BIOLOGICAL TAXONOMY AND ONTOLOGY DEVELOPMENT: SCOPE AND LIMITATIONS

NICO M. FRANZ[1] AND DAVID THAU[2]

[1]Department of Biology, Call Box 9000, University of Puerto Rico, Mayagüez, PR 00681-9000 USA, nico.franz@upr.edu; [2] Department of Computer Science, University of California at Davis, Kemper Hall, One Shield Avenue, Davis, CA 9561 USA, thau@learningsite.com

Abstract. — The prospects of integrating full-blown biological taxonomies into an ontological reasoning framework are reviewed. Traditionally ontological representations of taxonomy have adopted the model of a single and static hierarchy. This model is contrasted with a more realistic situation involving dynamic revisions of particular groups and alignments among alternative taxonomic perspectives. Taxonomic practice is bound by a range of epistemological constraints and linguistic conventions that run orthogonal to the logical background from which ontological entities and relationships originate, resulting in severe challenges for ontological representation and reasoning. In particular, the purported existence of a single hierarchy in nature forces taxonomists to gradually approximate this hierarchy and make frequent rearrangements in light of new evidence. The evolvability of taxa implies that taxon-defining features may be lost in subordinate members or independently gained across multiple sections of the tree of life. As a result, many terms for phenotypic properties are phylogenetically underdetermined and have limited hierarchical transitivity. The standard approach of defining taxa both in reference to properties (intensional) and members (ostensive) undermines the individual/class dichotomy that sustains conventional ontologies, and compromises the reasoning capabilities associated with this distinction. Neither the use of Linnaean ranks in taxonomy nor the 250-year legacy of nomenclatural adjustments have obvious analogues in the ontological realm. Lastly, the piece-meal appearance of full-blown taxonomic information makes it necessary to use expert alignments to obtain a comprehensive static perspective. In light of these limitations, research along the taxonomy/ontology interface should focus either on strictly nomenclatural entities and relationships or on ontology-driven strategies for aligning multiple taxonomies, but not on building static networks for large portions of the tree of life. The prospects of using ontology-based services in taxonomy will largely depend on the ability of the taxonomic expert community to present its products in ways that are more compatible with ontological principles than concurrent practice.

Key words. — classes, evolvability, individuals, intension, legacy integration, nomenclature, ontology alignment, ostension, phylogeny, taxonomic concepts.

Biologically oriented ontologies are regarded as essential for integrating a vast range of biological data (Goble and Stevens, 2008). Having originated in the biomedical sector (Smith et al., 2007), the development of such ontologies is now branching out into other disciplines including behavior (Midford, 2004), ecology (Jones et al., 2006; Madin et al., 2008), evolution (Mabee et al., 2007), anatomy (Ramírez et al., 2007; Mikó and Deans, 2009; Vogt et al., 2009; Balhoff et al., 2010; Dahdul et al., 2010; Mungall et al., 2010), and taxonomy (Schulz et al., 2008). Nevertheless, the impact of ontology-based information services has been uneven, in part because of differences in

needs and resources among disciplines. It may seem particularly intriguing that the field of systematics[1] was *not* among the first to shift towards using ontologies, even though species names are prime vehicles for transmitting biological data and taxonomic hierarchies appear very similar to ontology networks. In spite of such an apparent match, actual representations of taxonomy in information services such as GenBank are rather simplistic (cf. Wheeler et al., 2008); and therefore are unable to support

---

[1] The terms *systematics* and *taxonomy* are used interchangeably herein, and each encompasses phylogenetics as a subdiscipline that informs the establishment of classifications (Wheeler, 2004).

semantically complex searches or inferences across alternative taxonomic perspectives (Page 2004, 2007; Thau and Ludäscher, 2007; Dahdul et al., 2010). This is unfortunate because a thoroughgoing ontology for biological taxonomy could play the same overarching role in the bio-ontological domain that taxonomy plays in biology (Schulz et al., 2008).

This paper examines to what extent a full-blown representation of biological taxonomy in the ontological domain is possible. Using Schulz et al.'s (2008) pioneering work as a point of departure, we first observe that there are two fundamentally different models to create ontological representations of taxonomy; viz. strictly nomenclatural and full-blown taxonomic representations. Each model serves a different purpose and user domain. We furthermore distinguish between static and dynamic perspectives of taxonomic classifications, arguing that only the latter is well suited to represent the reasoning needs of taxonomists. We then describe a number of ontological representation challenges posed by biological taxonomy under the full-blown dynamic model; including epistemological limitations, linguistic conventions, and alignment challenges. Our review concludes with a series of practical recommendations for advancing the taxonomy/ontology interface, with an emphasis on research tasks for the taxonomic expert community.

NOMENCLATURAL *VERSUS* FULL-BLOWN TAXONOMIC REPRESENTATIONS

Biological taxonomy integrates a wide variety of heterogeneous data, including information on physical specimens and voucher samples, taxonomic names and nomenclatural relationships, phenotypic and genotypic properties of taxa, and hierarchical phylogenetic groupings and classifications. Some user communities are well served by focusing on a subset of this immense data pool. For instance, taxonomic experts might have great use for an ontology of strictly *nomenclatural* relationships (homonymy, synonymy, typification, etc.) among accepted and rejected Linnaean names (Scoble, 2004; Kennedy et al., 2005; Page, 2006; Franz and Peet, 2009; Huber and Klump, 2009). Such nomenclatural entities and relationships have a tractable history, and representations of this history in an ontological

network are useful – up to a point – for identifying and integrating taxonomic legacy information (Berendsohn et al., 2003). Nevertheless, nomenclatural relationships such as "is a misapplied name for" or "is a pro parte synonym for" are often not sufficient to permit reasoning among of alternative and succeeding taxonomic perspectives (Franz, 2005a). To our knowledge no comprehensive ontology of nomenclatural entities and relationships exists, making this a worthwhile if semantically limited pursuit.

Other examples of ontologies with a strong taxonomic component include phenotype ontologies for model organisms (Smith et al., 2007). By design, these phenotype networks fall short of representing the depth and breadth of taxonomic information across large sections of the tree of life.

In what follows, we will deliberately concentrate on ontologies that aim to represent authoritative, expert-produced phylogenies and classifications, as opposed to just nomenclatural relationships. We refer to such ontologies as *full-blown* if they permit inferences ranging from queries about the taxonomic identity of specimens to reasoning about evolutionary properties across multiple lineages (see Dahdul et al., 2010, for a rationale of this approach). In this context, the ontological framework developed by Schulz et al. (2008) is remarkable and pioneering effort. The authors (pp. i314-i315) intend "to provide classes and classificatory criteria to categorize the foundational kinds of biology, without any restriction to granularity, species, developmental stages or states of structural well- or ill-formedness". To this end they utilize a set of complementary ontological relationships (cf. fig. 2 on p. i318); including *is_a* (example: a taxon quality is a quality), *part_of* (example: the Mammalia class region is part of the Animalia kingdom region), and *instance_of* (example: the individual "Clyde" is an instance of the species *Elephas maximus*); as well as more taxonomic-specific relationships such as *inheres_in* (example: the quality of having an elephant heart inheres in the species *Elephas maximus*); *derives_from* (example: the HeLa cell line derives from human cells [but exists separately from the human population]), *has_granular_part* (example: the population of Thai elephants has a granular part "Clyde"), and *located_in* (example: a species

quality is located in the *Elephas maximus* species region). Suitable combinations of these relationships allow Schulz et al. (2008) to connect individual organisms and populations (particulars) to their respective species (universals), link their properties (qualities) to particular species, and integrate these qualities up to higher levels (regions) in the taxonomic hierarchy (see Fig. 1). Schulz et al.'s (2008) ontology is therefore a first approximation of the kind of full-blown ontological representation of biological taxonomy that is our focus hereinafter.

STATIC *VERSUS* DYNAMIC PERSPECTIVES

Ontologies are necessarily biased to suit particular representation and reasoning demands (McCray, 2006). Accordingly, ontologies for biological taxonomy may focus on truly ontological aspects of taxa *in nature* (example question: *what* are species?); or alternatively, on the epistemological challenges of representing taxa *in systematic practice* (example question: *how* did Rivera delimit *Mus musculus* in her 1980 publication?).[2] On one side of the spectrum, ontology creators must address such contentious issues as whether taxa are real in a suitable sense (Ereshefsky, 2002; Lee, 2003; Johansson, 2006; Merrill, 2010), what natural boundaries exist between species and populations (Hey, 2006; Schulz and Hahn, 2007), and whether taxa are historical individuals or natural kinds (Ghiselin, 1997; Boyd, 1999). Different perspectives on these theory-laden issues will lead to different ontological implementations, as is reflected in Schulz et al.'s (2008) elaborate attempts to represent taxa as meta-properties, class hierarchies, populations, qualities, or qualia (regions of qualities). At the same time the authors place little emphasis on how actual taxonomies come about, relying instead on use cases that consider taxonomic hierarchies as a relatively simple and stable (Fig. 1).

Ontology development on the other side of the spectrum focuses on representing concepts and relationships inherent in taxonomic publications (e.g., Yoon and Rose, 2001; Pullan et al., 2005; Thau and Ludäscher, 2007; Franz et al., 2008;
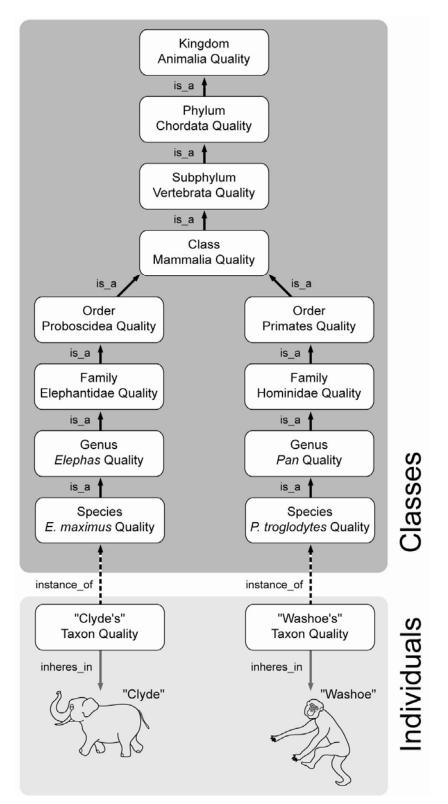
Thau, 2008; Thau et al., 2008, 2009; Franz and Peet, 2009). These efforts are motivated by the need to integrate alternative taxonomic classifications and biological linked to them. The goal is therefore to extract whatever ontology-compatible information is encrypted in a published taxonomy. Unfortunately most taxonomic works display some form of logical inconsistency (Thau and Ludäscher, 2007; Franz, 2009; Vogt et al., 2009); including regional and taxonomic sampling biases, incomplete semantic treatments, implicit references to other works, and mosaic (as opposed to strictly hierarchical) distributions of phenotypic features (Fig. 2). Ontological representations must capture the underlying similarities and differences between such multiple perspectives. In short, they are more concerned with representing what taxonomists *mean* than with what taxa *are.*

The two disparate views of the taxonomy/ontology interface have strong implications for ontology development and usage. Specifically, the various solutions presented in Schulz et al. (2008) all adhere to a *static* perspective of taxonomy: ontological entities and relationships are regarded as correct and stable enough to permit reasoning across a single taxonomic hierarchy. The static perspective assumes that at any given time there is a community-wide agreement as to how taxonomic entities are constituted and connected in nature. Schulz et al. (2008) are not alone in employing this view, which also underlies proposals to establish a unitary taxonomy (cf. Scoble, 2004) and is commonplace in on-line taxonomic information services including GenBank and the Catalogue of Life (Page, 2004; Berendsohn and Geoffroy, 2007; Franz et al., 2008).

