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

P.E. Midford1,2 ([peter.midford@gmail.com](mailto:peter.midford@gmail.com))

J.P. Balhoff2,3 ([balhoff@nescent.org](mailto:balhoff@nescent.org))

D.C. Blackburn4([dblackburn@calacademy.org](mailto:dblackburn@calacademy.org))

W.M. Dahdul2,5 ([wasila.dahdul@usd.edu](mailto:wasila.dahdul@usd.edu))

T.A. Dececchi5 ([alex.dececchi@usd.edu](mailto:alex.dececchi@usd.edu))

N. Ibrahim6 ([nibrahim@uchicago.edu](mailto:nibrahim@uchicago.edu))

H. Lapp2 ([hlapp@nescent.org](mailto:hlapp@nescent.org))

J.G. Lundberg7 (lundberg@ansp.org)

P.M. Mabee5 ([pmabee@usd.edu](mailto:pmabee@usd.edu))

P.C. Sereno6 ([dinosaur@uchicago.edu](mailto:dinosaur@uchicago.edu))

M. Westerfield8 ([monte@uoneuro.uoregon.edu](mailto:monte@uoneuro.uoregon.edu))

T.J. Vision2,3 ([tjv@bio.unc.edu](mailto:tjv@bio.unc.edu))

1Department of Ecology and Evolutionary Biology, University of Kansas

2National Evolutionary Synthesis Center

3Department of Biology, University of North Carolina

4Department of Vertebrate Zoology and Anthropology, California Academy of Sciences

5Department of Biology, University of South Dakota

6Department of Organismal Biology and Anatomy, University of Chicago

7Academy of Natural Sciences, Philadelphia

8Institute of Neuroscience, University of Oregon

**Abstract**

One of the most common ways to organize biodiversity data is by species placed in hierarchical classification of named taxonomic groups. Computational analysis and semantic integration of such data requires species taxonomies that are expressed in ways suitable for analytical processing. Yet, even for a group as well studied as vertebrates, there is currently no single expansive and authoritative taxonomy that includes extinct and extant taxa, links valid nomenclature to synonyms and common names, and that preserves stem and crown group relationships across the group. Compiling such a resource by integrating from multiple taxonomic sources is challenging due to the lack of standardization among taxonomies. To meet the needs of large-scale comparative biology research, in particular the Phenoscape project (<http://phenoscape.org/>), we created the [Vertebrat](http://purl.obolibrary.org/obo/vto.owl)e Taxonomy Ontology, a semantically defined taxonomic resource spanning both extant and fossil vertebrates. The information in VTO is derived from the integration of existing authoritative sources. It currently contains 104,746 valid terms, which are annotated with 83,992 synonyms, 99,933 taxonomic authorities (i.e. author, year) and 20 taxonomic ranks. We describe our approaches to the key challenges in constructing the VTO: (1) extracting and merging names, synonyms, and identifiers from different highly heterogeneous sources; (2) grafting together hierarchies and higher-level taxonomic node names from conflicting sources; and (3) automating this process so it is repeatable and can accommodate updates in the source taxonomies. The VTO is actively used in the Phenoscape project for integrating evolutionary phenotypes across the vertebrate systematics literature.  It is the basis for algorithms that use this information to assign profiles of phenotypic diversification to particular named and ranked groups in the taxonomy, and it is used to discover candidate genes for evolutionary transitions in vertebrate phenotypes.

KEYWORDS: evolutionary biology, fossils, taxonomy, vertebrates

**Background**

Semantic integration of data for biodiversity research requires taxonomic ontologies. In conjunction with the Phenoscape project we constructed an integrated taxonomic ontology: the Vertebrate Taxonomy Ontology (VTO;<http://purl.obolibrary.org/obo/vto.owl>). The Phenoscape project aims to bring together rich and diverse species-associated morphological data with genetic and genomic information and to integrate phenotypes and developmental data from model organisms with that from non-model systems, including extinct taxa. These goals have necessitated extending the taxonomic breadth of existing resources, such as the NCBI taxonomy, to include both extant and extinct taxa. To achieve this goal we adopted the principles of phylogenetic systematics/classification in this work. As of April 2013 the VTO contained 104,746 terms.  These terms are annotated with 83,992 synonyms, 99,933 references, and 20 ranks. This ontology is free and available to all users under the Creative Commons Zero (CC0) public domain waiver.

