

Calculating structural complexity in phylogenies using ancestral ontologies

Martín J. Ramírez^{a,*} and Peter Michalik^b

^aMuseo Argentino de Ciencias Naturales “Bernardino Rivadavia” - CONICET, Av. Angel Gallardo 470, C1405DJR, Buenos Aires, Argentina;

^bZoologisches Institut und Museum, Ernst-Moritz-Arndt-Universität, J.-S.-Bach-Str. 11/12, D-17489, Greifswald, Germany

Accepted 20 February 2014

Abstract

Complexity is an important aspect of evolutionary biology, but there are many reasonable concepts of complexity, and its objective measurement is an elusive matter. Here we develop a simple measure of complexity based on counts of elements, incorporating the hierarchical information as represented in anatomical ontologies. Neomorphic and transformational characters are used to identify novelties and individuated morphological regions, respectively. By linking the characters to terms in an anatomical ontology a node-driven approach is implemented, where a node ontology and a complexity score are inferred from the optimization of individual characters on each ancestral or terminal node. From the atomized vector of character scorings, the anatomical ontology is used to integrate the hierarchical structure of morphology in terminals and ancestors. These node ontologies are used to calculate a measure of complexity that can be traced on phylogenetic trees and is harmonious with usual phylogenetic operations. This strategy is compared with a terminal-driven approach, in which the complexity scores are calculated only for terminals, and optimized as a continuous character on the internal nodes. These ideas are applied to a real dataset of 166 araneomorph spider species scored for 393 characters, using Spider Ontology (SPD, <https://bioportal.bioontology.org/ontologies/SPD>); complexity scores and transitions are calculated for each node and branch, respectively. This result in a distribution of transitions skewed towards simplification; the transitions in complexity have no apparent correlation with character branch lengths. The node-driven and terminal-driven estimations are generally correlated in the complexity scores, but have higher divergence in the transition values. The structure of the ontology is used to provide complexity scores for organ systems and body parts of the focal groups.

© The Willi Hennig Society 2014.

One of the most important topics in evolutionary biology is understanding the mechanistic aspects of morphological change and the origin of morphological novelties (Newman and Müller, 2000; Carroll, 2001; Moczek, 2008; Müller, 2010; Peterson and Müller, 2013). For this, bioinformatics helps with the integration of knowledge from different disciplines, such as genetics, development, comparative anatomy, and evolution. The phenotypic data contained in phylogenetic datasets combined with gene expression patterns from model organisms hold promise for the detection of candidate genes responsible for morphological changes (Mabee et al., 2007a,b, 2012; Coulet et al., 2008;

Washington et al., 2009), provided that both disciplines refer to a common anatomical reference. Anatomical ontologies have been developed for the dual purpose of serving as a common reference for interoperability between research areas (e.g. genomics and comparative anatomy), and for automatic reasoning (i.e. computing) over the structured data to produce new inferences (Wolstencroft et al., 2007; Vogt, 2009; Balhoff et al., 2010; Walls et al., 2012). How these reasoning capabilities are performed over evolutionary trees has just started to be explored (e.g. Mabee et al., 2007a; Prosdocimi et al., 2009; Gaudet et al., 2011). Here we will investigate some operations to propagate anatomical knowledge over phylogenetic trees, using ontologies as representations of the morphology of species and hypothetical ancestors, and

*Corresponding author:

E-mail address: ramirez@macn.gov.ar

phylogenetic datasets as data sources. Using those concepts we test the reasoning capabilities and accuracy of anatomical representations with the calculation of a structural complexity index on phylogenetic trees.

Measuring complexity

The origin and evolution of complexity is an important and much debated topic in biology (Valentine et al., 1994; McShea, 1996; Bell and Mooers, 1997; Aburomia et al., 2003; Freeling and Thomas, 2006; Vogel and Chothia, 2006; Kim and Caetano-Anollés, 2011). However, there are many ways to conceive complexity, and the operational aspects of measuring it have been elusive (Carroll, 2001; Donoghue and Purnell, 2005). The simplest measures are the mere counting of elements present in an organism, organ, or any other entity under comparison. More complex measures incorporate the relationships between elements (e.g. number of hierarchical levels, counts of structural and functional relationships; see McShea, 1996; for a review). Despite this diversity, measures of complexity are conceptually similar in diverse fields, such as ecology, anatomy, and information science (McElhinny et al., 2005; McShea, 1991; Zhang et al., 2010).

Here we focus on the so-called narrower view of complexity, which reduces the problem to only evaluating the number of different elements or interactions present in an organism (McShea, 1991). Under this view, complex organisms have many different types of parts or interactions, while simple ones have few types of parts or simple interactions. This approach to evaluating complexity is simple enough for applications in narrowly defined problems, and allows for quantitative testing of hypotheses (e.g. comparing morphological and genetic complexity, or complexity of different body parts; Carroll, 2001). McShea (1991) classified four types of complexity, from the combination of two basic dichotomies: object vs. process, and hierarchical vs. non-hierarchical organization. Because of the nature of our data, namely comparative morphology as expressed in phylogenetic datasets, we will only consider here the object or structural aspects of complexity, thus taking into account the number of diversified morphological elements and their hierarchical arrangement.

There are other measures of complexity based on the geometric properties of shapes or contours, typically used for simple, comparable structures in closely related species (for example, fractal dimension and Fourier harmonics; e.g. McLellan and Ender, 1998; Rowe and Arnqvist, 2012). We focus here on a more qualitative approach, allowing the treatment of evolutionary novelties and losses of structures; in our strategy, any given structure in an organism may or may not have a comparable counterpart in other organisms. We base our complexity measure on the

presence or absence of individualized structures and morphological regions. As will be shown below, finely grained characters (such as counts of modular elements, degree of sclerotization, size and proportions, or surface texture), if not quantified, are nevertheless useful to individualize anatomical elements.

Phylogenetic datasets, novelties, and individuation

We will start by exploring what constitutes a separate anatomical element, and how the phylogenetic dataset helps determine its individuality. As stated above, in our strategy we will concentrate on counting individuated structures, rather than on their shapes, textures or other qualities. First, some characters express directly the presence or absence of structures and are straightforward for this purpose. These were called *neomorphic characters* by Sereno (2007), and conform to the concept of evolutionary novelties (Müller and Wagner, 1991) of prime importance for a complexity measure. A gross classification of kinds of novelties, as stated by Müller and collaborators, is helpful to illustrate the diversity of concepts covered by neomorphic characters in phylogenetics (Müller, 2010; Peterson and Müller, 2013; see Müller and Wagner, 1991: 238 for examples in mammals). Type I morphological novelties account for primary anatomical architectures, such as those defining the body plans of different animal phyla (example characters are presence or absence of tight junctions, chitinous cuticle, etc.). Type II novelties cover the appearance of discrete new elements in an already defined body plan (examples in spiders are the spinnerets, cheliceral venom gland, and male copulatory bulb). Type III novelties result from the individuation of pre-existing, serially homologous elements (e.g. the specialized setae aligned to form a comb in the hind legs, in the spider calamistrum; the sensory forelegs in whip-spiders). Neomorphic characters readily identify novel structures that will be counted in our complexity measure, but not all novelties are expressed in that way.