Schulz et al.'s (2008) model use case (Fig. 1) shows a hierarchy passing from the kingdom Animalia through six intermediate ranks to the species *Elephas maximus.* Yet in context of taxonomy's century-long history, this classification may only represent a momentary snapshot, ignoring previous disagreements or ongoing debates about the identity of taxonomic entities. When instead classifications are modeled so as to accurately reflect their nature as intermittent summaries of ongoing research, the result is a complex network of competing hypotheses whose relative support changes over time. For instance, Vane-Wright (2003) reviews a series of alternative

---

[2] This distinction permits, but does not presuppose, that concepts used in systematic practice may correspond closely to 'true' and causally sustained entities and relationships in nature. See Merrill (2010) for further discussion.

Fig. 1. Idealized taxonomic hierarchy as depicted in Figure 1 of Schulz et al. (2008). Taxon *qualities* inhere in individual organisms (or parts thereof), and at the same time represent *instances* of species- and higher-level taxon qualities.

perspectives on species identities in the *Amauris* butterfly complex.[3] He concludes (p. 10):

> *[I]f we are to move to any system approaching Godfray's unitary ideal, then we will have to come to terms with either making arbitrary choices among competing classifications [...], or accept the additional burden of arriving at some compromise or consensus at the outset. [...] In my experience, the taxonomic community does not seem to be ready to accept the idea of compromise or consensus.*

The extent to which taxonomies evolve over time is easily underestimated. Geoffroy and Berendsohn (2003) reanalyzed 12 succeeding classifications of German mosses from 1927 to 2000 (Koperski et al., 2000). Of the 1548 taxonomic entities recognized in 2000, 44.5% were unstable, 20% were potentially unstable, and an additional 22.2% had undergone nomenclatural changes in comparison to previous treatments; leaving only 13.3% of the entities consistent in name and circumscription over a 73-year period. Franz et al. (2008) compared eight succeeding classifications of North American vascular plants from 1933 to 2006. Only 55% of the currently valid concepts remained stable over the examined time period. These numbers are expected to decrease further when analyzing more diverse and understudied lineages such as invertebrates (cf. IISE, 2009). In other words, the static model used by Schulz et al. (2008) assumes a level of stability that has so far been unattainable for many taxonomic groups.

The *dynamic* perspective relaxes the frequently unjustified requirement of taxonomic stability. Under this view, names, specimens, descriptions, illustrations, proposed homologies, molecular evidence, and other kinds of core taxonomic information are potentially varying elements of taxonomic concepts that evolve and replace each other over time (Berendsohn, 1995). A particular perspective, published by an author or group of authors at a specific time and place, makes explicit or implicit reference to some or all of these elements. Alternative perspectives are either congruent or (partially) incongruent with each other. The ontological challenge, then, is to accurately represent the elements of alternative taxonomies and connect them in ways that are logically consistent and permit reasoning across these taxonomies (Thau and Ludäscher, 2007; Thau, 2008; Thau et al., 2008, 2009).

To achieve this goal, taxonomic experts must provide an initial, non-exhaustive set of concept relationship mappings or *alignments* (Koperski et al., 2000; Franz and Peet, 2009). Given two taxonomies and the initial expert alignments, ontological reasoning can perform a series of useful inferences; including the implementation of global taxonomic constraints (non-emptiness, etc.), checks for logical consistency, logical repairs, inferences of implied relations, removal of redundant relations, and the merging of aligned concepts (Thau et al., 2008, 2009). However, as currently implemented the dynamic perspective focuses solely on the issue of concept congruence and otherwise treats the contents of taxonomic concepts as black boxes. This narrow approach is not sufficient to facilitate a full-blown ontological representation of taxonomy.

ONTOLOGICAL REPRESENTATION CHALLENGES

In allowing taxonomies to evolve over time, the dynamic perspective is best suited to represent the products and interests of the taxonomic expert community. The static approach, on the other hand, is effective when taxonomy is virtually a constant, which is only a reasonable assumption in well circumscribed lineages such as model organisms. As illustrated above, each view has its inherent strengths and limitations. Moreover, both dynamic and static ontological representations of taxonomy have so far focused on use cases that are strongly simplified in comparison to full-blown taxonomic publications. They have been developed by following the semantic restrictions of ontological languages used in computer science (e.g., OWL Web Ontology Language[4]). Because of their origin in description logic (Baader et al., 2004), these languages do not necessarily respond to special representation demands posed by biological taxonomies. They were certainly not conceived to address occurrences of varying species concepts (Wheeler and Meier, 2000) or evolving perspectives of higher-level taxa.

---

[3] Clark et al.'s (2009) model of taxonomy as an eScience similarly allows multiple perspectives to be represented and updated dynamically following peer review.

[4] http://www.w3.org/TR/owl-features/.

While it is necessary to explore the taxonomy/ontology interface with existing ontology languages like OWL, we also need to understand which phenomena they *cannot* represent, even though these phenomena may play an important role in taxonomy. By reviewing the full range of representation demands from the viewpoint of taxonomic practice, we may thus identify features that prominent ontological languages can, or cannot, handle. Clearly, both sorts of features must be understood to advance the development of ontologies for taxonomy.

In the following sections we describe a non-comprehensive set of full-blown taxonomic phenomena that constitute severe representation challenges in the ontological realm. These phenomena are often interconnected. They are organized here starting with three kinds of epistemological constraints (Sections I to III), continuing with special linguistic conventions (Sections IV and V), and ending with challenges related to the alignment of taxonomies (Sections VI and VII). Throughout, the example of a phylogeny and classification of the weevil genus *Apodrosus* Marshall (Girón and Franz, 2010) is used to illustrate the challenges being discussed (Fig. 2; Table 1).

Sections I to III – Epistemological Constraints
I. NATURE HAS THE LAST LONG WORD

Biological taxonomies are proposed and used by humans and in that sense they are socially mediated constructs (Bowker and Star, 1999). However, taxonomists working under the widely accepted paradigm of evolution have no intention to construct wholly artificial systems. Instead, they strive to *discover* natural phylogenetic relationships among the taxa under study (Hennig, 1966). Most will agree that evolution on Earth has unfolded along a linear time scale, resulting in a unique sequence of events and causally sustained relationships – a *single,* true phylogeny.[5] This phylogeny is what systematists aim to reconstruct and represent in their classifications.

Taxonomists often argue about the quantity and quality of character evidence and the validity of inference methods placing a certain group close to another one. For instance, the twisted-wing parasites (Strepsiptera) are notoriously difficult to place among the insect orders; with beetles and flies being the two most favored possibilities (Whiting, 1998; Bonneton et al., 2006). Despite such disagreements, it is understood that there are no multiple correct answers, and no answer can prevail mainly for conventional reasons. Indeed, the naturalness of taxa – their unique origin and coherence through time – is thought to exist independently of human classifications.

Biological taxonomy differs in this sense from more conventional ontologies of (e.g.) pizzas or human diseases (Horridge et al., 2004; Du et al., 2009). In many research disciplines outside of taxonomy, it is epistemologically permissible to maintain multiple alternative hierarchies. As a result, the process of generating standard ontologies more readily leads to agreements about the context-specific utility of entities and their relationships. For instance, it is more straightforward to stipulate the truth content of the relationship "CheeseTopping *is_a* PizzaTopping" or "Carcinoma of the Large Intestine *is_a* Carcinoma", than it is to assert that "Strepsiptera *is_a* Halteria" (= a member of the haltere-bearing lineage that presumably includes flies; cf. Whiting, 1998). Unlike the former two assertions, which may be useful or not depending on the inferential context, the validity of the latter assertion depends on the final outcome of complex investigations of *homology* – a special and highly theory-laden kind of similarity resulting from common ancestry (e.g., Wagner, 2001; Franz, 2005b; Assis and Brigandt, 2009). While systematists are allowed at any time to posit such a relationship about the Strepsiptera, its truth content is subject to continuous inductive testing of phylogeny and its inferential reliability is an *a posteriori* phenomenon.

Boyd (2000) uses the term *bicamericalism* to describe the process of delimiting biological taxa. This means that *both* our linguistic conventions *and* the causal structure of the world make up the content of taxa – therein understood as natural kinds – but the latter has the "final word". Indeed, the impact of nature's causal relationships on taxonomy is so strong that classifications require continuous revision in light of new evidence. Taxonomic research is an ongoing quest to approximate phylogeny where new insights may

---

[5] This is an oversimplification, particularly when applied to microbial lineages where hereditary information is often passed on horizontally (cf. Bapteste et al., 2009; Boto, 2009; Fraser et al., 2009). The main point nevertheless remains valid enough for the present context.

Fig. 2. Phylogeny and classification of the weevil genus *Apodrosus* Marshall (Girón and Franz, 2010). Species that are not part of *Apodrosus* represent three tribes: Anypotactini (*A. bi caudatus* and *P. scans orius*); Polydrusini (*C. ni grocinctus, P. coni cus, P. mutabilis* and *P. peninsularis*); and Sitonini (*S. californicus*), the latter chosen as the "root". The full scientific names are provided and select phylogenetically informative characters are mapped onto the internal branches. Grey rectangles represent synapomorphic transformations and white rectangles represent homoplasious transformations. The numbers above and below each rectangle correspond to character numbers and states, respectively. For explanations of letters A to N see Table 1.
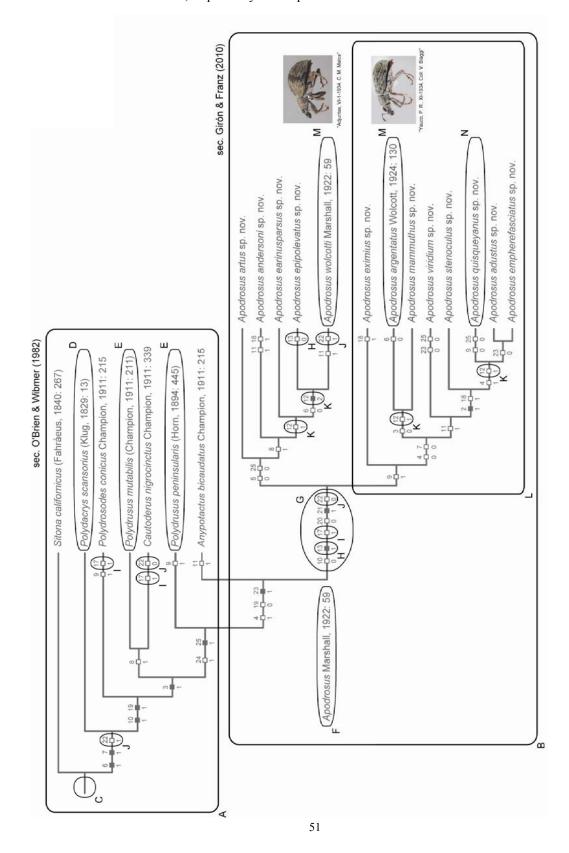

51

Table 1. Examples and explanations of linguistic phenomena (letters A to N in Fig. 2) with relevance to the taxonomy/ontology interface.