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

**Construction and Content**

**Selection of sources**

The first step in constructing the VTO was to identify existing taxonomy sources that could be merged to ensure inclusion of all vertebrates. No existing single source had the requisite coverage. For instance, the NCBI taxonomy, while taxonomically broad and commonly used for bioinformatics applications, is largely limited to species with published genetic data, and thus excludes many extant and nearly all extinct taxa (Figure 1). Selected sources must provide complementary information covering major vertebrate clades both extinct and extant, in addition to basal chordates. Each source must be electronically available and have compatible terms of use for mixing and distribution. Our process included obtaining permission to use and redistribute the possibly copyrighted information aggregated from our chosen source taxonomies. As a starting point for the VTO, we selected four taxonomic resources based on taxonomic coverage, authority and availability: (1) the TTO, a taxonomic ontology for fishes developed during the first phase of the Phenoscape project; (2) AmphibiaWeb; (3) the vertebrate portion of the NCBI taxonomy covering extant amniotes; and (4) the Paleobiology Database (PaleoDB).

The VTO does not seek to publish new taxonomic names, and thus we only import names from our source taxonomies. In keeping with the practice recommended for OBO Foundry ontologies [1], it is possible to modify the structure and taxonomic scope of the VTO by curator requests submitted through the term request tracker (<http://purl.obolibrary.org/obo/vto/tracker>), such as for correction of misspellings or hierarchy errors. If a curator seeks to add taxa to VTO, they are encouraged to instead add them directly to the source taxonomies following their prescribed curation methodologies. These will be incorporated into the VTO during periodic updates.

**NCBI Taxonomy**

The vertebrate portion of the NCBI taxonomy was chosen as the basal hierarchy, because it is a curated consensus view of taxonomic relationships [2]. It offers broad coverage, at a coarse level, and it provides valuable linkages to molecular data for vertebrates. Its expansive nature, covering all extant organisms, provides a framework onto which clade-specific and granular hierarchies such TTO and AmphibiaWeb can be grafted (see below). As of May 2, 2013, the NCBI contained 63,987 taxonomic names for vertebrates, including unidentified specimens (e.g., ‘Squamate sp.’). Of these, the VTO incorporates 46,649 names, discounting unidentified specimens.

**Teleost Taxonomy Ontology**

The TTO is a taxonomy ontology based on the Catalog of Fishes (CoF, [3]) and modified by contributions from taxonomic experts as part of the Phenoscape project [4]. The TTO is continually updated to reflect both changes in CoF and to incorporate additional taxa encountered during the process of curation in Phenoscape [5]. These include taxa known only from fossils, subjective names, misspellings, and names used as placeholders in manuscripts or publications before a formal taxonomic treatment is available (e.g., “*Danio* aff*. dangila*, Fang 2003 [6]” or “*Agoniates* sp. Toledo-Piza 2000 [7]”). Specimens that are not given a species designation are assigned by TTO to a nonspecific taxon that includes a citation (e.g., “*Eigenmannia* sp*.* (Fink and Fink 1981) ”). This practice taxonomically places specimens that are useful for data integration, but that are excluded from traditional taxonomies such as CoF due to uncertainty in species affinities at time of publication.

A tool called TTOUpdate (http://phenoscape.org/wiki/TTOUpdate\_tool) was developed by the Phenoscape project to automatically update TTO with each new release of the CoF. As of May 2, 2013, the TTO contains 38,640 taxonomic terms for valid species and higher taxa and 60,028 synonyms (taxonomic synonyms and vernacular names). VTO does not retain taxonomic information from TTO for clades outside Actinopterygii or Chondrichthyes, relying instead on NCBI for this hierarchy, thus it excludes 77 taxa (~0.002% of TTO).

**AmphibiaWeb**

For living amphibians, we chose to graft the AmphibaWeb taxonomy (obtained from <http://amphibiaweb.org/amphib_names.txt>) onto the NCBI backbone.  The hierarchy and taxon sampling of the AmphibiaWeb taxonomy is more expansive than that of NCBI for this clade. It is updated frequently, available online, and widely used by the community. As of April 2013, AmphibiaWeb contains 7,854 taxonomic names, all of which are incorporated into VTO.