Sereno (2007) distinguished a second type of character, expressing variation in a quality for a given structure (e.g. hind trochanter length: short or long). In this case the structure (hind trochanter) is uniform, and the character records variation in a quality (length); these are called *transformational characters*. Many Type III novelties are expressed in the grammar used for transformational characters. For example, the character “posterior median eyes tapetum orientation: parallel or orthogonal to each other” seems to be a transformational character, but in fact indicates the occurrence of a complicated polarized light analyser used as a compass for navigation (Dacke et al., 1999) defining a large clade of spiders (Ramírez, in press). The same is true for a vast grey area of novelties, such as the change of con-

text of pre-existing structures, as for example the invagination of the posterior wall of the uterus externus and lining of female genital glands to conform a posterior receptacle in dysderoid spiders (Forster and Platnick, 1985), and the clustering of plesiomorphic elements to conform new structures (e.g. a definite patch of chemosensory setae on the male palpal tarsus) (see Müller and Wagner, 1991). An opposite situation occurs in characters recording the distribution of non-localized, modular structures on body regions. For example, the character “femoral trichobothria: absent or present” at first glance looks neomorphic, but because trichobothria are always present on some articles of the spider legs, the extension of their distribution on the femora is best understood as a quantitative change rather than an evolutionary novelty.

The understanding and classification of novelties is an open topic in biology (see Brigandt and Love, 2012), and consequently we have not attempted to sort out novelty types in our dataset, or even draw precise distinctions between neomorphic and transformational characters. In fact, the methods we apply below are useful to study whether some of these distinctions are meaningful in specific cases (e.g. testing the phylogenetic depth or diversity of clades defined by certain novelties). However, we implemented an explicit use of characters, both neomorphic and transformational, as a means to identify the evolutionary individuality of morphological regions. An example is the case of serially homologous appendages. Spiders have four pairs of legs, mostly similar to each other. In principle, we can suppose that construction of the legs consists of the repeated deployment of the same genetic and developmental mechanisms on both sides of four consecutive segments (see Klingenberg, 2008); thus, the mere number, either three, four, or five pairs of legs, does not add much to the complexity of the organism (see also Cisne, 1974; Wills et al., 1998; Adamowicz et al., 2008). However, some very flat spiders have an elongate hind trochanter (probably helping them to walk sideways beneath tree bark), thus implying that there is some mechanism that individuates the hind leg and, more specifically, its trochanter, to make it much longer than the remaining leg trochanters. The transformational character expressing the variation in length of the hind trochanter marks the individuality of that body part, which adds to the complexity of the organism. In this way, we use neomorphic and transformational characters to recognize body regions that can undergo independent evolutionary change, and count those in our complexity score.

Complexity scores on tree nodes

Because we are interested in an evolutionary context, we need a measure that can be computed not

only for observable organisms, but also for the hypothetical ancestors in a phylogenetic tree, in a way that is consistent with accepted methods of phylogenetic inference. One commonly used strategy involves the construction of a complexity measure for observed species, and an optimization resulting from the measure on the internal nodes of the tree. We call this strategy *terminal-driven*. For example, Kuntner et al. (2009) and Tataric and Cassis (2010) scored several elaborations of male and female genitalia (bumps, pockets, curves, etc.) as independent characters, identified simple and complex states, and counted how many of the complex states were present in each terminal, to obtain a complexity score for every terminal. They then optimized the scores as a continuous character on the tree to obtain the ancestral values, a strategy probably dictated by the comparative test they used (phylogenetically independent contrasts; Felsenstein, 1985). A different approach to the same problem, which we call *node-driven*, is applying the calculation used for terminals also to the internal nodes of the tree, but on the optimized values instead of the observations. The complexity score for each internal node is the sum of all the characters optimized as present in that node (e.g. Aburomia et al., 2003; Kim and Caetano-Anollés, 2011).

By its design, the node-driven strategy produces complexity values exactly correlated with the phylogenetic inference of ancestral character states. We should expect that both ways of estimation produce roughly similar results, but it is easy to imagine particular situations where this does not hold, both for the ancestral nodes and for transitions on branches (Fig. 1). Because of the more convincing consistency with the estimation of ancestral states, we will use here a node-driven calculation. The inferred values can be easily integrated in phylogenetic correlation tests based on optimized values (see Giannini and Goloboff, 2010).

Ontologies and hierarchical complexity

In the previous section we proposed a strategy to detect and count individuated structures amounting to the non-hierarchical aspects of complexity, both for terminals and for ancestors on a tree (Fig. 1). In this section we will incorporate the hierarchical dimension. Quantitative methods for phylogenetic analyses require that characters are logically independent of each other, thus imposing an atomization of the anatomy. This allows the algorithms to reconstruct hypothetical ancestors, one character at a time, which is extremely efficient in the tree-evaluation phase. However, to capture the hierarchical component of complexity, we need a representation of the entire morphological configuration for each hypothetical ancestor, rather than piecemeal hypotheses for isolated characters. For this