Letter	Taxonomic language phenomena
A	*Peripheral* taxonomic information ("outgroup"). Tribes and genera are not fully represented with regards to their properties or constituent members. The *names and classification* of species are in accordance with O'Brien and Wibmer's (1982) checklist. The sources of the species *concepts*, whether original or revisional, remain unspecified.
B	*Core* taxonomic information ("ingroup"). Feature-based diagnoses and lists of all species-level members (for the genus) and specimens (for the species) are provided.
C	Unspecified root, presumably connecting to the subfamily Entiminae sec. Alonso-Zarazaga and Lyal (1999) (though see A).
D	The name *P. scansorius* has two synonyms: (1) *P. luctuosus* Dejean (1837: 278), a *nomen nudum* (invalid name not bearing description); and (2) *P. modestus* Gyllenhal (1834: 131), a heterotypic junior synonym. The synonymies are listed in O'Brien and Wibmer (1982), however this information is not referenced in Girón and Franz (2010). Nor is it clear from either publication who established these. *Polydacrys scansorius* was originally described in the genus *Sitona* Germar, thus (Klug) is in parentheses. This is also not mentioned in the 2010 publication.
E	The two *Polydrusus* species are non-monophyletic in the tree. Yet due to the fragmentary nature of the peripheral taxonomic information, no classificatory changes are undertaken.
F	Girón and Franz' (2010) concept of *Apodrosus*; including 13 member species (ostensive) and the following diagnosis (intensional; in part): "*Apodrosus* is a genus of relatively small sized (3-7 mm), often metallic colored, exclusively (western) Caribbean entimine weevils with phanerognathous mouthparts, without a postocular lobe and vibrissae, and with the humeri and wings being well developed. According to Marshall (1922: 59), the genus shares with the strictly continental *Polydrusus* and other polydrusine genera 'its more salient characteristics', including a laterally situated antennal scrobe and connate claws. However, *Apodrosus* can be distinguished from *Polydrusus* and other polydrusine genera by a particular combination of characters including a median furrow on the head; a large, bare, and smooth triangular area formed by the epistome on the rostrum; the presence of premucro; the presence of a median fovea on the ventral sternum VII; and an either J- or Y-shaped female spermatheca [...]." The diagnosis includes a mix of features with a strong phylogenetic signal (e.g. "spermatheca J- or Y-shaped") or just mainly for general recognition (e.g. "size small"). The 2010 definition differs dramatically from its 1922 predecessor: it is more restrictive intensionally yet includes more species ostensively. The resulting alignment is *Apodrosus* sec. Girón and Franz (2010) INT ⊊ AND OST ⊋ *Apodrosus* sec. Marshall (1922).
G	Features that define *Apodrosus* according to Girón and Franz (2010). Only one character (21) represents a unique and unreversed synapomorphy, at least within the narrow context of this revision.
H	Example of a reversed synapomorphy for the genus *Apodrosus*; character 13, state 1: presence of a median posterior fovea on sternum VII, secondarily lost ("absent") in *A. epipolevatus*.
I	Example of a homoplasious character; character 17, state 1: tegminal plate of male terminalia complete, convergently present in the outgroup taxa *P. conicus* and *P. peninsularis*, as well as in *Apodrosus* where it is unreversed.
J	Example of repeatedly reversed character; character 22, state 1: presence of a projection on the cornu of the spermatheca, with an inferred phylogenetic sequence (root) 0 → 1 → 0 → 1 (*A. wolcotti*).
K	Example of a homoplasious character, character 12: elytral strial interval X slightly (1) or strongly (2) produced, convergently present in three clades within *Apodrosus*, with a secondary transformation in the *A. epipolevatus-A. wolcotti* clade.
L	Example of a monophyletic, unranked and unnamed species group within *Apodrosus* ("*A. eximius-A. empherefasciatus* clade").
M	Only *A. argentatus* and *A. wolcotti* were known prior to the 2010 revision. The number of species has increased more than five-fold. The two previously known species are no longer considered sister taxa, but instead are nested within different clades. These two species are maintained as valid; their diagnoses (intensional) and specimen material (ostensive; including distribution ranges) have been refined and updated. The type specimens were not seen, nor are they explicitly cited in the 2010 publication.
N	*Apodrosus quisqueyanus* is one of 11 new species added to the genus. This species is a new ostensive element of the updated genus circumscription, with property-based and type specimen-based definitions of its own.

52

lead to radical realignments even of higher-level categories (e.g. Cavalier-Smith, 2004). This iterative inferential process differs dramatically from the logic-driven, stipulative ontology building that is prevalent in the information sciences (Baader et al., 2004). The epistemological challenge of generating reliable statements of homology often compromises our ability to produce stable taxonomies and develop ontologies based on them.

The sheer magnitude of the task of building a full-blown ontology differentiates taxonomy from other domains. Large portions of the tree of life remain insufficiently explored, particularly in megadiverse lineages such as arthropods (5-10 million species estimated; Ødegaard, 2000), fungi (up to 1.5 million species estimated; Schmit and Mueller, 2007), and microbial organisms (up to 10 million species estimated; Sogin et al., 2006; though see Fraser et al., 2009, for a discussion of problems related to species delimitation in bacteria). Some 18,500 new species were added to the global count in the year 2007 alone (IISE, 2009). With only 1.8 million species named – likely less than 10% of the total species richness on Earth – the task of completing the global inventory will take several additional centuries (Wheeler, 2004). Each species contains vast amounts of phylogenetic information that will require expert analysis to achieve a reliable classification. Discoveries of "missing link" fossils may lead to rearrangements of the sequence of deeper splits in the hierarchy (e.g. Franzen et al., 2009). On the other hand, working solutions for smaller and well known groups such as mammals appear feasible (~ 5420 species and 37,400 synonyms; Wilson and Reeder, 2005).

II. ENTITIES WITH EVOLVING PROPERTIES

Because biological taxonomies organize the products of evolutionary history, they have to accommodate certain phenomena that characterize species and higher-level taxa, such as the evolvability of traits (Boyd, 1999). Two of the most critical concepts related to the generation of phylogenies are (1) homology, the similarity of traits resulting from common ancestry (see above); and (2) *homoplasy,* the similarity of traits resulting from convergence or reversal (Schuh and Brower, 2009). According to Hennig (1966), natural taxa are characterized by synapomorphies, i.e. derived

homologous features whose unique origin in the tree of life has been inductively corroborated. The silk-spinning organs ("spinnerets") that characterize all spiders are a textbook example of a synapomorphy.

Synapomorphies may be unreversed or reversed within the taxon they define. In the latter case, a feature that defines the taxon as a whole, and presumably evolved in its ancestor, has subsequently been modified at the genetic level and is phenotypically absent ("lost") in one or more of the taxon's younger subgroups. In other words, the defining feature is *not* obviously present in all descendants of the ancestor. Instead, its presence – manifested in a modified state at the molecular level – must be *inferred* on the basis of the overall tree structure (see also Fig. 2; Table 1). This is how we can accurately classify the phenotypically limb- and digit-less snakes as members of the natural taxon Tetrapoda. While this sort of classificatory practice poses no deep problems for human recognition and communication, a feature that is phenotypically absent (though inferred as present in an altered genotypic state) in a subordinate member of the superordinate class defined by it, may be difficult to represent in the language of description logics. At the very least, the condition of *transitivity* of properties from higher to lower level members in the hierarchy is violated.[6]

Convergent properties present additional problems. A shared property that is phenotypically the "same", or even rooted in the "same" transformations at the molecular level, is *not* considered homologous if phylogeny indicates an independent origin in two taxa with no recent common ancestor. To provide an example, both flies and twisted-wing parasites have one pair of their thoracic wings modified into *halteres,* which are stalked and terminally knob-like structures that function as balancing organs during flight. In flies these structures are attached to the third thoracic segment, whereas in the twisted-wing parasites they occur on the second thoracic segment. Whether or not these structures are considered homologous depends in part on the outcome of future phylogenetic studies. If it turns out that flies

---

[6] Molecular phylogeneticists frequently represent homology statements in a probabilistic framework (Nielsen, 2002; Sober, 2002), which constitutes an additional representation challenge (J. Felsenstein, pers. comm.).

and twisted-wing parasites are sister taxa (Whiting, 1998), then they may jointly be named "Halteria" and defined by the homologous presence of halteres. In the opposite case (Bonneton et al., 2006), "halteres" are no longer a synapomorphy of a natural taxon. Instead we would have to take into account the *phylogenetic contextuality* of this feature – (1) "halteres of flies" and (2) "halteres of twisted-wing parasites" – and use each descriptor separately to characterize the respective groups.

The degree of homoplasy of a property across the tree of life frequently depends on its descriptive precision (Wenzel and Carpenter, 1994; Proctor, 1996; Rieppel, 2007; Franz and Engel, 2010). In the aforementioned example, "halteres" is a rather specific term that may ultimately refer to a homologous structure with a single evolutionary origin. The term is at the lower end of the range of levels of homoplasy commonly used in taxonomy. At the other end, we observe terms referring to broadly delimited properties, for instance "petal color red" or "ventral side of prothoracic tibia with a row of triangular teeth". Such widely circumscribed properties may have hundreds of independent origins in the evolution of angiosperms and insects, respectively. Within a particular lower-level group, "petal color red" or "presence of a protibial row of teeth" might well map onto a homologous state. When used at more inclusive levels, however, these descriptors will become increasingly homoplasious and phylogenetically uninformative, picking out sets of taxa that have no recent common ancestor. Thus the referential validity of broadly defined phenotypic terms fundamentally depends on a precise specification of the phylogenetic context.