**Paleobiology Database**

The Paleobiology Database (PaleoDB, [http://paleodb.org](http://paleodb.org/)) is a curated collection that attempts to cover the entirety of the fossil record including taxonomy, specimen locations, and stratigraphic distributions [8]. Its primary use is as a repository of occurrence data to allow for large-scale paleobiogeographic analyses; thus, in addition to named biological species, it includes data on trace fossils and unidentified body fossils. The PaleoDB provides not only a listing of all currently published taxon names, but also includes other identifiers (such as long-obsolete synonyms and trace fossil taxa) that are not needed for the purposes of our work using the VTO. As of May 2013 PaleoDB contained 33,164 extinct vertebrate taxa (out of its 39,775 total vertebrate taxa), across all ranks and including trace fossils. Of these VTO incorporates 18,961 with the rest being obsoleted based on criteria listed below.

**Constructing the VTO**

To merge the desired portions of each source hierarchy into a single tree, we developed a Taxonomy Ontology Tool (<https://github.com/NESCent/Taxonomy-Ontology-Tool>) that allows different sources to provide the hierarchy at different nodes. The tool is also capable of merging lists of synonyms from multiple sources based on matches between primary names. From the starting point of the NCBI hierarchy, we replaced certain sub-hierarchies in the one provided by NCBI with the corresponding more complete hierarchies provided by other sources. Specifically, we used TTO to replace the NCBI taxonomy under the nodes ‘Actinopterygii’ and ‘Chondrichthyes’, and we replaced the node ‘Amphibia’ and its descendants with content from AmphibiaWeb .

For integrating the PaleoDB, we took a different approach because it required first generating a separate taxonomy that was subsequently merged with the VTO composed of NCBI with the grafted sections from Amphibiaweb and TTO. Unlike the TTO or AmphibiaWeb, the contributions of the PaleoDB are not restricted to a single subgroup of the hierarchy, but are scattered across the entire topology. We approached PaleoDB’s taxonomic information as if it lacked a common root and then added each taxon to the lowest ranking existing node within the VTO tree, followed by splicing out invalid taxa (as stated in PaleoDB) in a post-processing step. This can lead to taxa lacking parents in the hierarchy. We treated such taxa by simply attaching them to the root, from where curators can subsequently regraft them. PaleoDB also includes taxonomically invalid names and names associated with non-diagnostic material for biological species identification (e.g., trace fossils or ootaxa ). These were obsoleted based on both expert input and details from the literature. Reasons for obsolescence were recorded in the “comment” annotation in the VTO.

**VTO modeling pattern**

The VTO uses the same modeling pattern as that used by C. Mungall for rendering the NCBI taxonomy in OBO format, from which it is converted to OWL (http://purl.obolibrary.org/obo/ncbitaxon.owl). In this pattern, each taxon is modeled as an ontological class, and ranks are assigned to taxon classes using the ‘has\_rank’ annotation property (declared in a separate taxonomic rank vocabulary, see below).

**Taxonomic rank vocabulary**

In order to have a consolidated vocabulary to annotate taxon terms with their taxonomic rank, we constructed a new vocabulary of taxonomic rank terms, the Taxon Rank vocabulary (<http://purl.obolibrary.org/obo/taxrank.owl>). The vocabulary augments the rank terms used by the NCBI taxonomy with those proposed in the Biodiversity Information Standards (TDWG) TaxonRank vocabulary (http://rs.tdwg.org/ontology/voc/TaxonRank). The resulting vocabulary contains 59 terms and links back to the corresponding terms in the source vocabularies. We maintain the vocabulary separately from VTO to promote reuse by other projects.

**Synonyms**

Taxonomic names can undergo multiple status revisions; as a consequence, taxonomic names encountered in legacy literature may not be the currently used names. Further, at any given time, multiple authors may use different scientific names for designating the same biological species; for example, two scientific names of the American Bullfrog are presently in common use (i.e., *Rana catesbeiana* and *Lithobates catesbeianus*). Supporting integration of species-related annotations across the literature required coverage of synonyms that allows for discovery of candidate taxon terms in the absence of an exact match. This categorization is orthogonal to one reflecting name status as a result of taxonomic revisions, such as whether a name is a junior or senior synonym. Additionally, failure to find a match for a taxon term in literature text can be due to misspelling or to referral by common (vernacular) name. Including these as synonyms was aided by the support of the OBO format for ontology author-defined 'type' tags for synonyms. This allowed us to define tags distinguishing these kinds of synonyms. The ability to include common name synonyms also allows us to make data organized by the VTO more readily accessible to non-expert users.