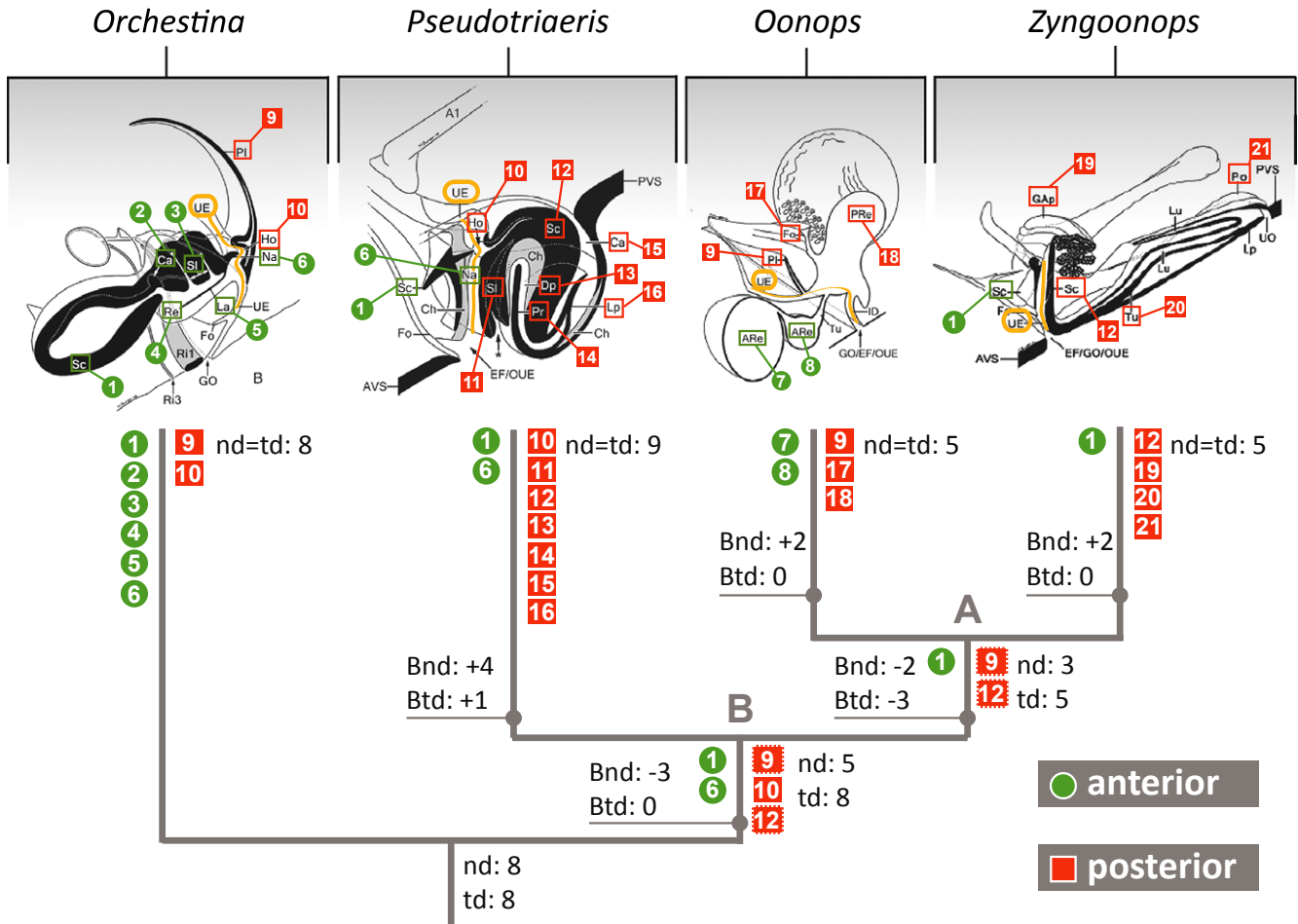


Fig. 1. Example showing the optimization of a “complexity” character as a continuous variable (terminal-driven) producing inaccurate results. Only non-hierarchical complexity is considered in this example. The female spider genitalia of goblin spiders (Oonopidae) are characterized by a high structural diversity in the anterior (features marked with green circles) as well as the posterior (features marked with orange boxes) section, as delimited by the uterus externus (yellow). The optimization of the characters as independent variables (nd, node-driven) infers three and five structures present in nodes A and B, respectively, while the continuous variable (td, terminal-driven) would wrongly suggest five and eight. This occurs because each terminal attained its complexity score by a combination of losses and additions of non-homologous structures. The transitions on branches (Bnd and Btd) are similarly discrepant. Examples are modified from the detailed work on oonopid genitalia by Matthias Burger (*Orchestina*: Burger et al., 2010; *Pseudotriaeris*, *Oonops*: Burger, 2012; *Zyngoonops*: Burger, 2011). For simplicity, optimizations use *Orchestina* as ancestral value. Ambiguous optimizations (dashed borders) are considered as present.

we will use an ontology to represent the hierarchical structure of morphology, and extend its application to integrate, for each ancestral node, the vector of ancestral character states in a hierarchical morphological representation (Fig. 2).

The logical representation and hierarchical structure of anatomy is a topic common to every biological discipline employing anatomical knowledge. As is becoming standard, we use an anatomical ontology for this purpose. Ontologies are used to represent knowledge on anatomy, genes, biological functions, and processes, and work as a bridge between domains such as comparative anatomy and genomics (Ashburner et al., 2000; Mabee et al., 2007a,b), as well as providing a normalized structure for data integration and repository-

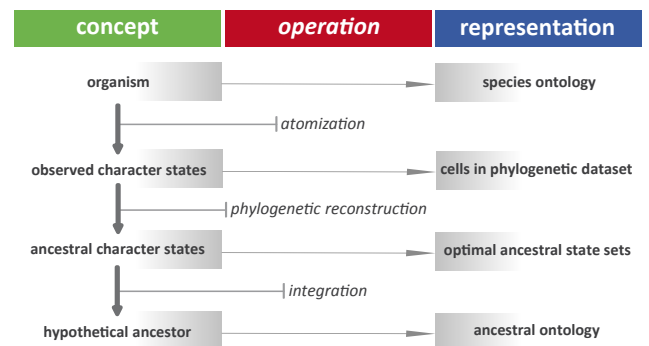


Fig. 2. Schematic view of the concepts, operations and representations involved in the inference of hierarchically structured morphological configuration of hypothetical ancestors.

ries within a given domain (see Vogt et al., 2013). In phylogenetics they have been shown to be efficient aids in the querying of relevant information in datasets (Ramírez et al., 2007). An ontology is similar to a glossary (all terms have a definition, their names are unique, and their synonyms are listed) with the addition of a logical structure of well-defined relationships between terms, which allows for reasoning and computation.

Organismic anatomy has a high degree of nestedness of regions within regions, and this is observable both in the developmental sequences and in their phenotypic results (Müller and Wagner, 1991). This organization has a correlate in phylogenetic characters describing phenotypic variation affecting large regions, or smaller regions nested within (e.g. the sclerotization of an entire tagma, such as the spider prosoma, versus a bulging of a small area in the prosoma, such as the clypeus). In Spider Ontology (SPD, <https://bioportal.bioontology.org/ontologies/SPD>), these hierarchical relationships are represented by the relations *part of* and *is a*. For example, the clypeus is part of the carapace, which in turn is part of the cephalic area, which is part of the prosoma. Similarly, a claw tuft seta is a tenent seta (other types include chemosensory, tactile, scale, and macroseta). From this structure and the properties of the relationships (i.e. transitivity), it can be reasoned that the clypeus is part of the prosoma, and the claw tuft seta is a seta. Several conceptual axes are usually represented in the same ontology (e.g. Dahdul et al., 2012). For example, a claw tuft seta is also part of the claw tuft, which is part of the tarsus (note that a given term, such as the claw tuft seta, may have more than a single parent: claw tuft and tenent seta); similarly, the heart is part of the opisthosoma and also part of the circulatory system. The ability to represent multiple hierarchies in a single ontology is especially convenient because it is then straightforward to derive inferences for specific body parts or organ systems, and comparisons between them. There are many relationships used in diverse ontologies, representing development (*develops from*), spatial arrangement (*adjacent to*), processes (*regulates*), etc. (Smith et al., 2005; The Gene Ontology Consortium, 2012). Here we focus on just two relationships (*is a*, *part of*), which are convenient to represent the main aspects of comparative anatomy usually represented in phylogenetic datasets.