In spite of the above, it is common practice to *reuse* broadly defined phenotypic terms without making the phylogenetic context explicit enough ("eyes globular", "hind legs saltatorial", "wingless", etc.; see also Fig. 3). The result is a phylogenetic underdetermined and multi-referential terminology. This problem also affects the utility of ontologies being defined for morphological structures of plants (Avraham et al., 2008), fish (Dahdul et al., 2010), spiders (Ramírez et al., 2007), wasps (Mikó and Deans, 2009), and other taxonomic groups (Smith et al., 2007). Even though each of these controlled vocabularies includes many taxon-specific terms, the terms are not consistently embedded in a phylogenetic

context that recognizes convergent occurrences in the tree of life. They are perhaps best thought of as a powerful set of phenotype-categorizing metadata (Vogt et al., 2009). As such, they stand to make great contributions to the standardization of taxonomic descriptions (Vogt, 2009), but will not automatically facilitate ontological reasoning within a phylogenetic framework. As argued above, any morphological term that is not mapped to a unique and unreversed property in the tree of life must be further annotated with a taxonomic delimiter ("halteres *of flies* ").[7] Only then does the term pass on from the strictly diagnostic to the phylogenetic language realm. The need for additional, context-specifying qualifiers for properties is much less prevalent in other ontological networks.

III. INDIVIDUAL- AND CLASS-LIKE COMPONENTS

The philosophical literature is testimony to a longstanding debate as to whether species and higher-level taxa are individuals or classes (= natural kinds). This discussion permeates Schulz et al.'s (2008) ontologies and motivates their introduction of inherent taxon *qualities* (see above). But beyond the development of ontologies, the individuals-versus-kinds debate has had little impact on the practice of classifying taxa (Keller et al., 2003; though see de Queiroz and Gauthier, 1992). As summarized by Brigandt (2009: 78, 95):

> *[…] a species or a higher taxon can be construed both as an individual and a natural kind, i.e. both views are metaphysically compatible. Yet one conceptualization can be pragmatically preferable depending on the epistemic considerations that are in play in a certain scientific context. Taxa are best construed as natural kinds when they are viewed as taxonomic units, while it is preferable to view taxa as individuals when they are conceived of as units of evolutionary change. […] The upshot of my discussion for the individualism vs. kinds debate is that the relevant question is not so much into which metaphysical category species and higher taxa fall, but how biological accounts of taxa (such as species concepts) underwrite*

---

[7] One reviewer stated that OWL allows such context specification through addition of a sufficient condition; and furthermore, that morphological ontologies are singled out unfairly here for not performing a service they were never designed for. Both objections are valid to a degree, though neither refutes the point that standard taxonomic practice and its ontological implementation are poorly matched up with the need to support phylogenetic inferences.

Fig. 3. Example of a full-blown taxonomic representation of a species of damselfish, *Chromis circumaurea* sec. Pyle et al . (2008: 15), including (A) intensional components (diagnosis, description), ostensive components (type specimens), as well as (B) links to images, DNA data, and globally unique identifiers for individual specimens (cf. Hyam, 2009; Page, 2009). Reproduced with permission of the authors and journal.

Chromis circumaurea, **new species**
urn:lsid:zoobank.org:act:8ADC4817-8F1C-4C88-8B8A-5372A84CAEC9

Gold-rim Chromis

(Figs. 3a –3c, Table 4; Morphbank[105]; GenBank[106]; Barcode[107])

A

Holotype. BPBM 40836[108] (98.2 mm SL), Caroline Islands; Yap, S end; "Magic Kingdom" (9°26'3.41"N, 138°2'5.96"E): among boulders on sloping shelf above deep drop-off, 98–100 m, hand net, R.L. Pyle and B.D. Greene, 20 April 2007 [PCMB 3080[109]].

  Paratypes. BMNH 2007.10.31.3[110] (102.4 mm SL) [PCMB 3081[111]]. CAS 225757[112] (97.6 mm SL) [PCMB 3078[113]]. MNHN 2007-1924[114] (92.5 mm SL) [PCMB 3076[115]]. USNM 391138[116] (94.2 mm SL) [PCMB 3077[117]]. WAM P.32900-001[118] (96.6 mm SL) [PCMB 3079[119]]. All with same data as holotype.

  Diagnosis. Dorsal rays XIV,12–13 (usually 13); anal rays II, 13–14 (usually 13); pectoral rays 18–19; spiniform caudal rays 3; tubed lateral-line scales 16–17; gill rakers 6–7+20–21 (total 26–27); body depth 1.68–1.86 in SL; color when fresh mahogany brown with bright yellow distally on spinous portion of dorsal fin; soft portion of dorsal fin, caudal fin, and anal fin bright yellow.

  Description. Dorsal rays XIV,13 (12 in one paratype); anal rays II,13 (14 in one paratype); all dorsal and anal rays branched, the last to base in some specimens; pectoral rays 19 (18–19), the upper 2 and lowermost unbranched; pelvic rays I,5; principal caudal rays 8+7=15; upper and lower procurrent caudal rays 5, the anterior 3 spiniform, the posterior 2 segmented and unbranched; tubed lateral-line scales 16|17 (16–17); posterior midlateral scales with a pore or deep pit 8 (5–8); scales above dorsal fin to origin of dorsal fin 3.5 (3–3.5); scales below lateral line to origin of anal fin 10 (9–10.5); gill rakers 6+21=27 (6–7+20–21= 26–27); surpaneural (predorsal) bones 3; vertebrae 12+13.

B

105.   http://www.morphbank.net/Show/?id=197038
106.   http://www.ncbi.nlm.nih.gov/sites/entrez?term=Chromis%20circumaurea&cmd=Search&db=nuccore
107.   http://www.barcodinglife.org/views/taxbrowser.php?taxon=Chromis+circumaurea
108.   http://zoobank.org/urn:lsid:zoobank.org:specimen:AC204B49-93B7-4BEE-890B-7BF07C1EF592
109.   http://nsdb.bishopmuseum.org/urn:lsid:bishopmuseum.org:bioobject:F8DCA388-8031-4856-ACB0-CE7D1149FCC1
110.   http://zoobank.org/urn:lsid:zoobank.org:specimen:03C05B9B-816B-4904-AF54-AF55FF33CA83

classifications and generalizations, shed light on the unity of taxa across time, and permit explaining their ability to undergo change a s a unit – all of which are epistemic issues.

Taxonomists have adopted a *hybrid a pproach* to circumscribe taxa, i.e. one that uses both *intensional* (property-based) and *ostensive* (member-based) components (Fig. 2; Table 1). This practice is ingrained in the Linnaean tradition (Farber, 1976; Stevens, 1984), where species definitions are fixated by the combination of a verbal diagnosis (intensional, often accompanied by illustrations) *and* the designation of a type specimen (ostensive).

The practice of pointing to a type is especially critical at the species level (Fig. 3). Taxonomic disagreements and nomenclatural synonymies are often resolved in direct reference to the identity of type specimens (e.g. Gardner and Hayssen, 2004). Typification is also mandatory at the genus level, but becomes increasingly less prevalent as one climbs up the hierarchy to the level of family, order, class, etc. While it is common to list *all* examined specimens when defining a new species (e.g., Franz, 2010), there is virtually no use in designating a specimen to typify megadiverse lineages such as the class Insecta. In practice the latter is sufficiently well defined by listing a set of diagnostic properties or synapomorphies (Grimaldi and Engel, 2005).

Thus, we observe a gradual shift from ostensive to intensional components in accordance with the inclusiveness of the taxon being defined. This convention matches up well with human cognitive abilities and inference needs (Brigandt, 2009; Franz, 2009). For instance, it is not necessary for humans to examine specimens of every species of the scarab beetle superfamily Scarabaeoidea in order to reliably recognize them as such. A generic illustration of the synapomorphic, asymmetrically lamellate antennal club is sufficient for this purpose. Similarly, two experts talking about the weevil genus *Perelleschus* Wibmer & O'Brien may understand each other even though each has only seen specimens from Central and South America, respectively, which share no common species (Franz and O'Brien, 2001). Our cognitive tendency to shift towards "loosely typified" intensional definitions explains why regional and taxonomic sampling biases are acceptable in higher-level

phylogenetic analyses. These definitions have greater predictive value and are especially useful for making wide-reaching inferences about the identity of a taxonomic group – past, present, and future, yet their referential precision is compromised by increasing levels of homoplasy and evolutionary transformation (Section II). Ostensive definitions are more accurate but offer few inferential benefits beyond specific identification of a set of specimens or taxa.[8]

Schulz et al. (2008) are highly responsive to the individuals-versus-kinds problem (see also Gangemi et al., 2001). While recognizing the complex interaction of ostension and intension in taxonomy, the authors discard several of their initial proposals; viz. taxa as meta-properties, hierarchies, and populations. Instead, they propose to represent taxa as *qualities;* in the sense that an individual organism or part thereof has an inherent quality of pertaining to a taxon (individual → species). That taxon, in turn, has the quality of pertaining to a higher-level taxon (species → genus, etc.). The addition of *quality regions* seemingly offers a workable transition from the realm of particulars to universals. However, this solution leads to a sort of "molecular essentialism" where taxon-level qualities can inhere in an isolated sequence of nucleic acids (P. E. Midford, pers. comm.). Moreover, the approach ignores why taxonomists single out specific properties to characterize taxa within a larger lineage (Hennig, 1966; Wheeler and Meier, 2000), and fails to recognize that taxonomists use utilize ostensive and intensional elements flexibly and inconsistently at varying levels (e.g., Sereno, 2005; Franz and Peet, 2009; Schuh and Brower, 2009). Therefore the taxa-as-qualities proposal remains too simplistic in comparison to actual practice. In the end, any approach that decides one way or the other with regards to the purported individual/class dichotomy falls short of the reasoning powers that humans derive from hybrid definitions of taxa.

Sections IV and V – Special Linguistic Conventions
IV. LINNAEAN RANKS

Linnaeus (1758) advocated the strict use of ranks for taxonomic names – a convention that has

---

[8] From an ontological viewpoint it is even arguable whether preserved specimens count as members of their respective taxa, given that they have lost key molecular or behavioral (P. E. Midford, pers. comm.).

now persevered for more than 250 years (Schuh, 2003). Particularly names for mid- to higher-level taxa incorporate standardized terminations indicating their rank. In zoology, for instance, the names of tribes end with *-ini,* subfamilies end with *-inae,* families end with *-idae,* and so on. The botanical nomenclature has at least 13 fixed terminations for ranks up to the level of division (*-phyta, -mycota*) (ICBN, 2006). Although the number of ranks is in principle indeterminate, the nomenclatural Codes only regulate the application of endings for a relatively small number of ranks, not reaching beyond the level of family in zoology (ICZN, 1999). But the use of more finely tiered ranks is pervasive in taxonomy, where prefixes such as super-, sub-, and infra- provide additional levels of resolution to accommodate increasingly more bifurcated trees and refined phylogenetic classifications (e.g. McKenna and Bell, 1997). A single classification may include (1) ranked names (i) with or (ii) without standardized endings; (2) informal names that map onto unranked sections of the hierarchy (e.g. "paleoherbs"; cf. Nixon and Carpenter, 2000); and (3) sections that are not named at all (unnamed clades, taxa of uncertain position, etc.).

