically available and either have compatible terms of use or have been
86 made available for use and redistribution with permission.

87 VTO is currently built upon two resources with broad taxonomic coverage of
88 the vertebrates, and two resources with richer coverage of particular subgroups.
89 The National Center for Biotechnology Information (NCBI) taxonomy provides
90 the hierarchical backbone for extant taxa. Because NCBI taxonomy largely
91 includes species associated with archived genetic data, it excludes many extant
92 and nearly all extinct taxa (Figure 1). To complement this, we also incorporated
93 taxonomic information across the vertebrates from the Paleobiology Database
94 (PaleoDB). The Teleost Taxonomy Ontology (TTO) and AmphibiaWeb (AWeb)
95 were incorporated to provide a more authoritative hierarchy and a richer set of
96 names for specific taxonomic groups.

97

98 ***National Center for Biotechnology Information (NCBI) Taxonomy***

99 The NCBI taxonomy is a curated consensus view of taxonomic relationships [1].

100 It offers broad coverage with a coarse hierarchy and provides valuable linkages to

101 molecular data. As of May 2013, the NCBI contained 65,747 taxonomic names
102 for vertebrates.

103

104 ***Teleost Taxonomy Ontology (TTO)***

105 The TTO is a taxonomy ontology based on the Catalog of Fishes (CoF, [2]). It is
106 subsequently modified by contributions from taxonomic experts as part of the
107 Phenoscape project [3]; to date, 754 terms have been added (mostly genera,
108 species, and some extinct taxa) that are not present in CoF. The TTO has been
109 continually updated to reflect both changes in CoF and to incorporate additional
110 taxa encountered during the process of curation in Phenoscape [4]. These include
111 taxa known only from fossils, subjective names, misspellings, and names used as
112 placeholders in manuscripts or publications before a formal taxonomic treatment
113 is available (e.g., “*Danio* aff. *Dangila* (Fang 2003) [5]” or “*Agoniates* sp.
114 (Toledo-Piza 2000) [6]”). Specimens that are not given a species designation are
115 assigned to a nonspecific taxon in TTO that includes a citation to the curated
116 publication (e.g., “*Eigenmannia* sp. (Fink and Fink 1981) ”). This practice enables
117 reasoning to be applied to specimens described in the literature that are excluded
118 from traditional taxonomies such as CoF due to uncertainty in species affinities at
119 the time of publication. A tool called ‘TTOUpdate’
120 (http://phenoscape.org/wiki/TTOUpdate_tool) was developed by the Phenoscape
121 project to update TTO automatically with each new release of the CoF. As of May
122 2013, the TTO contained 38,640 taxonomic terms for valid species and higher
123 taxa and 60,028 synonyms (taxonomic synonyms and vernacular names). VTO

124 does not retain taxonomic information from TTO for clades outside
125 Actinopterygii or Chondrichthyes, relying instead on NCBI for the hierarchy for
126 this small number of taxa (164 taxa).

127

128 *AmphibiaWeb (AWeb)*

129 For living amphibians, we have chosen to graft the AmphibiaWeb taxonomy
130 (obtained from http://amphibiaweb.org/amphib_names.txt) onto the NCBI
131 backbone. The hierarchy and taxon sampling of the AmphibiaWeb taxonomy is
132 more expansive than that of NCBI for this clade which contains fewer than 5,800
133 named taxa. It is updated frequently, available online, and widely used. As of
134 May 2013, AmphibiaWeb contained 7,854 taxonomic names, all of which are
135 incorporated into VTO.

136

137 *Paleobiology Database (PaleoDB)*

138 The Paleobiology Database (PaleoDB, <http://paleodb.org>) is an expert curated
139 collection that attempts to cover the entirety of the fossil record including
140 taxonomy, specimen locations, and stratigraphic distributions [7]. Its primary use
141 is as a repository of occurrence data to support large-scale paleobiogeographic
142 analyses; thus, in addition to named extinct and some extant biological species, it
143 includes data on trace fossils and unidentified body fossils. The PaleoDB provides
144 not only a listing of all currently published taxon names, but also other identifiers
145 (such as long-obsolete synonyms and trace fossil taxa) that were not included in
146 the VTO because they would not be encountered in the relevant literature. We did

147 not include taxa that are invalid, synonyms, or difficult to place because they
148 lacked a parent classification. As of May 2013, we have incorporated 35,937 of
149 the 39,180 vertebrate taxa (including trace fossils) of all ranks in PaleoDB, 28,451
150 of which are extinct.