We use SPD, a multi-species ontology, aimed to represent the anatomy of any spider species. From the inception, SPD arose to accommodate the terms employed in phylogenetic characters (see Ramírez et al., 2007), and later gradually incorporated further terms for more general purposes. A multi-species ontology is a general consensus representation of the anatomical elements in a lineage (e.g. Teleosts: Dahdul

et al., 2010b; Hymenoptera: Yoder et al., 2010; amphibians: Maglia et al., 2006; flowering plants: Ilic et al., 2007; neuroanatomy of invertebrates: Richter et al., 2010; see Dahdul et al., 2010b for a discussion about the design of a multi-species ontology). Ideally, such an ontology includes all the anatomical terms in use for a taxon, although many of them do not occur in any given organism. For example, in spiders the serrate accessory claws and the claw tufts are clusters of setae that occur at the tip of the leg tarsi in some species, but so far have not been found together in any single species. In contrast, other ontologies represent one species only, usually a model organism (e.g. *Drosophila melanogaster*: Drysdale, 2001; zebrafish: Sprague et al., 2006; *Xenopus laevis* anatomy and development: Segerdell et al., 2008; human: Rosse and Mejino, 2003; fission yeast: Harris et al., 2013; *Tribolium castaneum*: Dönitz et al., 2013).

We could imagine an analogue of the single-species ontology as a subset of a multi-species ontology, where all the elements not instantiated in a selected species (i.e. not present) were pruned away. For example, the ontology of *Araneus diadematus* (an orb-weaver spider) will include the serrate accessory claws but not the claw tufts. For this reason, it can be quickly understood that neomorphic characters point explicitly to elements that must be pruned in certain species when scored as absent, along with all of its substructures. We use those pruned ontologies to calculate complexity scores for terminals, and by the same logic of the node-driven approach, extend this idea to apply to hypothetical ancestors, represented as internal nodes on the tree. Using character optimization, we infer all the elements that are present in an ancestor, and produce a subset of the ontology to accommodate them, including their relationships and the complete path to the top-level class. For the sake of brevity, we term these subsets of the ontology as *ancestral* and *terminal*, or in general *node ontology* to apply both to terminals or to ancestors. The ancestral ontology is a representation of the whole anatomy of the hypothetical ancestor, as accurate as the multi-species ontology, and reproducing its reasoning capabilities. For example, while the phylogenetic inference tells us that some ancestor has claw tuft setae and scopular setae, the ancestral ontology will infer that both are kinds of tenent setae, which occur on the appendages of the prosoma, and that the tenent setae are instantiated in the ancestor.

A similar approach, as implemented here, is used to propagate annotations of gene function over phylogenetic trees in the PAINT application (Gaudet et al., 2011), although this implementation suggests ancestral assignments for subsequent manual curation. Our application deals with the much simpler problem of a pre-curated dataset, and hence allows for a fully automated assignment of ontology terms to ancestors.

Note also that Mabee et al. (2007a,b), while exploring the links between phylogenetic datasets and genomics using ontologies, arrived at an equivalent grammar for characters and their states as that of Sereno (2007).

In our approach, the phylogenetic matrix defines the scope of the study. We have only accounted for ontology elements that are referred to by at least one character in the phylogenetic matrix, either neomorphic or transformational. Once a term is included, we have also included the complete path of parent terms and relationships connecting to the root. Operationally, we have implemented two simple rules to build a terminal or ancestral ontology. Starting from the multi-species ontology:

(i) load all the terms referred in the dataset, and also all terms in their path to the root; this defines the scope of the study. Then, for each terminal or ancestor,

(ii) if an element is absent, all its descendants are absent as well (e.g. if calamistrum is absent, then calamistrum seta is also absent).

We then implemented three simple rules to decide what elements are counted in the complexity score. For every terminal or internal node, a term is counted if (1) it is represented in the node ontology, and (2) it fulfils one or more of the following criteria:

(a) it could be absent in some other organism (inferred by association with a neomorphic character);

(b) it may have some individuated phenotypic variability (inferred by association with a transformational character);

(c) it has more than one child term (inferred from the ontology; the region is a conceptual cluster of other elements).

Criterion (a) counts elements that are evolutionary novelties of some taxa in the dataset, or to all of the taxa if the character is uniformly scored (e.g. spinnerets present for all spiders). Criterion (b) counts elements that can undergo independent evolutionary change, and thus have some degree of individuation. Criterion (c) will filter out intermediate terms with a single child, which probably express that there may not be a real or necessary distinction between parent and child terms; this has a correspondence with the quality control metric in BioPortal (Noy et al., 2009) detecting classes with only one subclass, and also with the method “*include-ComputedIntermediates*” of OntoFox (Xiang et al., 2010). Building of the node ontology after rules (i) and (ii) is also equivalent to the method “*includeAllIntermediates*” in OntoFox, using the root of the ontology as the top-level class (see also Courtout et al., 2011).

Data and methods

We used a dataset of 166 spider species belonging to 49 families, scored for 393 characters (Ramírez and

Michalik, 2014), most of them from morphology and a few from behaviour and silk structure. This study aimed to resolve the relationships of two-clawed spiders, and has numerous outgroups spanning most clades of Araneomorphae, especially entelegynes, the most diverse clade of spiders.

We used version 0.9.9 of the Spider Ontology available through the OBO Foundry portal (<http://purl.bioontology.org/ontology/SPD>). The ontology was edited in OBO-Edit (Day-Richter et al., 2007). All characters were mapped to ontology terms similarly as in Mabee et al. (2007a,b), Sereno (2007), Dahdul et al. (2010a), and Balhoff et al. (2010). Because we are not interested in specific qualities, we have used a simpler mapping of characters to ontology terms than the more elaborate Entity-Quality syntax implemented in Phenex (Balhoff et al., 2010; Dahdul et al., 2010a). Instead of mapping every character state to one or more entity terms plus one or more quality terms, we only mapped the entire character to the entity terms in the ontology.

For simplicity, we used a single, fully resolved tree, the reference tree obtained in Ramírez and Michalik (2014), although arbitrarily resolving a single trichotomy at the base of Eutichuridae. For the calculations of ancestral states and the production of coloured trees we used custom scripts written for TNT (Goloboff et al., 2008). Ontology terms were represented as instances for ancestors or species when the corresponding neomorphic character was optimized either as present or as ambiguous. To evaluate the effect of ambiguous optimizations, a global calculation of the count of terms with a stricter rule (ambiguous as absent) was also computed for terminals. Node ontologies (for ancestors or terminals) were represented as separate OBO files. To compute complexity scores for terminals with missing entries, we used the optimization of the immediate ancestor. For the parsing of the ontology files and calculation of complexity scores and ancestral ontologies we used a custom script written in Visual Basic as implemented in Sax Basic in IMatch (www.photools.com), interacting with an MS-Access database also containing the mapping of characters to ontology terms.

All the data and scripts used for this study are available in the Dryad repository, along with raw and formatted data of the results (Ramírez and Michalik, 2014).