Adding a rank ending to a taxonomic name might seem nothing more than an arbitrary convention with limited significance for ontological reasoning. Yet it is precisely this convention which allows humans to make countless implicit and inter-subjectively reliable inferences about taxonomic relationships without necessarily having to visualize a reference tree. Many mid- to higher-level taxa have conspicuous and well conserved synapomorphies. The coupling of the features with a ranked name and standardized ending reinforces a mental association between them. The cognitive pay-offs are aptly characterized in this example by Platnick (2001: 8-9; see also Platnick, 2009):

> *I was wandering around John Murphy's garden out in Hampton, and came across a nice jumping spider. Now, jumping spiders, the family Salticidae, are probably the easiest of all spider families to recognize. With their large anterior median eyes, their excellent vision, the often highly exuberant and ornamented morphology that males use in their elaborate courtship displays, and their prowess at jumping on prey several body-lengths away, salticids are quite*

> *distinctive. [...] If you visit the [World Spider Catalog] site, you'll find a summary table that shows, for each of the 109 currently recognized spider families, the numbers of currently valid genera and species, including, at the very end of the list, the salticids, with 4,834 species. Using the Linnaean hierarchy, when I identified the spider in John's garden as a salticid, I was asserting that John's spider is more closely related to any single species currently included within the Salticidae than it is to any single species that is currently excluded from that family. In other words, if my identification, and the current classification, are both correct, then John's spider is more closely related to salticid species #1 than it is to any of the 32,752 spider species currently excluded from the Salticidae. It is also more closely related to salticid species #2 than it is to any non-salticid spider. So, assuming that the spider from John's garden belongs to one of the currently known 4,834 salticid species (and this being England, that's certainly a fair assumption), then my identification enables 4,833 (other salticids) times 32,752 (non-salticids) three-taxon statements. So by placing the animal as a salticid, the current Linnaean hierarchy allows me to make 158,290,416 three-taxon statements about it, within spiders alone. If I were to expand the arena to include all arthropods, or all life, the number of implied three-taxon statements would, for all practical purposes, approach one third of infinity – the other two-thirds would be prohibited. That's none too shabby, for a single word – Salticidae (admittedly, in a context provided – solely – by the Linnaean hierarchy, and the mutual exclusivity of equally ranked names it requires).*

Even if we concede that Linnaean ranks alone are not sufficient to convey the level of speaker expertise portrayed in this example, it is clear that using ranks has immense inferential advantages. Put simply, two taxa that have the same rank ending within a single coherent classification cannot share any subordinate members (see also Thau et al., 2009). This perfect nestedness allows humans to make countless accurate inferences about the placement of subordinate members into these non-overlapping taxa – Platnick's (2001) three-taxon statements – without having to memorize their diagnostic feature or exact position in the overall hierarchy. In addition, and despite the fact that rank assignments are not evolutionarily comparable across the tree of life (cf. Avise and Johns, 1999), many taxon/rank

couplings *do* remain stable enough so that relevant biological phenomena become tied them over time (e.g., Salticidae ↔ large anterior median eyes, prowess at jumping, elaborate courtship). Students of biology often begin by recognizing orders, then families, and later on lower ranked entities such as genera and species. At each level they establish meaningful cognitive links between ranked names and biological traits; e.g. members of the Formicidae (ants) are highly eusocial, the workers are female and wingless, etc. Whenever (minor) rank changes occur, it is relatively easy for humans to adjust by making a limited number of rank reconnections (e.g. subfamily [-*inae*] → tribe [-*ini*]) while keeping much of the learned name/information association intact.

Linnaean ranks and standardized endings have no obvious match in conventional ontologies. To address this issue, Schulz et al. (2008) created a secondary hierarchy of rank classes that interfaces with the primary taxon quality hierarchy. However, as the authors concede themselves (p. i320): "the meaning of the taxonomic rank classes […] is somewhat counterintuitive, since every instance of *SpeciesQuality* is also an instance of *GenusQuality* and so on. They are, therefore, not suited to comprehensively represent the meaning of Species as disjoint from Genus, Kingdom, etc." Other attempts to incorporate ranks into ontologies have opted to render them "ontologically weak"; treating a taxon as an *instance_of* its rank as opposed to the stronger *is_a* relationship which would give ranks the status of meta-classes and associated reasoning powers (Dahdul et al., 2010; P. E. Midford, pers. comm.). In short, translating the inferential benefits of ranks into the ontological realm has so far remained elusive.

V. NOMENCLATURAL AND TAXONOMIC LEGACY

Taxonomy is bound in its use of names by a legacy that reaches back some 250 years to Linnaeus' (1758) *Systema Naturae.* Many taxon names and definitions were first published in that work. Linnaeus (1758) also laid out a set of rules for naming taxa, including his binomial system and advocacy of ranks. In zoological taxonomy in particular, these rules were expanded by Strickland et al. (1843) who introduced the Law of Priority and other requirements regarding the typification of names (Farber, 1976). These efforts have gradually evolved into the current nomenclatural

Codes (e.g., ICZN, 1999). The process of amending the Codes continues in response to new threats to the stability of names and associated information (e.g., ICZN, 2008).

Application of the rules of nomenclature for nearly 250 years has transformed the taxonomic literature into a continuous chain of publications with quasi-legal status (Minelli, 2003). Any new publication must recognize relevant nomenclatural and taxonomic precedents. Indeed, new taxonomies are mostly communicated through existing names whose meanings were established in earlier publications. Often these meanings are revised, expanded, or contracted in complex ways. As taxonomic revisions accumulate and supersede each other over time, they tend to create a network of many-to-many relationships among valid names, invalid synonyms, and past and present meanings (Koperski et al., 2000; Geoffroy and Berendsohn, 2003; Franz, 2005a; Kennedy et al., 2005; Franz et al., 2008). The situation is compounded by the fact that nomenclatural and full-blown taxonomic relationships are established in different and frequently non-congruent ways; the former being determined strictly on the basis of the identity of *type specim ens,* whereas the latter involve comparison of *diagnostic features* and other kinds of phylogenetic information. The trajectories of nomenclatural and taxonomic relationships among names are therefore semi-independent and must be modeled separately to record partial name/meaning disjunctions over time (Koperski et al., 2000; Kennedy et al., 2005; Franz and Peet, 2009).

The history of taxonomy is littered not only with an indelible record of nomenclatural relationships but with thousands of published taxonomies that are partially incomplete, outdated, or of questionable quality and validity. Nevertheless, some cross-section of this body of work represents our best present-day knowledge of nature's hierarchy (cf. Maddison et al., 2007). Many old classifications have been fully replaced by more recent revisions. Yet even the least regarded publications (cf. Jäch, 2006) are part of taxonomy's enduring ledger and must be linked in some ways to a more widely accepted view. It is generally not permissible in taxonomy to purge low quality work from the record or to ignore names coined in obscure treatments (cf. Godfray, 2002).

In this sense, taxonomy presents legacy integration difficulties that go beyond those of more conventional ontologies. While it is considered best practice in any domain to link a new ontology to a relevant predecessor (Euzenat and Shvaiko, 2007; Smith et al., 2007; Sahoo et al., 2008), there are no quasi-legal requirements to do so. No laws of priority or rules for typifying classes exist that, if violated, would render the new ontology invalid. There are also no clear analogies to the consistent separation of type-based nomenclatural versus full-blown taxonomic relationships among the elements of succeeding ontologies. Lastly, it is permissible in most domains to ignore a previous ontology if it no longer has any utility. The result is a more uniform framework for reasoning that is not overly compromised by idiosyncratic entities and relationships of the past.

Sections VI and VII – Alignment Challenges
VI. STATIC ALIGNMENTS – CORE *VERSUS* PERIPHERAL INFORMATION

For reasons given in Section I, no team of authors is capable of publishing a full-blown taxonomy that spans across the entire tree of life. To the extent that such "complete" taxonomies *are* available, they represent compilations of multiple treatments published on subsections of the tree (Scoble, 2004). They tend to have a limited amount of information associated with each taxon name (e.g.,[9]), and typically provide no information on types and synapomorphies, or even links to sources in the primary taxonomic literature. This means that all published classifications are somehow are incomplete, due either to insufficient breadth (not all taxa covered) or depth (not all information provided of the covered taxa), or both.

New taxonomic contributions must adopt a piece-meal approach, focusing on select taxa and sources of data; e.g., morphological traits and/or molecular sequence information. Beyond this *core* taxonomic focus, each new treatment usually connects to *peripheral* taxonomic information stemming from relevant predecessors. To provide an example, the most recent revision of the weevil genus *Cotithene* Voss (Franz, 2008) contains a near-complete account of core taxonomic information on the genus itself and all eight constituent species; including differential diagnoses and detailed lists of specimens designated to represent each taxon. On the other hand, this work makes only peripheral and implicit statements about the taxonomy of related genera, without listing specimens or even naming all species per genus. Mostly there are pointers to other treatments which contain this information. In contrast, a separately published genus-level reclassification of the tribe which contains *Cotithene* offers much less taxonomic information on this genus in particular (Franz, 2006). Instead the focus is on inferring synapomorphies that define monophyletic groups within the tribe. Beyond this narrow focus there is only minimal information on other tribes from which or into which certain genera are transferred (Franz, 2009). Thus the "semantic joints" to non-focal data remain vague (see also Fig. 2; Table 1).

The piece-meal nature of taxonomic publications and incurred differential focus on core versus peripheral information result in severe challenges for ontology building. Specifically, if the goal is to represent the *entire* tree of life based on both intensional and ostensive data, then information from many independently published sources must be linked together to obtain just a single static perspective. Primarily ostensive classifications of catalogues (e.g. Alonso-Zarazaga and Lyal, 1999) would have to be integrated with exemplary phylogenetic studies (cf. Prendini, 2001) so that names and concepts listed in the former may be further defined through synapomorphies. Phylogenetic mid-level analyses must be linked to species descriptions and specimen data provided in taxonomic revisions, and so on.

Considering the semantic complexity of the information to be integrated, the goal of a single full-blown ontology of the tree of life becomes almost as difficult as the underlying research itself. Most critically, the integration process must involve *third-party expert assessments* in order to connect core results from multiple publications at their peripheral "edges" and thereby construct a contiguous information-rich network. In light of the inherent vagueness of these "edges", the assessments require expertise of the historical and regional context in which each publication was produced. Different expert teams may propose different ways in which to integrate such works.

---

[9] http://www.catalogueoflife.org/.

Therefore the challenge of developing a single tree of life ontology rapidly devolves into making intersubjective assessments of concept equivalence among partial ontologies. In practice this translates into matching taxonomies at their edges using taxonomic concept relationships. This challenge is further discussed in the next section.

VII. DYNAMIC ALIGNMENTS

As seen above, the assembly of a full-blown static ontology for taxonomy and dynamic integration of alternative taxonomies both require some measure of ontology alignment. Thau et al. (2009) provide an overview of the logical and computational challenges involved in optimizing this process. Here we will concentrate on the *semantic* challenges posed by the initial expert alignments.