151

152 **Constructing the VTO**

153 We have followed the basic principle of phylogenetic taxonomy that supraspecific
154 taxon names should, whenever possible, represent monophyletic groups. The
155 construction process of the VTO is initiated by importing the NCBI hierarchy. We
156 developed a Taxonomy Ontology Tool ([https://github.com/NESCent/Taxonomy-](https://github.com/NESCent/Taxonomy-Ontology-Tool)
157 [Ontology-Tool](https://github.com/NESCent/Taxonomy-Ontology-Tool)) to graft specialized taxonomies in place of nodes in this hierarchy
158 while merging lists of synonyms from multiple sources (based on matches
159 between primary names). The TTO is used to replace the NCBI taxonomy under
160 the nodes ‘Actinopterygii’ and ‘Chondrichthyes’, and AmphibiaWeb to replace
161 the node ‘Amphibia’ and its descendants (Figure 2). This results in the portion of
162 the VTO relevant to extant taxa and creates a framework within which to add
163 extinct taxa.

164 The PaleoDB taxonomy must be grafted at many different nodes, and there is
165 considerable taxonomic uncertainty for some fossil taxa. Thus, the procedure for
166 incorporating PaleoDB involves more expert interaction than with either the TTO
167 or AmphibiaWeb taxonomy. PaleoDB subtrees are grafted at the lowest ranking
168 node inclusive of that subtree within the proto-VTO. Taxa marked as “disused” in
169 PaleoDB are then marked in VTO as obsoleted, and any non-deleted children of

170 disused taxa are attached at the root. Names associated with non-diagnostic
171 material for biological species identification (e.g., track fossils, eggshells, or
172 scales) are also not included. Based on published taxonomic treatments, we have
173 augmented parts of the hierarchy by moving or merging a small number (~50) of
174 nodes. These changes, along with the rationale and the reference, were recorded
175 in the “comment” annotation section for each modified VTO term.

176

177 **Modeling of taxa and rank**

178 We provide information about taxonomic rank for those taxon terms for which a
179 rank has been provided in one of the source taxonomies. However, terms need not
180 have a rank and we note that rank-free taxonomies are becoming common in the
181 literature. To annotate taxon terms with their taxonomic rank, we constructed a
182 new vocabulary of taxonomic rank terms, the Taxon Rank vocabulary
183 (<http://purl.obolibrary.org/obo/taxrank.owl>). The vocabulary consolidates the rank
184 terms used by the NCBI taxonomy with those proposed in the Biodiversity
185 Information Standards (TDWG) TaxonRank vocabulary
186 (<http://rs.tdwg.org/ontology/voc/TaxonRank>). The resulting vocabulary contains
187 59 terms with links back to the corresponding terms in the source vocabularies.
188 We maintain this vocabulary separately from VTO to promote reuse by other
189 projects.

190 The VTO follows the same modeling pattern used to render the NCBI
191 taxonomy in OBO format and convert from OBO to OWL
192 (<http://purl.obolibrary.org/obo/ncbitaxon.owl>). In this pattern, each taxon is

193 modeled as an ontological class, and ranks are assigned to taxon classes using the
194 'has_rank' annotation property declared within the Taxon Rank vocabulary.

195

196 **Synonyms**

197 Taxonomic names can undergo multiple status revisions. As a consequence,
198 taxonomic names encountered in legacy literature may not be those in current use.

199 Further, at any given time, multiple authors may use different scientific names for

200 designating the same biological species; two scientific names of the American

201 Bullfrog, for example, are in common use today (i.e., *Rana catesbeiana* and

202 *Lithobates catesbeianus*). Supporting integration of species-related annotations

203 requires inclusion of all synonyms encountered in the literature irrespective of the

204 official taxonomic status of that name (e.g., junior synonym, spelling variant).

205 Names are excluded from the VTO when they are unavailable based on criteria in

206 the International Code of Zoological Nomenclature [8].