Results

Mapping of characters to entity terms in the ontology

We summarize in Table 1 a few exemplar cases to illustrate the mapping of characters to terms in the ontology. While some characters refer to a single

Table 1
Example cases of mapping of characters to terms in the ontology

Example	Character	Mapped ontology terms	Notes
1	Cribellum: present or absent	Cribellum (SPD:0000115)	Neomorphic
2	Calamistrum: absent or present	Calamistrum (SPD:0000237)	Neomorphic. Correlated with cribellum
3	Calamistrum organization: linear or oval	Calamistrum (SPD:0000237)	Transformational
4	Cribellate silk axial lines: present or absent	Cribellate band axial line (SPD:0000406)	Neomorphic. Silk structure
5	Colulus: absent or present	Colulus (SPD:0000117)	Neomorphic. Homologous (reduction) of cribellum (see Case 1)
6	ALE–PLE lens distance: separated or juxtaposed	Posterior lateral eye (PLE, SPD:0000051) Anterior lateral eye (ALE, SPD:0000049)	Transformational. Composition of two localized entities
7	Anterior eye row curvature: notably procurved or approximately straight	Anterior median eye (SPD:0000048) Anterior lateral eye (SPD:0000049)	Transformational. Cluster not in SPD (“anterior eye row”), mapped to localized components
8	Femoral trichobothria: absent or present	Femur (SPD:0000375) trichobothria (SPD:0000313)	Transformational. Composition of one localized, and one non-localized entity
9	<i>Hortipes</i> sensor: absent or present	Metatarsus I (SPD:0000087) metatarsus II (SPD:0000095)	Neomorphic: Entity not in SPD (“ <i>Hortipes</i> sensor”), mapped where it occurs
10	Heart ostia pairs: four, three, or two Heart ostium 5: present or absent Heart ostium 4: present or absent Heart ostium 3: present or absent	Heart ostium 5 (SPD:0000607) Heart ostium 4 (SPD:0000606) Heart ostium 3 (SPD:0000605)	Neomorphic, mixed. Binary-recoded in three characters

entity (monadic; examples 1–5 in Table 1), others refer to relationships of two localized entities (relational; examples 6–7), the distribution of non-localized entities over a localized one (example 8), organs that occur on more than one body part (example 9), and multistate characters applying to several entities (example 10). Examples 1 and 2 refer to two entities that are functionally correlated and always occur together (cribellum and calamistrum, see below), and thus to preserve independence, the phylogenetic dataset only has one of the neomorphic characters (for the cribellum, in this case). To simplify the mapping of characters to terms we scored additional “bookkeeping” neomorphic characters (all inactive for the phylogenetic analysis) to document presence or absence of structures that otherwise had to be inferred using contextual, sometimes complex methods. This included the binary recoding of multistate characters that apply to several entities (example 10). Example 6, a relational character expressing the separation between two eye lenses, refers to two terms in the ontology (the anterior and posterior lateral eyes). Even if they are modular

(spiders have eight eyes), each eye can be individuated and homologized with a specific eye in other species, and thus eyes are *localized* entities. Example 7 is similar, but it is phrased as applying to a cluster (the anterior eye row, widely used in taxonomy of spiders) that is not represented in the ontology; instead, the character is mapped to its components (the median and lateral anterior eyes). Example 8 applies to two entities, but one is localized (the femur, the third leg article) and the other (the trichobothria), as discussed above, are non-localized setae that may occur in many leg articles; correspondences in trichobothria are not proposed one-to-one between distantly related spiders. Example 9 illustrates the issue of granularity, with a neomorphic character for a sensory organ found in the metatarsi of the first and second legs in spiders of the genus *Hortipes* (Bosselaers and Jocqué, 2000). Because of its narrow occurrence, the organ is not represented in the ontology, and hence the character was mapped to those metatarsi only, with the same effect as a transformational character. We have also included a few terms from behaviour (cribellate silk carding

movements), web (preying web and orb-web architecture), and silk (kinds of silk fibres composing the cribellate band; example 4). These are sufficiently stereotyped to make clear-cut characters and are customary in spider phylogenetics, hence the structural terms are already in the ontology. These examples are useful to demonstrate that the same reasoning can be extended to other fields of comparative biology beyond morphology.

Homology matters

While the top levels of the Spider Ontology are structural divisions (e.g. portion of organism substance, acellular anatomical structure, organ system; see Haendel et al., 2008), and some high-level terms are defined in terms of function (e.g. digestive system, sensory organ), as the terms become more specific, homology notions become more necessary for their application to concrete organisms. This includes simple operations, such as identifying the spinnerets in a spider, or more elaborate ones, such as the labelling of all gland spigots in an entelegyne spider, for which there may be seven types in the same individual (see Coddington, 1989). In the first case, the user may not even be conscious that she or he is resorting to homology, but in the second, the operation is more explicit.

As the Spider Ontology arose to manage the morphological concepts used in phylogenetic datasets, it is natural that it incorporated much of the pre-processed homology correspondences on its structure and definitions, to make room for the variety of form and function that the same organ may have in different organisms. In this way, the ontology accommodates the vast majority of homology statements currently accepted in spider systematics. We have not represented, however, the most hypothetical homology relations in the ontology, especially when they involve important structural reorganizations. For example, the colulus is not represented by a neomorphic character in the original matrix, because it is considered a remnant of the cribellum (Table 1, examples 1 and 5), which in turn is considered a radical modification of the anterior median spinnerets as it occurs in liphistiid spiders (not represented in this dataset; see Platnick and Gertsch, 1976). Once we added a bookkeeping character for the colulus (case 5), it refers to a different entity than the cribellum, although they are homologous. As explained by Dahdul et al. (2010b), the ontology can represent both entities in their structural aspects (e.g. their position and components), while the homology statements could be recorded in a separate table, with their respective sources of evidence (see also Vogt et al., 2010). Other homology relationships not represented in the ontology are, for example, the median tracheae—3rd abdominal entapophyses (Purcell,

1909; Ramírez, 2000)—and copulatory bulb—palpal claw (see Coddington, 1990).

Aligned ontologies

Figure 3 shows a section of the multi-species ontology side by side with one ancestral and two terminal ontologies. Given that all our node ontologies originate as a subset of the same multi-species ontology, the top-down alignment is trivial (for the more complex case of aligning pre-existing ontologies of diverse organisms, see Parmentier et al., 2010; Dahdul et al., 2012; Niknejad et al., 2012; Bertone et al., 2013). The corresponding terms were aligned to highlight that some body regions are differentially enriched with structures across taxa. In this case we focus on metatarsus IV, individualized from the other serially homologous metatarsi, and on the leg tarsus.

In our phylogenetic dataset there are no characters directly mapped to metatarsus IV. However, we do have the character “calamistrum, absent or present”, and the calamistrum is part of metatarsus IV. Thus, when the calamistrum is present, we infer that metatarsus IV is individuated, and hence represented in the ontology as a separate structure. Then, for example, *Hypochilus* adds a score of 1 in complexity, because of having a calamistrum (Fig. 3b).

When examining whether metatarsus IV should add to the complexity score, it is clear that *Hypochilus* has no peculiarity about it other than the presence of a calamistrum (in this dataset at least); it looks otherwise just like any other leg metatarsus, and accordingly lacks any child term other than the calamistrum, which is already accounted for. Hence, metatarsus IV, although represented in the terminal ontology of *Hypochilus*, does not add to the complexity score; it is a term with a single child without any grouping purpose.

In the ancestor of Cycloctenidae (Fig. 3c) no calamistrum is present, nor any other peculiarity in metatarsus IV that could make it different from metatarsi I–III, and hence it is not represented in the ontology, and does not add to the complexity score. The ontology, however, has a generic term “metatarsus”, for any metatarsi I–IV, which is included in every species ontology, and in all cases adds to the complexity score. This generic metatarsus has several characters pointing to it, which makes it distinctive from other leg articles such as the tibia or tarsus.