Koperski et al. (2000) were the first to use a vocabulary of five terms derived from set theory in order to align two separately published taxonomic concepts C_1 and C_2 with each other; as follows (symbols according to Thau et al., 2009): (1) congruence ($C_1 \equiv C_2$), (2) proper inclusion ($C_1 \subsetneqq C_2$), (3) proper inverse inclusion ($C_1 \supsetneqq C_2$), (4) partial overlap ($C_1 \oplus C_2$), and (5) exclusion (C_1 ! C_2). The terms have since been employed sporadically by other authors (e.g., Gradstein et al., 2001; Güntsch et al., 2003; Kennedy et al., 2006; Weakley, 2006; Graham and Kennedy, 2007; Craig and Kennedy, 2008; Franz et al., 2008; Krings, 2008). Franz and Peet (2009) subsequently showed that this vocabulary must be expanded to account for different relationships based on whether the intensional or ostensive subcomponents of two concepts are compared. An example for such a scenario is given in Section III: two experts working in Central America (C_1) and South America (C_2) agree on the diagnosis of *Perelleschus* based on jointly recognized synapomorphic features, but each lists a mutually exclusive set of species in the corresponding regional treatment of the genus. Accordingly, the intensional alignment of the two concepts is $C_1 \equiv C_2$ whereas the ostensive alignment is C_1 ! C_2. The intensional congruence indicates that the concepts are in agreement as to what species generally belong to *Perelleschus* – past, present, and future. The ostensive exclusion, in turn, reflects the fact that the two experts happen to work with a regionally biased and non-overlapping set of species. Only the combined alignments capture how the two concepts relate to each other with in terms of their predictive content and explicitly included members.

Franz and Peet (2009) showed how the expanded vocabulary can be applied to express (1) how taxonomic (C) and nomenclatural (N) relationships are interconnected (e.g. $C_1 \equiv C_2$ *AND* N_1 is a heterotypic synonym for N_2); (2) whether there is uncertainty in an alignment (e.g. $C_1 \equiv C_2$ *OR* $C_1 \subsetneqq C_2$); (3) how to negate a relationship (e.g. C_1 *NOT* $\equiv C_2$); (4) if there is a way to reconcile two classifications through addition or subtraction of concepts on one side (e.g. $C_1 \equiv C_2 + C_3$); or (5) whether two authors stipulate different feature-based definitions (INT) of a taxon even though they examined the same set (OST) of subordinate members (e.g. C_1 *INT* $\oplus C_2$ *AND* C_1 *OST* $\equiv C_2$). Depending on the richness of the source data, such combined alignments can represent partial concept matches and implicit judgments of errors in the aligned taxonomies. These capabilities are not yet available in an ontological environment.

Even though alignments are essential for connecting concepts occurring in alternative taxonomies, they cannot replace expert judgment as to whether certain instances of congruence or non-congruence are *significant* with respect to a particular integration task. One might ask, for instance, whether two feature-based definitions of a taxon pick out a sufficiently similar set of subordinate members when each is applied outside its geographic context. The answer will vary according to case-specific standards of "sufficiently similar": what is similar enough for an ecological study may not suffice for an analysis of adaptive radiation, and so on (cf. Peterson and Navarro-Sigüenza, 1999). In this regard it is difficult to image that taxonomic concept alignments will ever become fully automated (see also Geoffroy and Güntsch, 2003; Thau and Ludäscher, 2007; Thau et al., 2009).

SUMMARY AND OUTLOOK

As reviewed above, taxonomic practice is bound by a range of epistemological constraints and linguistic conventions that run *orthogonal* to the logical background from which ontological entities and relationships originate (Baader et al., 2004). The enormous challenge of reconstructing

the tree of life compromises the goal of creating an ontology that is comprehensive and reliable enough to permit reasoning about taxa and their properties. The inherent evolvability of taxa poses genuine ontological challenges in the philosophical sense of the term, because it undermines the individual/class dichotomy and associated reasoning capabilities that sustain conventional ontologies (cf. Schulz et al. 2008). In addition, the idiosyncratic yet indelible 250-year legacy of nomenclatural and taxonomic changes has resulted in an immense network of names and concepts that can only be aligned with an expanded vocabulary whose application requires expert input. For better or worse, we should recognize that much of taxonomy's cumulative body of work is not well aligned with the requirements for ontological representation and reasoning.

Our analysis could explain why the development of ontologies for taxonomy has lagged behind in comparison to other disciplines. To the extent that taxonomic research is still focused on acquiring and interpreting *primary data* to generate a natural classification, the prospects of reliably applying these data in a *metadata*-driven ontological framework will remain limited. Many ongoing projects in taxonomy are motivated precisely by the insight that the existing "ontology" for a particular lineage – i.e., the previously established classification – is incomplete or at least partly wrong (Vane-Wright, 2003). This is not an optimal foundation for positing stable, logic-based definitions and relationships among taxa.

In light of these limitations, we suggest that a full-blown and static representation of taxonomic information for large portions of the tree of life is *not* the most fruitful path to advance research along the taxonomy/ontology interface. Instead, our efforts should concentrate (1) on representing strictly nomenclatural relationships (Huber and Klump, 2009), and (2) on improving ontology-driven vocabularies and algorithms for producing alignments between multiple taxonomies (Franz et al., 2009; Thau et al., 2009). Taxonomic experts stand to benefit from each of these developments because they will facilitate the identification and integration of taxonomic legacy information. Both types of services may become stepping stones towards a dynamic ontological network representing the products of taxonomic research.

Conversely, full-blown and static ontologies for major organismal lineages will neither serve taxonomy nor its users in the long term. The utility of such ontologies is limited to smaller groups where the assumption of taxonomic stability is reasonable (cf. Smith et al., 2007; Dahdul et al., 2010; Mungall et al., 2010). In either case, researchers should for the most part refrain from resolving "deep questions" about the ontological nature of taxa because the answers will vary according to the preferred inferential context (Brigandt, 2009).

We furthermore suggest that the prospects of utilizing ontological reasoning in taxonomy will largely depend on the ability of the expert community to present phylogenies and classifications in ways that are more compatible with ontological principles than concurrent practice. Minimally, this means: (1) adopting strict conventions for linking new core taxonomic information to (provisionally accepted) peripheral information so that the relevant context of the new contribution is fully defined; (2) using lineage-specific phenotype ontologies for taxonomic descriptions while specifying the phylogenetic context of the descriptive terms in use (cf. Ramírez et al., 2007; Mikó and Deans, 2009; Dahdul et al., 2010); (3) presenting all nomenclatural and taxonomic novelties in an ontology-compatible format, including intensional and ostensive definitions (see also Sereno, 2009); and (4) providing intensional and ostensive alignments to entities in relevant preceding taxonomies (Franz and Peet, 2009; Thau et al., 2009).

The implementation of these practices will require a wider acceptance of the taxonomic concept approach (Berendsohn, 1995; Koperski et al., 2000; Franz et al., 2008). In particular, taxonomists will have to become more disciplined in recognizing acts of *authoring* or *citing* concepts as well as *identifying* specimens or subordinate concepts to them (Franz and Peet, 2009). Adopting this approach may also force taxonomists to have more control over the economics of maintaining a cyberinfrastructure for the publication, continuous versioning, and cross-linking of such concepts. While these goals are worthy of pursuit, they will likely remain elusive in the short term.

Neither taxonomists nor developers of ontologies should be under the illusion that full-blown ontologies for taxonomy will soon be

dynamically assembled and updated without also putting in place robust mechanisms for recognizing individual expert contributions (cf. Clark et al., 2009). Although phylogenies and classifications may represent no more than a means to an end for user communities (Dahdul et al., 2010), they constitute the primary intellectual products of taxonomists. The latter rely on signaling their authorship of these products in order to advance their academic careers. Schulz et al.'s (2008) ideal of an overarching ontology-based framework for organizing all organismal information implicitly requires that the field of taxonomy regains a more powerful role among the biological disciplines.

REFERENCES

Alonso-Zarazaga, M. A., Lyal, C. H. C. 1999. A World Catalogue of Families and Genera of Curculionoidea (Insecta: Coleoptera) Excluding Scolytidae and Platypodidae. Entomopraxis, Barcelona.

Assis, L. C. S., Brigandt, I. 2009. Homology: homeostatic property cluster kinds in systematics and evolution. Evol. Biol. 36: 248-255.

Avise, J. C., Johns, G. C. 1999. Proposal for a standardized temporal scheme of biological classification for extant species. Proc. Natl. Acad. Sci. U.S.A. 96, pp. 7358-7363.

Avraham, S., Tung, C. W., Ilic, K., Jaiswal, P., Kellogg, E. A., McCouch, S., Pujar, A., Reiser, L., Rhee, S. Y., Sachs, M. M., Schaeffer, M., Stein, L., Stevens, P., Vincent, L., Zapata, F., Ware, D. 2008. The Plant Ontology Database: a community resource for plant structure and developmental stages controlled vocabulary and annotations. Nucleic Acids Res. 36, D449-D454.

Baader, F., Horrocks, I., Sattler, U. 2004. Description logics. In: Staab, S., Studer, R. (Eds.), Handbook on Ontologies. Springer, Berlin, pp. 3-28.

Balhoff, J. P., Dahdul, W. M., Kothari, C. R., Lapp, H., Lundberg, J. G., Mabee, P. M., Midford, P. E., Westerfield, M., Vision, T. J. 2010. Phenex: ontological annotation of phenotypic diversity. PLoS ONE. (in press).

Bapteste, E., O'Malley, M. A., Beiko, R. G., Ereshefsky, M., Gogarten, J. P., Franklin-Hall, L., Lapointe, F. J., Dupré, J., Dagan, T., Boucher, Y., Martin, W. 2009. Prokaryotic evolution and the tree of life are two different things. Biol. Direct. 2009, 4:34, 1-20.

Berendsohn, W. G. 1995. The concept of "potential taxa" in databases. Taxon 44, 207-212.

Berendsohn, W. G., Döring, M., Geoffroy, M., Glück, K., Güntsch, A., Hahn, A., Kusber, W.-H., Li, J.-J., Röpert, D., Specht, F. 2003. The Berlin Taxonomic Information Model. Schrift. Veget. 39, 15-42.

Berendsohn, W. G., Geoffroy, M. 2007. Networking taxonomic concepts – uniting without "unitary-ism". In: Curry, G., Humphries, C. (Eds.), Biodiversity Databases – Techniques, Politics, and Applications. CRC Taylor & Francis, Baton Rouge, pp. 13-22.

Bonneton, F., Brunet, F. G., Kathirithamby, J., Laudet, V. 2006. The rapid divergence of the ecdysone receptor is a synapomorphy for Mecopterida that clarifies the Strepsiptera problem. Insect Mol. Biol. 15, 351-362.

Boto, L. 2009. Horizontal gene transfer in evolution: facts and challenges. Proc. Royal Soc. B, 1-9.

Bowker, G. C., Star, S. L. 1999. Sorting Things Out: Classification and Its Consequences. MIT Press, Cambridge, MA.