207 Additionally, failure to find a match for a taxon term in the literature can be

208 due to misspelling or use of common (vernacular) names. Common names are

209 valuable for making the data organized by the VTO more readily accessible to

210 non-expert users. As each source taxonomy has its own mechanism for including

211 common or vernacular names, each of these sets have been merged into the VTO.

212 For instance, in the TTO, approximately 14,400 English common names were

213 generously provided by Fishbase [9]. We take advantage of the author-defined

214 'type' tags in OBO for synonyms to define tags that distinguish the different kinds

215 of synonyms such as 'COMMONNAME' and 'MISPELLING'. The VTO

216 includes an additional annotation property 'is_extinct' to indicate taxa that are
217 known only from fossil evidence or if such designation is present in a source
218 ontology.

219 We have followed the principle that supraspecific taxon names should,
220 whenever possible, represent monophyletic groups. We have made and recorded
221 additional adjustments to the resulting hierarchy (especially coming from
222 PaleoDB) to ensure that the VTO is consistent with the phylogenetic hierarchy as
223 established by current research.

224

225 **Discussion**

226 Taxonomic ontologies are required for semantic integration of biodiversity data.
227 In the case of Phenoscope, such integration allows us to link naturally occurring
228 phenotypes among diverse taxa to phenotypes resulting from genetic
229 manipulations in model organisms [3]. Curation of phenotypes in Phenoscope
230 entails translating phenotype descriptions into the Entity-Quality (EQ) formalism,
231 and assigning these EQ descriptions to appropriate taxa (at any rank) [4]. This
232 application motivated us to develop a single taxonomic ontology for vertebrates,
233 the VTO.

234 The VTO is built as a simple hierarchy of ontology classes, which allows
235 straightforward data aggregation via subsumption reasoning. For example, data
236 referencing the ontology classes 'Rodentia' or 'Primates' should be returned from
237 a query using 'Mammalia'. We note, however, that linking phenotypic data to

238 ontology classes can introduce reasoning inconsistencies when the phylogenetic
239 history of the features involves character reversal or evolutionary loss [10].

240 While treating taxa as classes closely follows the traditional view of taxonomy,
241 it is not the only way in which taxa may be modeled. An alternative approach
242 would be to view taxonomic entities as historical individuals (e.g., Ghiselin 1974
243 [11]). Others have explored ontological models that attempt to capture the
244 complex interplay between evolutionary relationships and the practice of
245 taxonomic classification [12,13]. This approach may better represent features that
246 change over evolutionary time. For future work, we intend to make a fuller
247 comparison of the consequences of these two approaches.

248

249 **Use in the Phenoscape Knowledgebase**

250 The VTO was developed to be an integral part of the second-generation
251 Phenoscape Knowledgebase (KB) that is currently under development. Some of
252 the user interface functionality can be seen in prototype form in the first-
253 generation Phenoscape KB (<http://kb.phenoscape.org>), which uses the TTO and is
254 limited to data from fishes. There is a display page for each taxon that includes its
255 immediate taxonomic parents and children, as well as synonyms, extinction status,
256 and links to other source data for that taxon (e.g., Fishbase, Wikipedia, etc.).
257 Taxon display pages also include a browsable taxonomy tree. Taxa are included
258 in the results returned by queries for phenotypes (e.g., ‘all species with
259 phenotypes in which the basihyal is absent and the pectoral fin is triangular’).
260 Taxa can also be used to scope queries (e.g., ‘all Cyprinidae with triangular fins’)

261 and appear as elements in faceted queries. The Phenoscape KB additionally
262 displays summary phenotype statistics for taxa such as the degree of annotation
263 coverage and phenotypic variation on a simplified taxonomic tree.

264

265 **Maintenance and Revision**

266 In keeping with the practice recommended for OBO Foundry ontologies [14], it is
267 possible to modify the structure and taxonomic scope of the VTO by curator
268 requests submitted through the term request tracker
269 (<http://purl.obolibrary.org/obo/vto/tracker>), such as for correction of misspellings
270 or hierarchy errors. If a curator needs to add taxa to or suggest taxonomic
271 rearrangements of the VTO, they are encouraged to contribute this information
272 directly to the source taxonomies following their prescribed curation
273 methodologies. These will then be incorporated back into the VTO when it is
274 updated. Questions and discussions pertaining to the VTO can be directed to obo-
275 taxonomy listserv (<https://lists.sourceforge.net/lists/listinfo/obo-taxonomy>).