The calamistrum is used to comb a special type of silk for entangling prey. Many spiders abandoned web construction to become active hunters, which is reflected in some simplification of the spinning apparatus (see below; Fig. 5) and in the loss of the calamistrum (Fig. 3c,d). Of these cases, some spiders

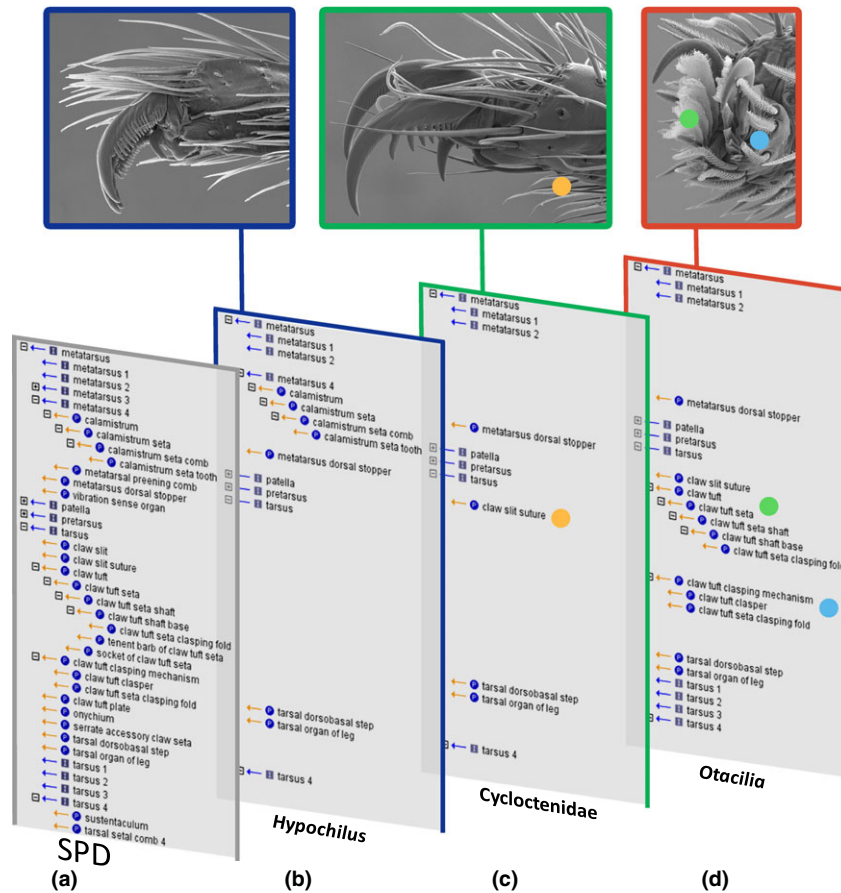


Fig. 3. A section of a multi-species ontology of spiders focusing on metatarsus IV, and in the generalized leg tarsus, aligned with species and ancestral ontologies. For simplicity, instances are not shown in the figure. (a) The Spider Ontology. (b) *Hypochilus pococki*, a spider species with calamistrum and without tenent setae. (c) Ancestral ontology of Cycloctenidae, the ancestor of *Cycloctenus nelsonensis*, and *Toxopsiella minuta*, both without calamistrum or tenent setae. (d) *Otacilia* sp. MJR-410, without calamistrum but with tenent setae.

developed clusters of adherent setae on the tip of the tarsi (the claw tuft), which allow them to walk on smooth surfaces and handle their prey (see Wolff et al., 2013; Ramírez and Michalik, 2014). This is reflected in an increase of complexity in the tarsi (Fig. 3d), with several elements in the claw tuft that can evolve independently, as reflected by 11 characters in the dataset.

Overall complexity

Of the 600 terms in the Spider Ontology, a subset of 350 is sufficient to accommodate all the 243 anatomical elements referred to in the phylogenetic dataset, plus the necessary terms to complete a path to the root. The overall complexity ranges from 184 to 236 individuated elements, and both extreme values are reached in terminals rather than in ancestors. The most complex terminal is *Acanthoctenus*, a cribellate spider with complex spinning organs, unusual by also having claw tufts. The simplest is *Ariadna*, a haplogyne

spider with relatively simple genitalia, simple spinning organs, and a limited repertoire of setae. The complexity in ancestors ranges from 200 to 235. Most of the node ontologies, either for ancestors or for terminals, are unique (i.e. combinations of morphological parts not inferred or observed in any other node). Table 2 shows the sources contributing to the complexity scores for some exemplar terminals and internal nodes.

Figure 4 shows the overall complexity scores reconstructed on the entire tree. The area near the root of the tree has low complexity values, corresponding to the lack of differentiation in many organ systems (spinning organs with few spigot types, unelaborated genitalia, lack of tenent setae). There is a stepwise increase in overall complexity in the Entelegynae (origin of the fertilization duct in female genitalia), the Divided Cribellum Clade (novel structures on the chelicerae and their articulation with the carapace, probably associated with a closer engagement with prey, rather than the distant prey manipulation of specialized web builders). The highest overall complexity

Table 2
Count of ontology terms in selected nodes (corresponding to species or ancestors)

Node or terminal name	(a) Terms in node ontology	(b) Leaves	(c) Non-trivial intermediates	(d) Total complexity
<i>Hypochilus pococki</i>	254	100	105	205
<i>Ariadna boesenbergi</i>	232	88	96	184
<i>Acanthoctenus cf. spinipes</i>	283	116	120	236
<i>Trachelopachys ammobates</i>	269	108	115	223
Root (Araneomorphae)	272	113	114	227
Cycloctenidae	264	109	109	218

(a) Total count. (b) Terms in the node ontology without descendants (leaves). (c) Non-terminal terms in the ontology that are referred by characters, or have more than one child. (d) Total complexity score (b + c).

is reached by members of the RTA clade retaining the plesiomorphic cribellum and calamistrum and usually a prey catching web, but incorporating many of the above mentioned novelties. From these, several independent losses of the cribellum and calamistrum lead to simplification, partially compensated for by the origination of tenent setae in *Dionycha* (and palpimanoids).

We calculated the transitions in complexity over branches (Fig. 4a,b); positive values are increases in complexity, negative are simplifications. The modal peak shows that many of the branches are static, but the distribution is skewed towards simplification. The transitions in complexity have no apparent correlation with overall character branch lengths (Fig. 4c), and the same is true when only the neomorphic characters are considered (not shown).

To compare the terminal-driven and node-driven approaches, we created a continuous character with the complexity scores for terminal species, and obtained the optimizations at the internal nodes. Comparing these optimizations with the values calculated from the ancestral ontologies shows a good correlation, but also a wide margin of discrepancies (Fig. 4d). The divergences are, however, much higher when the transitions in complexity over branches are compared (Fig. 4e), rather than the overall values at nodes. This is remarkable, as most of the comparative methods focus on transitions.