Aapproximately 14,400 English common names were made available by Fishbase [12] for the development of the TTO, and are therefore now also included in the VTO. A similar process occurred with names from AmphibiaWeb. Each source taxonomy has a mechanism for including common or vernacular names, though the details differ slightly. Because our use-cases do not require distinguishing the taxonomic status of various synonyms, we made no effort to include those as well. Names that are unavailable based on ICZN are excluded. The VTO adds an additional annotation property ‘is\_extinct’ to indicate taxa that are known only from fossil evidence or if such designation is present in a source ontology.

**Nodal naming and definition of clades**

Given the different species domains of the source taxonomies, there are potentially not only different names but also different definitions for particular nodes in the vertebrate taxonomy. While both the NCBI taxonomy and PaleoDB include some representation of both extinct and extant taxa, the overlap is superficial and does not resolve problems related to how large clades are defined. Without a common definitions of what constitutes these clades, integrating taxonomies into a single ontology can lead to confusion and mistaken inferences for data annotated with it.

We illustrate this here specifically for Tetrapoda (see [13] for an in-depth analysis of the problem), though we note that similar issues arise for other major groupings, including Mammalia [14] and Aves [15]. There are 3 primary ways to define taxonomic groups: Total group, apomorphy-based, and crown group [16] (Figure 2). A total group definition of Tetrapoda includes all sarcopterygian fishes that are more closely related to living amphibians and amniotes than any either dipnoans or coelacanths [17]. This definition differs from the traditional definition of Tetrapoda as it includes many finned forms without digits. The apomorphy-based definition restricts the term Tetrapoda to the last common ancestor of all taxa that possess digits and all of its descendants [18,19]. Challenges to this system include the problem of missing data, for example in taxa near the transition without preserved limbs, and convergent evolution or the transitory stages of a character along the phylogenetic stem [17]. A crown group definition restricts the term Tetrapoda to only include the last common ancestor of living amphibians (Lissamphibia) and amniotes. This definition approximates the one in the NCBI taxonomy because it does not sample the diversity of fossil taxa bridging the node Sarcopterygii to Tetrapoda (Figure 2). This usage is problematic in that it is a restrictive use relative to that in the majority of the paleontological literature and the traditional definition of tetrapod [13].

We chose the apomorphy-based definition for Tetrapoda for use by the VTO because it retains the most commonly applied sense of the term both in colloquial usage and in the literature pertinent to the data we annotate using the VTO. Tetrapoda in VTO is defined as “The first species derived from Sarcopterygians that possessed digits, and all its descendants” [18,19].

**Discussion and Utility**

The VTO is built as a simple hierarchy of ontology classes. Although this closely follows a traditional view of taxonomy, arguments for treating taxonomic entities as historical individuals suggest alternative approaches (e.g., Ghiselin 1974 [9]). Others have explored different ontological models that attempt to capture the complex interplay between evolutionary relationships and the practice of taxonomic classification [10,11]. In construction of the VTO we have focused on our use cases, which are satisfied by a basic representation of the taxonomic hierarchy.

**Utility**

Taxonomic ontologies are required for semantic integration of ‘natural’ phenotypic variation among species with phenotypes associated with genetic manipulation of model organisms. Such integration enables reasoning from species phenotypes to candidate genes [4] and is the goal of the Phenoscape project (<http://phenoscape.org/>). Curation of phenotypes in Phenoscape entails translating phenotype descriptions into the Entity-Quality (EQ) formalism, and assigning these EQ descriptions to appropriate taxa that could be species, genera, families, or higher-level groupings [5]. This requires a single taxonomic ontology to aggregate and compute on taxon-associated data across studies.

**Analysis and Reasoning**

The basic taxonomic class hierarchy provided by VTO allows straightforward data aggregation via subsumption reasoning. For example, data referencing the ontology classes ‘Rodentia’ or ‘Primates’ should be returned from a query using ‘Mammalia’. However, linking phenotypic data to ontology classes can introduce reasoning inconsistencies when the phylogenetic history of the features involves character reversal or evolutionary loss [20]. Modeling taxa instead as ontological individuals may be better suited to represent features that change over evolutionary time, and indeed we are currently evaluating this alternative approach.

**Future Directions**

**Display**

A taxonomy ontology appears in the Phenoscape Knowledgebase (KB) as content as well as a way to navigate other content.  The TTO is currently used; the VTO will be used shortly. The Phenoscape KB includes a display page for each taxon in the TTO that includes its immediate taxonomic parents and children, as well as synonyms, extinction status, and links to other source data for that taxon (e.g., Fishbase, Wikipedia, etc.). Taxon display pages also include a browsable taxonomy tree. Taxa are included in the results returned by queries for phenotypes (e.g., all reported phenotypes involving fin shape, or phenotypes where the basihyal is absent and the pectoral fin is triangular), but it can also be used to scope queries (all Cyprinidae with triangular fins), as well as appearing as elements in faceted queries. The Phenoscape KB additionally displays summary phenotype statistics such as degree of annotation coverage and phenotypic variation on a simplified tree.