276

277

278 **Opportunities for inclusion of additional taxonomic sources**

279 There are additional sources for vertebrate taxonomic information that have not
280 been incorporated in the version of the VTO described here, but would be of
281 value to add in the future. For instance the Reptile Database ([http://www.reptile-
282 database.org/](http://www.reptile-
282 database.org/)) and the International Ornithologists' Union Bird List
283 (www.worldbirdnames.org) would provide taxa and synonyms not included

284 among the current sources as well as greater resolution to the hierarchy. The
285 IUCN Red List of Threatened Species (<http://www.iucnredlist.org/>) would
286 provide information about conservation status and potentially widen the
287 application of the ontology. We welcome inquiries from parties interested in
288 integrating these or additional taxonomic resources.

289

290 **Conclusions**

291 To fill the need for a single taxonomic ontology including both modern and
292 ancient vertebrate taxa, we have developed the Vertebrate Taxonomy Ontology
293 (VTO) by merging taxonomic information from a variety of expert sources. The
294 integration pipeline we have developed is repeatable and thus capable of
295 incorporating both updates from the source taxonomies and additional sources.
296 Although the development to date has been guided by the requirements of the
297 Phenoscope project, it is our hope that the VTO will be useful for integration of
298 diverse forms of data from vertebrates, and serve as a model for the development
299 of taxonomy ontologies in other groups of organisms.

300

301 **Availability and requirements**

302 The Vertebrate Taxonomy Ontology (VTO) and the vocabulary of Taxonomic
303 Ranks that it references are available via their permanent URLs under a Creative
304 Commons Zero (CC0) public domain waiver: VTO, in OWL format,
305 <http://purl.obolibrary.org/obo/vto.owl>, and in OBO format,
306 <http://purl.obolibrary.org/obo/vto.obo>; TAXRANK,

307 <http://purl.obolibrary.org/obo/taxrank.owl>. The VTO is a large ontology, and thus
308 viewing in a desktop OWL editor such as Protégé may require allocating
309 sufficient memory (2 GB at present). The VTO can also be browsed at the NCBO
310 BioPortal (<http://bioportal.bioontology.org/ontologies/50317/>).

311 The open-source software used to generate the VTO is available under the
312 MIT license at GitHub (<https://github.com/NESCent/Taxonomy-Ontology-Tool>),
313 and we welcome further development of this resource by the wider community.

314

315 **Competing interests**

316 The authors declare that they have no competing interests.

317

318 **Author's contributions**

319 Wrote the paper: All authors (PEM, TAD, JPB, WMD, NI, HL, JGL, PMM, PCS,
320 MW, TJV, DCB). Developed and updated ontologies: PEM, JPB, DCB, TAD, NI,
321 WMD, JGL. Developed the figures: TAD, PMM, JPB, PCS.

322

323 **Acknowledgments**

324 This material is based upon work supported by the National Science Foundation
325 (NSF, grants DBI-0641025, DBI-1062404, and DBI-1062542) and the National
326 Evolutionary Synthesis Center (NSF EF-0423641 and NSF EF-0905606). Any
327 opinions, findings, and conclusions or recommendations expressed in this
328 material are those of the author(s) and do not necessarily reflect the views of the
329 NSF. We also acknowledge support from the National Institutes of Health

330 (HG002659). We thank especially W. Eschmeyer and S. Blum for their generous
331 distribution of the Catalog of Fishes data, J. Gross and D. Wake for assistance
332 with the AmphibiaWeb data, N. Bailly and R. Froese for supplying common
333 names from FishBase, J. Alroy for providing data from the Paleobiology Database,
334 R. Mayden and K. Conway for providing information on the taxonomy of
335 Cypriniformes, M. Sabaj Pérez, M. Coburn, T. Grande, and E. Hilton for
336 providing information on the taxonomy of other teleost groups, and C. Mungall
337 for advice on OBO properties and our implementation of taxonomic ranks.
338

339 **References**

340 1 Federhen S: **The Taxonomy Project**. In: McEntyre J, Ostell J, editors.
341 The NCBI Handbook Bethesda (MD): National Center for Biotechnology
342 Information (US); 2002. Chapter 4.