The overall complexity score is composed of values from all the organ systems and body parts, but often we are interested in specific sources of complexity. By using the structure of the ontology, it is straightforward to dissect the individual signals of complexity scores (Fig. 5). The complexity of spinning organs has some correlation with the silk structures (webs and cribellate silk threads), although there are large variations in spinning organ complexity not related to web structure and complexity. By contrast, the tenent setae seem highly correlated with the abandoning of capture webs (as in Wolff et al., 2013).

Note that the global complexity values are not the sum over all organ systems; a given body part may

participate in several systems, but will be counted only once in the overall score. The ontology and the complexity scores from some systems involve structures that may seem counterintuitive at first. For example, in an ontology for setae, all terminal elements (leaves) are kinds of setae or structures thereof, but there are intermediate terms such as the basal article of the chelicerae and metatarsus IV, because some setae are specific to those articles (e.g. whisker setae and calamistrum, respectively). If these intermediate terms are counted or not in the complexity score depends, as above, on how many child terms they have, or the presence of characters from setae pointing to them.

Elaborations and limitations

Ambiguities and multiple trees

For this study we implemented a simple treatment of ambiguity. The states of neomorphic characters may be ambiguously optimized at a node; here we considered an element as present in this case, thus producing a maximum estimation. In our example, only 2% of the optimizations were ambiguous, and thus the effect is only marginal. To fully incorporate ambiguities in the calculations, it would be possible to report the range between minimum (i.e. considered absent when ambiguous) and maximum estimations, or the midpoint. If needed for a statistical test, the combination of possible reconstructions could also be randomized (e.g. as in Giannini and Goloboff, 2010). In the case of multiple optimal trees, the ancestral ontologies could be calculated for the nodes present in the strict consensus, again with maximum and minimum complexity estimations obtained from iteration over all the optimal trees.

Missing entries

In our method it is straightforward to calculate the species ontology and the complexity scores for terminals with missing entries. The character optimization

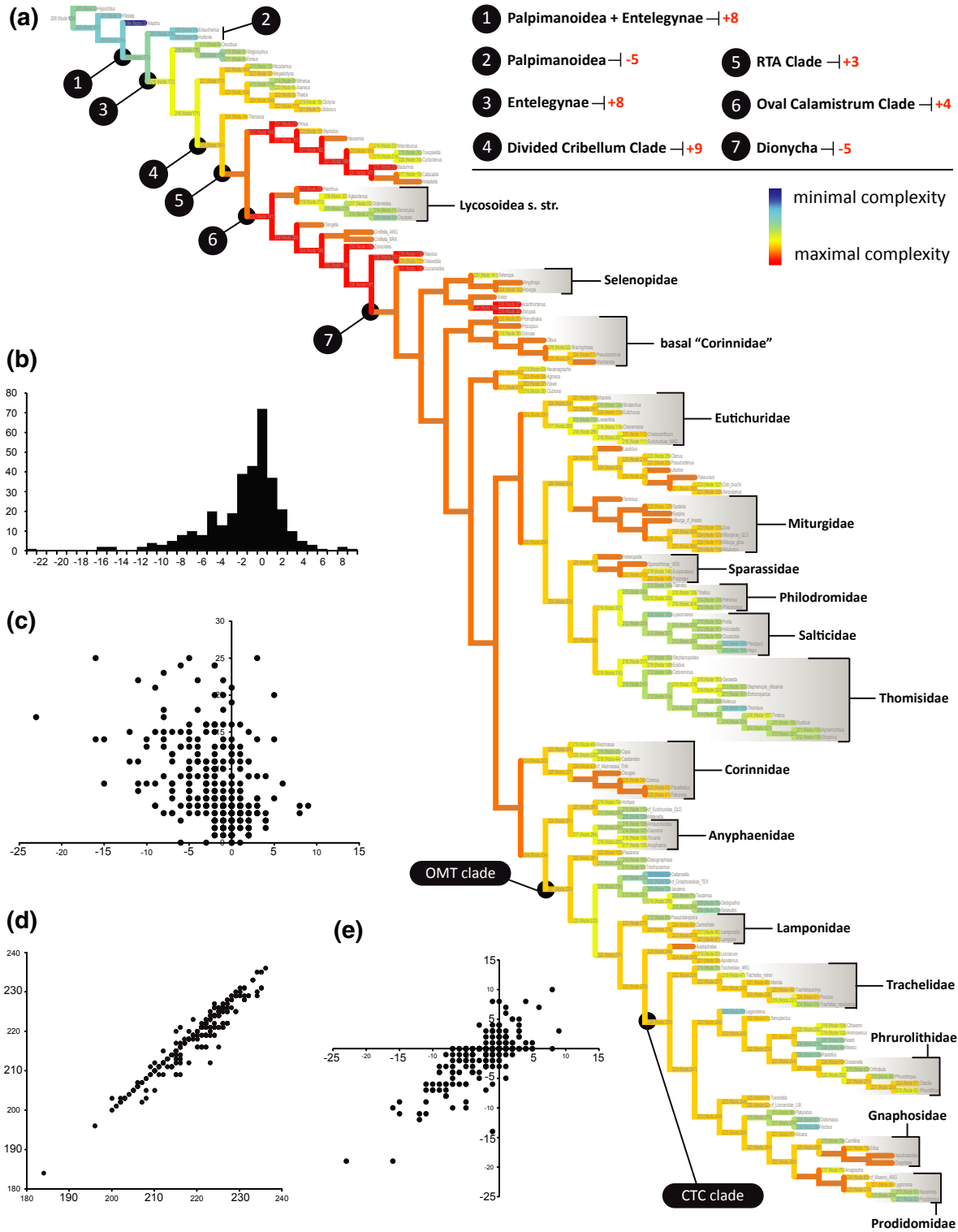


Fig. 4. (a) Overall complexity. (b) Histogram of frequencies of complexity transitions on branches. (c) Complexity transitions (x) vs. character branch length (y). (d) Comparison of node-driven (x) and terminal-driven (y) approaches for the calculation of overall complexity. (e) Same, for complexity transitions on branches.

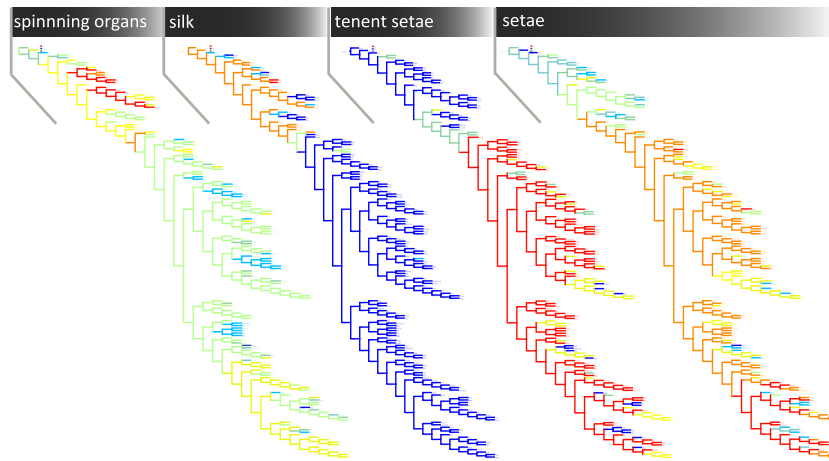


Fig. 5. Evolution of complexity for selected character systems. Colours represent the different degrees of complexity (same scale as in Fig. 4).

of the immediate ancestor produces a most-parsimonious inferred character state for those cases. For example, a species belonging to a clade with complex female genitalia will be inferred as such when the female is unknown.