Boyd, R. 1999. Homeostasis, species, and higher taxa. In: Wilson, R. A. (Ed.), Species: New Interdisciplinary Essays. Bradford Books, MIT Press, Cambridge, pp. 141-185.

Boyd, R. 2000. Kinds as the "workmanship of men": realism, constructivism, and natural kinds. In: Nida-Rümelin, J. (Ed.), Rationality, Realism, Revision: Proceedings of the Third International Congress of the Society for Analytical Philosophy, September 15-18, 1997 in Munich. Walter de Gruyter, Berlin, pp. 52-89.

Brigandt, I. 2009. Natural kinds in evolution and systematics: metaphysical and epistemological considerations. Acta Biotheor. 57, 77-97.

Cavalier-Smith, T. 2004. Only six kingdoms of life. Proc. Royal Soc. Lond. B 271, 1251-1262.Clark, B. R., Godfray, H. C. J., Kitching, I. J., Mayo, S. J., Scoble, M. J. 2009. Taxonomy as an eScience. Phil. Trans. R. Soc. A 367, 953-966.

Craig, P., Kennedy, J. 2008. Concept Relationship Editor: a visual interface to support the assertion of synonymy relationships between taxonomic classifications. In: Börner, K., Gröhn, M., Park, J., Roberts, J. (Eds.), Visualization and Data Analysis 2008, Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE), Vol. 6809, January 28-29, 2008, San Jose, CA; pp. 6809.06-6809.12.

Dahdul, W. M., Lundberg, J. G., Midford, P. E., Balhoff, J. P., Lapp, H., Vision, T. J., Haendel, M. A., Westerfield, M., Mabee, P. M. 2010. The Teleost Anatomy Ontology: anatomical representation for the genomic age. Syst. Biol. 59. (in press).

Du, P., Feng, G., Flatow, J., Song, J., Holko, M., Kibbe, W. A., Lin, S. M. 2009. From disease ontology to disease-ontology lite: statistical methods to adapt a general-purpose ontology for the test of gene-ontology associations. Bioinformatics 25, i63-i68.

Ereshefsky, M. 2002. Linnaean hierarchy: vestiges of a bygone era. Phil. Sci. 69, S305-S315.

Euzenat, J., Shvaiko, P. 2007. Ontology Matching. Springer, Berlin.

Farber, P. L. 1976. The type concept in zoology in the first half of the nineteenth century. J. Hist. Biol. 9, 93-119.

Franz, N. M. 2005a. On the lack of good scientific reasons for the growing phylogeny/classification gap. Cladistics 21, 495-500.

Franz, N. M. 2005b. Outline of an explanatory account of cladistic practice. Biol. Phil. 20: 489-515.

Franz, N. M. 2006. Towards a phylogenetic system of derelomine flower weevils (Coleoptera: Curculionidae). Syst. Entomol. 31, 220-287.

Franz, N. M. 2008. Revision, phylogeny, and natural history of *Cotithene* Voss (Coleoptera: Curculionidae). Zootaxa 1782 1-33.

Franz, N. M. 2009. Letter to Linnaeus. In: Knapp, S., Wheeler, Q. D. (Eds.), Letters to Linnaeus. Linnean Society of London, London, pp. 63-74.

Franz, N. M. 2010. Revision and phylogeny of the genus *Apotomoderes* Dejean (Coleoptera: Curculionidae: Entiminae). ZooKeys 49: 33-75.

Franz, N. M., Engel, M. S. 2010. Can higher-level phylogenies of weevils explain their evolutionary success? A critical reassessment. Syst. Entomol. 35. (in press).

Franz, N. M., O'Brien, C. W. 2001. Revision and phylogeny of *Perelleschus* (Coleoptera: Curculionidae), with notes on its association with *Carludovica* (Cyclanthaceae). Trans. Am. Entom. Soc. 127, 255-287.

Franz, N. M., Peet, R. K. 2009. Towards a language for mapping relationships among taxonomic concepts. Syst. Biodiv. 7, 5-20.

Franz, N. M., Peet, R. K., Weakley, A. S. 2008. On the use of taxonomic concepts in support of biodiversity research and taxonomy. In: Wheeler, Q. D. (Ed.), The New Taxonomy. Systematics Association Special Volume Series 74. Taylor & Francis, Boca Raton, pp. 63-86.

Franzen, J. L., Gingerich, P. D., Habersetzer, J., Hurum, J. H., von Koenigswald, W., Smith, B. H. 2009. Complete primate skeleton from the Middle Eocene of Messel in Germany: morphology and paleobiology. PLoS ONE 4(5), e5723, 1-27.

Fraser, C., Alm, E. J., Polz, M. F., Spratt, B. G., Hanage, W. P. 2009. The bacterial species challenge: making sense of genetic and ecological diversity. Science 323, 741-746.

Gangemi, A., Guarino, N., Masolo, C., Oltramari, A. 2001. Understanding top-level ontological distinctions. In: Gómez Pérez, A., Gruninger, M., Stuckenschmidt, H., Uschold, M. (Eds.), Ontologies and Information Sharing. Proceedings of the IJCAI-01 Workshop on Ontologies and Information Sharing. Morgan Kaufmann, San Francisco, pp. 26-33.

Gardner, A. L., Hayssen, V. 2004. A guide to constructing and understanding synonymies for mammalian species. Mamm. Spec. 739, 1-17.

Geoffroy, M., Berendsohn, W.G. 2003. The concept problem in taxonomy: importance, components, approaches. Schrift. Veget. 39, 5-14.

Geoffroy, M., Güntsch, A. 2003. Assembling and navigating the potential taxon graph. Schrift. Veget. 39, 71-82.

Ghiselin, M. T. 1997. Metaphysics and the origin of species. State University of New York Press, Albany.

Girón, J. C., N. M. Franz, N. M. 2010. Revision and phylogeny of the genus *Apodrosus* Marshall 1922 (Coleoptera: Curculionidae: Entiminae). Insect Syst. Evol. 41. (in press).

Goble, C., Stevens, R. 2008. State of the nation in data integration for bioinformatics. J. Biomed. Inf. 41: 687-693.

Godfray, H. C. J. 2002. Challenges for taxonomy. Nature 417, 17-19.

Gradstein, S. R., Sauer, M., Koperski, M., Braun, W., Ludwig, G. 2001. TaxLink, a program for computer-assisted documentation of different circumscriptions of biological taxa. Taxon 50, 1075-1084.

Graham, M., Kennedy, J. 2007. Visual exploration of alternative taxonomies through concepts. Ecol. Inform. 2, 248-261.

Grimaldi, D., Engel, M. S. 2005. Evolution of the insects. Cambridge University Press, Cambridge.

Güntsch, A., Döring, M., Geoffroy, M., Glück, K., Li, J., Röpert, D., Specht, F., Berendsohn W. 2003. The Taxonomic Editor. Schrift. Veget. 39, 43-56.

Hennig, W. 1966. Phylogenetic Systematics. University of Illinois Press, Urbana, IL.

Hey, J. 2006. On the failure of modern species concepts. Trends Ecol. Evol. 21, 447-450.

Horridge, M., Knublauch, H., Rector, A., Stevens, R., Wroe, C. 2004. A Practical Guide to Building OWL Ontologies Using the Protégé-OWL Plugin and CO-ODE Tools Edition 1.0. University of Manchester[10] [Accessed XII-20-2009].

Huber, R., Klump, J. 2009. Charting taxonomic knowledge through ontologies and ranking algorithms. Comp. Geosci. 35, 862-868.

Hyam, R. 2009. Taxa, taxon names and globally unique identifiers in perspective. To appear in: Descriptive Taxonomy: the Foundation of Biodiversity Research. Systematics Association, Cambridge University Press, 11 pp.[11] [Accessed XII-20-2009].

ICBN – International Code of Botanical Nomenclature (Vienna Code). Regnum Vegetabile 146, 1-568.

ICZN – International Commission on Zoological Nomenclature. 1999. International Code of Zoological Nomenclature. International Trust for Zoological Nomenclature, London.

ICZN – International Commission on Zoological Nomenclature. 2008. Proposed amendment of the International Code of Zoological Nomenclature to expand and refine methods of publication. Zootaxa 1908, 57-67.

IISE – International Institute for Species Exploration. 2009. SOS – State of Observed Species[12] [Accessed XII-20-2009].

Jäch, M. A. 2006. Taxonomy and nomenclature threatened by D. Makhan. Koleop. Rund. 76, 360.

Johansson, I. 2006. Bioinformatics and biological reality. J. Biomed. Inf. 39, 274-287.

Jones, M. B., Schildhauer, M. P., Reichman, J., Bowers, S. 2006. The new bioinformatics: integrating ecological data from the gene to the biosphere. Ann. Rev. Ecol. Evol. Syst. 37, 519-544.

Keller, R. A., Boyd, R. N., Wheeler, Q. D. 2003. The illogical basis of phylogenetic nomenclature. Bot. Rev. 69, 93-110.

Kennedy, J., Hyam, R., Kukla, R., Paterson, T. 2006. A standard data model representation for taxonomic information. OMICS 10, 220-230.

Kennedy, J., Kukla, R., Raschid, L., Paterson, T. 2005. Scientific names are ambiguous as identifiers for biological taxa: their context and definition are required for accurate data integration. In: Ludäscher,B., Raschid, L. (Eds.), Data Integration in the Life Sciences: Proceedings of the Second International Workshop, San Diego, CA, USA, July 20-22. DILS 2005, LNBI 3615, pp. 80-95.

Koperski, M., Sauer, M., Braun, W., Gradstein, S. R. 2000. Referenzliste der Moose Deutschlands. Schrift. Veget. 34, 1-519.

Krings, A. 2008. Revision of *Gonolobus s.s* . (Apocynaceae: Asclepiadoideae) in the West Indies. J. Bot. Res. Inst. Texas 2: 95-138.

Lee, M. S. Y. 2003. Species concepts and species reality: salvaging a Linnaean rank. J. Evol. Biol. 16, 179-188.

Linnaeus, C. 1758. *Systema Naturae p er Regn a Tria Naturae, Secundum C lasses, Or dines, Gene ra, Species, cum Ca racteribus, Differen tiis, Synonymis, Locis.* Vol. 1. 10th Ed. L. Salvii, Holmiae (Sweden).

Mabee, P. M., Ashburner, M., Cronk, Q., Gkoutos, G. V., Haendel, M., Segerdell, E., Mungall, C., Westerfield, M. 2007. Phenotype ontologies: the bridge between genomics and evolution. Trends Ecol. Evol. 22: 345-350.

Maddison, D. R., Schulz, K.-S., Maddison, W. P. 2007. The Tree of Life web project. Zootaxa 1668: 19-40.

Madin, J. S., Bowers, S., Schildhauer, M. P., Jones, M. B. 2008. Advancing ecological research with ontologies. Trends Ecol. Evol. 23: 159-168.