343 2 Eschmeyer WN: **Catalog of Fishes**. California Academy of Sciences
344 (<http://research.calacademy.org/research/ichthyology/catalog/fishcatmain.asp>).
345 (version 3/2013).

346 3 Mabee PM, Balhoff JP, Dahdul WM, Lapp H, Midford PE, Vision TJ,
347 Westerfield M: **500,000 fish phenotypes: The new informatics landscape of**
348 **evolutionary and developmental skeletal biology**. *J Appl Ichthy* 2012,
349 **28(3):300–305**.

350 4 Dahdul WM, Balhoff JP, Engeman J, Grande T, Hilton E, Kothari C, Lapp
351 H, Lundberg JC, Midford PE, Vision TJ, Westerfield M, Mabee PM:
352 **Evolutionary characters, phenotypes and ontologies: Curating data from the**

353 **systematic biology literature.** *PLOS ONE* 2010,
354 (<http://dx.doi.org/10.1371/journal.pone.0010708>)

355 5 Fang F: **Phylogenetic analysis of the Asian Cyprinid genus *Danio***
356 **(Teleostei, Cyprinidae).** *Copeia* 2003, 714–728.

357 6 Toledo-Piza M: **The Neotropical fish subfamily Cynodontinae**
358 **(Teleostei: Ostariophysi: Characiformes): A phylogenetic study and revision**
359 **of *Cynodon* and *Rhaphiodon*.** *Amer Mus Nov* 2000, **3286**: 1–88.

360 7 Uhen MD, Barnosky AD, Bills B, Blois J, Carrano MT, Carrasco MA,
361 Erickson GM, Eronen JT, Fortelius M, Graham RW, Grimm EC, O’Leary MA,
362 Mast A, Piel WH, Polly PD, Salla, LK: **From card catalogs to computer:**
363 **databases in vertebrate paleontology.** *J Vert Paleon* 2013: **33(1)**: 13–28.

364 8 International Commission on Zoological Nomenclature: International
365 Code of Zoological Nomenclature 4th edition, *International Trust for Zoological*
366 *Nomenclature* 1999.

367 9 Froese R, Pauly P: **FishBase.** World Wide Web electronic publication.
368 www.fishbase.org, version (04/2013).

369 10 Balhoff JP, Midford PE, Lapp H: **Integrating Anatomy and Phenotype**
370 **Ontologies with Taxonomic Hierarchies,** *Proceedings of the International*
371 *Conference on Biomedical Ontology*; July 26–30th 2011; Buffalo. 2011: 426–427.

372 11 Ghiselin M: **A radical solution to the species problem.** *Syst Zool* 1974.
373 **23**:536–544.

- 374 12 Schulz S, Stenzhorn H, Boeker M: **The ontology of biological taxa.** *OUP*
375 *Bioinform.* 2008, **24**: i313–21
- 376 13 Thau D, Franz N: **Biological taxonomy and ontology development:**
377 **Scope and limitations.** *Biodiv Inform* 2010, **7**:45–66.
- 378 14 Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg
379 LJ, Eilbeck K, Ireland A, Mungall CJ, OBI consortium, Leontis N, Rocca-Seraa P,
380 Ruttenberg A, Sansone SA, Scheuermann RH, Shah N, Whetzel PL, Lewis S:
381 **The OBO Foundry: coordinated evolution of ontologies to support**
382 **biomedical data integration.** *Nat Biotechnol* 2007, **25(11)**: 1251–1255.

383 **Figures**

384 **Figure 1** Venn diagram showing overlap across source hierarchies used in the
385 construction of the VTO. Colored numbers denote number of names used in
386 database (after removal of duplicates), numbers in black represent the number of
387 intersecting terms between hierarchies. Circles are not to scale.

388 AWeb=AmphibiaWeb, NCBI=NCBI taxonomy, PaleoDB= Paleobiology
389 Database (vertebrates only), TTO= Teleost Taxonomic Ontology.

390

391 **Figure 2** Illustration of the construction of the VTO from source ontologies.

392 AWeb and TTO are grafted (dashed lines) onto the NCBI backbone; PaleoDB
393 taxa are filtered prior to merging with NCBI. AWeb=AmphibiaWeb,
394 NCBI=NCBI taxonomy, PaleoDB=Paleobiology Database (vertebrates only),
395 TTO=Teleost Taxonomic Ontology.