**Additional data sources**

We continue to improve the VTO, striving to make it universally relevant for vertebrate researchers. To further this agenda we are exploring integration of data from International Ornithologists’ Union Bird List ([www.worldbirdnames.org](http://www.worldbirdnames.org/)), the IUCN Red List of Threatened Species (<http://www.iucnredlist.org/>) and the Reptile Database (<http://www.reptile-database.org/>). The inclusion of these resources would help expand the existing scope of amniotes in the VTO beyond those in the NCBI taxonomy and PaleoDB and include conservation status data (such as is\_extinct for modern extinction events) to broaden the appeal of this resource.

**Conclusions**

Taxonomies in a form that supports machine reasoning can be key ingredients for large-scale knowledge discovery engendered by integrating big data from genomics, evolutionary and systematic studies. Computation across diverse data associated with extant and fossil species often requires a more inclusive set of taxa than is presently represented in any single taxonomy. To fill this gap for vertebrates, we have developed a taxonomy ontology, the VTO, through merging and translating taxonomies from a variety of expert sources. The integration pipeline we have developed is repeatable, and is thus capable of incorporating updates from the source taxonomies. While the development to date has been guided by the requirements of the Phenoscape project, the VTO can be useful to any cross-vertebrate research endeavor. The methods to generate it are available in source code form under an open-source license on a public code repository, and as the VTO itself are thus open to adapting to the needs of other projects.

**Availability and requirements**

The Vertebrate Taxonomy Ontology (VTO) and the vocabulary of Taxonomic Ranks that it references are available via their permanent URLs: VTO, <http://purl.obolibrary.org/obo/vto.owl>; TAXRANK, <http://purl.obolibrary.org/obo/taxrank.owl>. The VTO is a large ontology, and thus when viewing in a desktop OWL editor such as Protégé may require allocating sufficient memory (2 GB at present) to the program.

**Competing interests**

None declared.

**Author’s contributions**

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

**Acknowledgments**

This material is based upon work supported by the National Science Foundation under Grant Numbers DBI-0641025, DBI-1062404, and DBI-1062542, and supported by the National Evolutionary Synthesis Center under NSF EF-0423641 and NSF #EF-0905606. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We also acknowledge support from NIH (HG002659). We thank especially W. Eschmeyer and S. Blum for their generous distribution of the Catalog of Fishes data, J. Gross and D. Wake for permission and assistance with the AmphibiaWeb data, and J. Alroy for providing data from the Paleobiology Database. R. Mayden and K. Conway provided information on the taxonomy of Cypriniformes. M. Sabaj Pérez, M. Coburn, T. Grande, and E. Hilton provided information on the taxonomy of other teleost groups. Chris Mungall provided advice on OBO properties and our implementation of taxonomic ranks.