Marshall, G. A. K. 1922. Some injurious Neotropical weevils (Curculionidae). Bull. Entomol. Res. 13, 59-71.

McCray, A. T. 2006. Conceptualizing the world: lessons from history. J. Biomed. Inf. 39, 267-273.

McKenna, M. C., Bell, S. K. 1997. Classification of Mammals Above the Species Level. Columbia University Press, New York, NY.

Merrill, G. H. 2010. Ontological realism: methodology or misdirection? Appl. Ontology 5, 79-108.

Midford, P. E. 2004. Ontologies for behavior. Bioinformatics 20, 3700-3701.

Mikó, I., Deans, A. R. 2009. *Masner,* a new genus of Ceraphronidae (Hymenoptera, Ceraphronoidea) described using controlled vocabularies. ZooKeys 20, 127-153.

Minelli, A. 2003. The status of taxonomic literature. Trends Ecol. Evol. 18, 75-76.

Mungall, C. J., Gkoutos, G. V., Smith, C. L., Haendel, M. A., Lewis, S. W., Ashburner, M. 2010. Integrating phenotype ontologies across multiple species. Genome Biol. 11: R2: 1-16.

Nielsen, R. 2002. Mapping mutations on phylogenies. Syst. Biol. 51: 729-739.

Nixon, K. C., Carpenter, J. M. 2000. On the other "phylogenetic systematics". Cladistics 16, 298-318.

O'Brien, C. W., Wibmer, G. J. 1982. Annotated checklist of the weevils (Curculionidae *sensu lat o*) of North America, CentralAmerica, and the West Indies (Coleoptera: Curculionoidea). Mem. Am. Entomol. Inst. 34, 1-382.

---

[10] http://www.co-ode.org/resources/tutorials/ProtegeOWLTutorial.pdf.

[11] http://www.hyam.net/blog/wp-content/uploads/2009/08/perspectives_09.pdf.

[12] http://species.asu.edu/files/IISE_SOS_2009.pdf.

Ødegaard, F. 2000. How many species of arthropods? Erwin's estimate revised. Biol. J. Linn. Soc. 71, 583-597.

Page, R. D. M. 2004. Phyloinformatics: towards a phylogenetic database. In: Wang, J. T. L., Mohammed, J. Z., Toivone, H. T. T., Sasha, D. (Eds.), Data Mining in Bioinformatics. Springer, London, pp. 219-241.

Page, R. D. M. 2006. Taxonomic names, metadata, and the Semantic Web. Biodiv. Inf. 3, 1-15.

Page, R. D. M. 2007. TBMap: a taxonomic perspective on the phylogenetic database TreeBASE. BMC Bioinformatics 8, 158, 1-11.

Page, R. D. M. 2009. bioGUID: resolving, discovering, and minting identifiers for biodiversity informatics. BMC Bioinformatics 10 (Suppl. 14): S5: 1-10.

Peterson, A. T., Navarro-Sigüenza, A. G. 1999. Alternate species concepts as bases for determining priority conservation areas. Conserv. Biol. 13, 427-431.

Platnick, N. I. 2001. From cladograms to classifications: the road to dePhylocode.[13] [Accessed XII-20-2009].

Platnick, N. I. 20099. Letter to Linnaeus. In: Knapp, S., Wheeler, Q. D. (Eds.), Letters to Linnaeus. Linnean Society of London, London, pp. 171-184.

Prendini, L. 2001. Species or supraspecific taxa as terminals in cladistic analysis? Groundplans versus exemplars revisited. Syst. Biol. 50, 290-300.

Proctor, H. C. 1996. Behavioral characters and homoplasy: perception versus practice. In: Sanderson, M. J., Hufford, L. (Eds.), Homoplasy: the Recurrence of Similarity in Evolution. Academic Press, San Diego, CA, pp. 131-149.

Pullan, M. R., Armstrong, K. E., Paterson, T., Cannon, A., Kennedy, J. B., Watson, M. F., McDonald, S., Raguenaud, C. 2005. The Prometheus Description Model: an examination of the taxonomic description-building process and its representation. Taxon 54, 751-765.

Pyle, R. L., Earle, J. L., Greene, B. D. 2008. Five new species of the damselfish genus *Chromis* (Perciformes: Labroidei: Pomacentridae) from deep coral reefs in the tropical western Pacific. Zootaxa 1671: 3-31.

de Queiroz, K., Gauthier, J. 1992. Phylogenetic taxonomy. Ann. Rev. Ecol. Syst. 23, 449-480.

Ramírez, M. J., Coddington, J. A., Maddison, W. P., Midford, P. E., Prendini, L., Miller, J., Griswold, C. E., Hormiga, G., Sierwald, P., Scharff, N., Benjamin, S. P., Wheeler, W. C. 2007. Text linking of digital images to phylogenetic data matrices using a morphological ontology. Syst. Biol. 56, 283-294.

Rieppel, O. 2007. The performance of morphological characters in broad-scale phylogenetic analyses. Biol. J. Linn. Soc. 92, 297-308.

Sahoo, S. S., Sheth, A., Henson, C. 2008. Semantic provenance for eScience: managing the deluge of scientific data, IEEE Internet Computing 12 (July/August 2009): 46-54.

Schmit, J. P., Mueller, G. M. 2007. An estimate of the lower limit of global fungal diversity. Biodivers. Conserv. 16, 99-111.

Schuh, R. T. 2003. The Linnaean system and its 250-year persistence. Bot. Rev. 69, 59-78.

Schuh, R. T., Brower, A. V. Z. 2009. Biological Systematics: Principles and Applications, 2nd Edition. Cornell University Press, Ithaca, NY.

Schulz, S., Hahn, U. 2007. Towards the ontological foundations of symbolic biological theories. Artif. Intell. Med. 39, 237-250.

Schulz, S., Johansson, I. 2007. Continua in biological systems. The Monist 90, 499-522.

Schulz, S., Stenzhorn, H., Boeber, M. 2008. The ontology of biological taxa. Bioinformatics 24, i313-i321.

Scoble, M. J. 2004. Unitary or unified taxonomy? Phil. Trans. R. Soc. Lond., Ser. B 359, 699-711.

Sereno, P. C. 2005. The logical basis of phylogenetic taxonomy. Syst. Biol. 54: 595-619.

Sereno, P. C. 2009. Comparative cladistics. Cladistics 25: 624-659.

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L. J., Eilbeck, K., Ireland, A., Mungall, C. J., The OBI Consortium, Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R. H., Shah, N., Whetzel, P. L., Lewis, S. 2007. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature Biotech. 25, 1251-1255.

Sober, E. 2002. Reconstructing the character states of ancestors: a likelihood perspective on cladistic parsimony. Monist 85: 156-176.

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M., Herndl, G. J. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc. Natl. Acad. Sci. U.S.A. 103, 12115-12120.

Stevens, P. F. 1984. Metaphors and typology in the development of botanical systematics 1690-1960, or the art of putting new wine in old bottles. Taxon 33, 169-211.

Strickland, H. E., Phillips, J., Richardson, J., Owen, R., Jenyns, L., Broderip, W. J., Henslow, J. S., Shuckard, W. E., Waterhouse, G. R., Yarrell, W., Darwin, C., Westwood, J. O. 1843. Series of propositions rendering the nomenclature of zoology uniform and permanent. In: Report of the Twelfth Meeting of the British Association for the

---

[13] http://www.systass.org/archive/events-archive/2001/platnick.pdf.

Advancement of Science held at Manchester in June 1842; 105-121.

Thau, D. 2008. Reasoning about taxonomies and articulations. In: Ph.D. '08: Proceedings of the 2008 EDBT Ph.D. Workshop. ACM, New York, NY, pp. 11-19.

Thau, D., Bowers, S, Ludäscher, B. 2008. Merging taxonomies under RCC-5 algebraic articulations. In: Conference on Information and Knowledge Management. Proceeding of the 2nd International Workshop on Ontologies and Information systems for the Semantic Web. ACM, pp. 47-54.

Thau, D., Bowers, S, Ludäscher, B. 2009. Merging sets of taxonomically organized data using concept mappings under uncertainty. In: Proceedings of the 8th International Conference on Ontologies, Databases, and the Applications of Semantics (ODBASE 2009). Lecture Notes in Computer Science. OTM 2009, pp. 1103-1120.

Thau, D., Ludäscher, B. 2007. Reasoning about taxonomies in first-order logic. Ecol. Inform. 2, 195-209.

Vogt, L. 2009. The future role of bio-ontologies for developing a general data standard in biology: chance and challenge for zoo-morphology. Zoomorphology 128, 201-217.

Vogt, L., Bartolomaeus, T., Giribet, G. 2009. The linguistic problem of morphology: structure versus homology and the standardization of morphological data. Cladistics 25, 1-25.

Vane-Wright, R. I. 2003. Indifferent philosophy versus almighty authority: on consistency, consensus and unitary taxonomy. Syst. Biodiv. 1, 3-11.

Wagner, G. P. (Ed). 2001. The Character Concept in Evolutionary Biology. Academic Press, San Diego, CA.

Weakley, A. S. 2006. Flora of the Carolinas, Virginia, and Georgia. Working Draft of January 17, 2006[14] [Accessed XII-20-2009].

Wenzel, J. W., Carpenter, J. M. 1994. Comparing methods: adaptive traits and tests of adaptation. In: Eggleton, P., Vane-Wright, R. I. (Eds.), Phylogenetics and Ecology. Academic Press, London, pp. 79-101.

Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., DiCuccio, M., Edgar, R., Federhen, S., Feolo, M., Geer, L. Y., Helmberg, W., Kapustin, Y., Khovayko, O., Landsman, D., Lipman, D. J., Madden, T. L., Maglott, D. R., Miller, V., Ostell, J., Pruitt, K. D., Schuler, G. D., Shumway, M., Sequeira, E., Sherry, S. T., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusov, R., Tatusova, T. A., Wagner, L., Yaschenko, E. 2008. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 36, D13-D21.

Wheeler, Q. D. 2004. Taxonomic triage and the poverty of phylogeny. Phil. Trans. R. Soc. Lond. B 359, 571-583.

Wheeler, Q. D., Meier, R. 2000. Species Concepts and Phylogenetic Theory: a Debate. Columbia University Press, New York, NY.

Whiting, M. F. 1998. Long-branch distraction and the Strepsiptera. Syst. Biol. 47, 134-138.

Wilson, D. E., Reeder, D. E. (Eds). 2005. Mammal Species of the World: a Taxonomic and Geographic Reference, 3rd Edition. Johns Hopkins University Press, Baltimore, MD.

Yoon, N., Rose, J. 2001. An information model for the representation of multiple biological classifications. In: Alexandrov, V. N., Dongarra, J. J., Juliano, B. A., Renner, R. S., Tan, C. J. K. (Eds.), Computational Science – ICCS 2001: International Conference, San Francisco, CA, USA. Lecture Notes in Computer Science. Springer-Verlag, Berlin, pp. 937-946.

---

[14] http://www.herbarium.unc.edu/flora.htm.