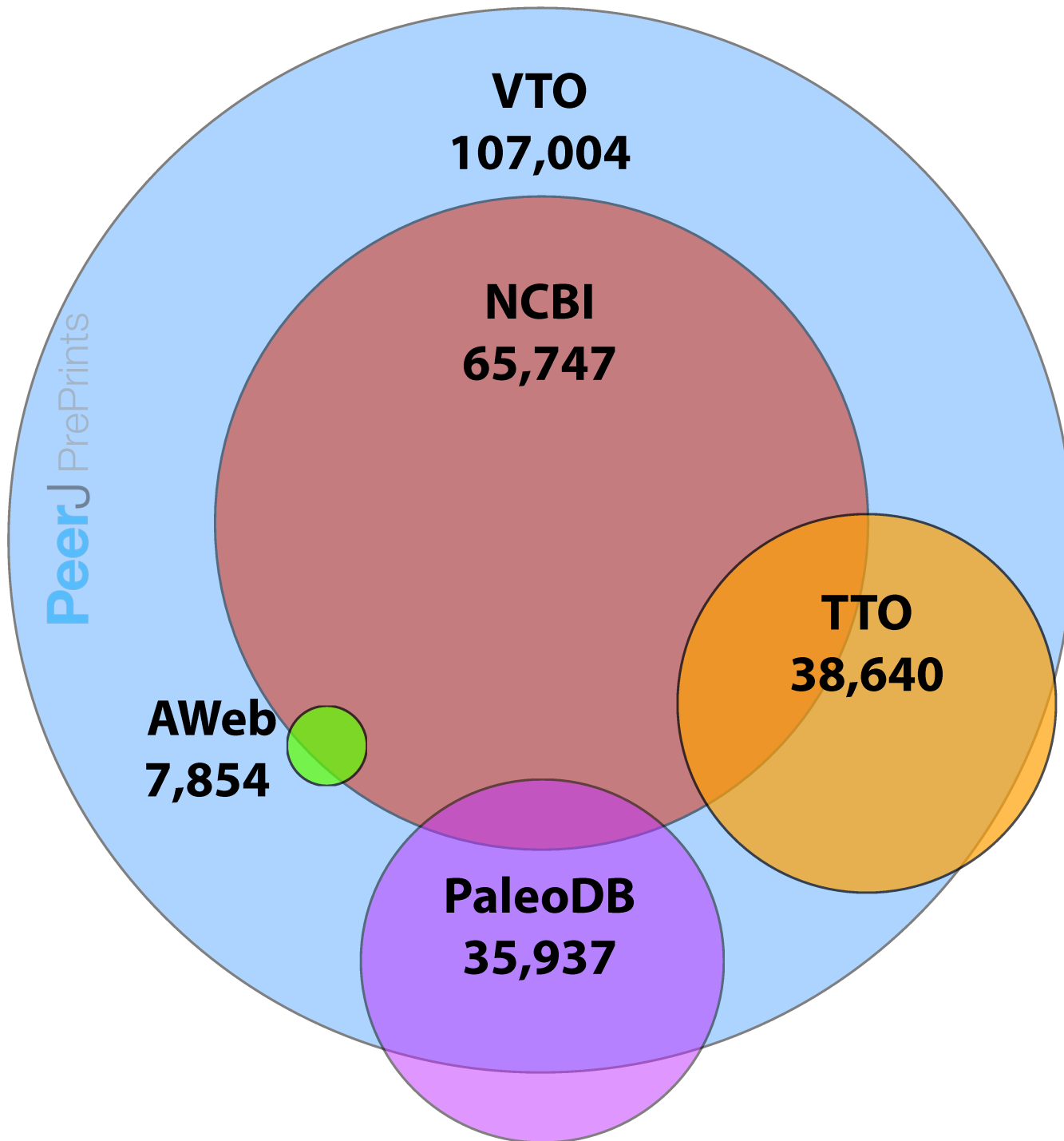


Figure 1

Extant Taxa

Extinct Taxa

PeerJ PrePrints

NCBI

PaleoDB

AWeb

TTO

Filters
Ichnotaxa
Ootaxa
Obsolete terms

Merge

Curatorial Review
Placement of extant nodes
in phylogenetic context of
fossil data

VTO

Figure 2