**References (inconsistencies below will be fixed)**

1. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ, OBI consortium, Leontis N, Rocca-Seraa P, Ruttenberg A, Sansone SA, Scheuermann RH, Shah N, Whetzel PL, Lewis S: **The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration.** *Nat Biotechnol* 2007, **25(11)**: 1251-1255,
2. Federhen S: **The Taxonomy Project.** In: McEntyre J, Ostell J, editors. The NCBI Handbook Bethesda (MD): National Center for Biotechnology Information (US); 2002-. Chapter 4.
3. Eschmeyer WN: **Catalog of Fishes**. California Academy of Sciences (<http://research.calacademy.org/research/ichthyology/catalog/fishcatmain.asp>). (version 3/2013).
4. Mabee PM, Balhoff JP, Dahdul WM, Lapp H, Midford PE, Vision TJ, Westerfield M: **500,000 fish phenotypes: The new informatics landscape of evolutionary and developmental skeletal biology.** *J Appl Ichthy* 2012, **28(3)**:300-305.
5. Dahdul WM, Balhoff JP, Engeman J, Grande T, Hilton E, Kothari C, Lapp H, Lundberg JC, Midford PE, Vision TJ, Westerfield M, Mabee PM: **Evolutionary characters, phenotypes and ontologies: Curating data from the systematic biology literature.** *PLOS ONE* 2010, (http://dx.doi.org/10.1371/journal.pone.0010708)
6. Fang F: **Phylogenetic analysis of the Asian Cyprinid genus *Danio* (Teleostei, Cyprinidae)**. *Copeia* 2003, 714-728.
7. Toledo-Piza M: **The Neotropical fish subfamily Cynodontinae (Teleostei: Ostariophysi: Characiformes): A phylogenetic study and revision of *Cynodon* and *Rhaphiodon*.** *Amer Mus Nov* 2000, **3286**: 1-88.
8. Uhen MD, Barnosky AD, Bills B, Blois J, Carrano MT, Carrasco MA, Erickson GM, Eronen JT, Fortelius M, Graham RW, Grimm EC, O’Leary MA, Mast A, Piel WH, Polly PD, Saila, LK: **From card catalogs to computer: databases in vertebrate paleontology.** *J Vert Paleon* 2013: **33(1)**: 13.-28.
9. Ghiselin M: **A radical solution to the species problem.** *Syst Zool* 1974. **23**:536-544.
10. Schulz S, Stenzhorn H, Boeker M: **The ontology of biological taxa**. *OUP Bioinform*. 2008, **24**: i313-21
11. Thau D, Franz N: **Biological taxonomy and ontology development: Scope and limitations.** *Biodiv Inform* 2010, **7**:45-66.
12. Froese R, Pauly P (2013): **FishBase.** World Wide Web electronic publication. www.fishbase.org, version (04/2013).
13. Laurin M, Anderson JA: **Meaning of the Name Tetrapoda in the Scientific Literature: An Exchange**. *Syst Biol* **2004**: 68-80.
14. Rowe T, Gauthier J: **Ancestry, paleontology, and definition of the name Mammalia**. *Syst Biol* 1992, **41**: 372-378
15. Padian K: **Basal Avialae**. In *The Dinosauria,* 2nd edition,Edited by Weishampel DB, Dodson P, Osmolska H. Berkeley: University of California; 2004: 210–231.
16. de Queiroz K, Gauthier J: **Phylogenetic taxonomy**. *Annu Rev Ecol Syst* 1992: **23**:449**–**480.
17. Ruta M, Coates MI, Quickie DLJ: **Early tetrapod relationships revisited**. *Biol Rev* 2003,**78**: 251-345.
18. Ahlberg PE, Clack JA: L**ower jaws, lower tetrapods—A review based on the Devonian genus *Acanthostega***. *Trans R Soc Edinb Earth Sci* 1998, **89**: 11-46.
19. Anderson JA: **Use of Well-Known Names in Phylogenetic Nomenclature: A Reply to Laurin**. *Syst Biol* 2002, **51**: 822-827.
20. Balhoff JP, Midford PE, Lapp H: **Integrating Anatomy and Phenotype Ontologies with Taxonomic Hierarchies**, *Proceedings of the International Conference on Biomedical Ontology*;  July 26-30th 2011; Buffalo. 2011: 426-427.
21. Shubin NH, Daeschler EB, Jenkins jr. FA: **The pectoral fin of *Tiktaalik roseae* and the origin of the tetrapod limb.** *Nature* 2008, **440:** 764-771.
22. Padian K, Chiappe LM: **The origin and early evolution of birds.** *Biol. Rev.* 1998, **73**: 1-42
23. Villee CA: In *Biology*. 7th Edition. 1977 Philadelphia W.B Sander Co.
24. Coates MI, Ruta M: **Skeletal changes in the transition from fins to limbs**. In: *Fins into limbs: Evolution, development, and transformation****.*** Edited by. Hall BK.Chicago: University of Chicago Press; 2007: 15-39.

**Figures**

**Figure 1**: Venn diagram showing overlap across source hierarchies used in the construction of the VTO. Coloured numbers denote number of names used in database (after removal of duplicates), numbers in black represent the number of intersecting terms between hierarchies. AWeb=AmphibiaWeb, NCBI=NCBI vertebrate taxonomy, PaleoDB= Paleobiology Database, TTO= Teleost Taxonomic Ontology. Circles are not to scale.

**Figure 2**: A simplified representation of sarcopterygian taxonomy and how differing definition for Tetrapoda influence nodal placement. 1=Total group definition of Tetrapoda as per [17]. 2=The apomorphy based definition as per [18]. This is the definition of Tetrapoda used in the VTO. 3=Crown group Tetrapoda corresponding closely to the usage in NCBI. Images modified from [21-24].