Coarse ontology and approximated mapping

A phylogenetic analysis is always a rough approximation. Only the main strokes of the complex transformations at molecular and developmental levels causing morphological change can be represented as clear-cut characters (when they are known), but most often the mechanistic basis is unknown. Similarly, the interdependences of anatomical parts are only approximately represented in the ontologies. For example, the highly asymmetric claws of gradungulids (Forster et al., 1987; Michalik et al., 2013), the above-mentioned metatarsal sensor of *Hortipes*, and the additional receptor on the tarsal organ of Oonopidae (Platnick et al., 2012) occur coordinately on legs I and II. These phenotypic variants indicate a level of individuation of *groups* of legs (I and II vs. III and IV), instead (or on top) of each leg varying independently, but those groups are not represented in the ontology version used here. An advantage of the hierarchical structure of the ontology is that these imperfect mappings have a reduced impact, limited to a small neighbourhood of terms.

Individuation and transformational characters

In our simple implementation the individuation of body regions via transformational characters is done for the entire dataset, but tracing the origin of the individuation to a specific clade is preferable. An example is the character “trochanter IV length: <1.5 times the length of trochanter III or at least 1.5 times as long as

trochanter III”. First, we recognize that the comparison with trochanter III is merely administrative; the comparison could be done with either trochanter I or II, and hence we do not map the character as expressing variation in trochanter III. Second, by outgroup comparison we distinguish that trochanter IV becomes elongated, and not trochanters I–III shortened, thus mapping the character to trochanter IV only. We could distinguish between “normal” and “modified” states, and map the increased complexity only to the few taxa with elongated trochanters, as if it were a neomorphic character (a Type III novelty, in this case, marking the origin of the individuation of the hind trochanter). However, at the moment we do not have the conceptual elaborations to implement this kind of distinction for all the transformational characters in the dataset. Imperfections aside, we want to emphasize that phylogenetic matrices have a tremendous value for the detection of individualization of repetitive structures that become subjects of independent evolution. Once a case of individuation is detected and selected for experimental study, phylogenetic matrices and trees are again useful for the selection of good candidates for contrasting experiments. For example, two closely related species differing in the occurrence of a single individuated structure (and little else) will be good experimental candidates to avoid other sources of variation and thus better isolate the origin of individuation (Müller and Wagner, 1991).

Conclusions and prospects

We believe that our implementation of ancestral ontologies as a subset of a multi-species ontology opens a powerful and intuitive way of extending the reasoning capabilities of ontologies in an evolutionary context. The ancestral ontologies integrate the atom-

ized character hypotheses in a coherent representation of the anatomy of hypothetical ancestors on trees. The combination of the phylogenetic dataset and trees with the ontology allows inferences of wide application in evolutionary biology, such as the evolution of complexity and the individuation of modular structures as we have demonstrated here. Using an ontology it incorporates the hierarchical dimension of complexity and, perhaps more importantly, allows for sound and fast calculations for subsystems in a variety of axes, such as morphological divisions of the organism or organ systems. We believe that the development of applications of anatomical ontologies specific for phylogenetics will unlock the potential of phylogenetic data in other fields using ontologies as a common reference. At the same time, and as discussed by Franz (2013), these will help determine which ontology design types, relations and reasoning rules are more convenient for the management and integration of phylogenetic data.

Acknowledgements

We thank Stefan Richter and Christian Wirkner for discussion and invitation to the symposium “Morphology, Ontologies and Systematics” at the 32nd Meeting of the Willi Hennig Society in Rostock. We thank Lars Vogt, Paula Mabee, and Matt Yoder for discussion and advice on conceptual and technical aspects of ontologies and comparative data. Two anonymous reviewers suggested substantial improvements to the manuscript. TNT has been made freely available thanks to the Willi Hennig Society. This study was supported by grants from ANPCyT PICT 1007, and CONICET PIP 3209 to M.J.R., and DFG Mi 1255/5-1 to P.M. Early stages of ontology development were supported by grants PICT 1393, the US National Science Foundation (EAR-0228699 and NSF-DEB-0956049), and a short-term stay at NESCENT to M.J.R.

References

- Aburomia, R., Khaner, O., Sidow, A., 2003. Functional evolution in the ancestral lineage of vertebrates or when genomic complexity was wagging its morphological tail. *J. Struct. Funct. Genomics* 3, 45–52.
- Adamowicz, S.J., Purvis, A., Wills, M.A., 2008. Increasing morphological complexity in multiple parallel lineages of the Crustacea. *Proc. Natl Acad. Sci. USA* 105, 4786–4791.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29.
- Balhoff, J.P., Dahdul, W.M., Kothari, C.R., Lapp, H., Lundberg, J.G., Mabee, P., Midford, P.E., Westerfield, M., Vision, T.J., 2010. Phenex: ontological annotation of phenotypic diversity. *PLoS ONE*, 5, e10500.
- Bell, G., Mooers, A.O., 1997. Size and complexity among multicellular organisms. *Biol. J. Linn. Soc. Lond.* 60, 345–363.
- Bertone, M.A., Mikó, I., Yoder, M.J., Seltmann, K.C., Balhoff, J.P., Deans, A.R., 2013. Matching arthropod anatomy ontologies to the Hymenoptera Anatomy Ontology: results from a manual alignment. *Database (Oxford)*, 2013, bas057.
- Bosselaers, J., Jocqué, R., 2000. *Hortipes*, a huge genus of tiny Afrotropical spiders (Araneae, Liocranidae). *Bull. Am. Mus. Nat. Hist.*, 256, 4–108.
- Brigandt, I., Love, A.C., 2012. Conceptualizing evolutionary novelty: moving beyond definitional debates. *J. Exp. Zool. B Mol. Dev. Evol.*, 318B, 417–427.
- Burger, M., 2011. Structure and function of the female reproductive system in three species of goblin spiders (Arachnida: Araneae: Oonopidae). *Invert. Biol.*, 130, 148–160.
- Burger, M., 2012. Genital morphology of female goblin spiders (Arachnida: Araneae: Oonopidae) with functional implications. *Acta Zool.*, 94, 249–371.
- Burger, M., Izquierdo, M., Carrera, P., 2010. Female genital morphology and mating behavior of *Orchestina* (Arachnida: Araneae: Oonopidae). *Zool.*, 113, 100–109.
- Carroll, S.B., 2001. Chance and necessity: the evolution of morphological complexity and diversity. *Nature*, 409, 1102–1109.
- Cisne, J.L., 1974. Evolution of the world fauna of aquatic free-living arthropods. *Evolution*, 22, 337–366.
- Coddington, J.A., 1989. Spinneret silk spigot morphology: evidence for the monophyly of orbweaving spiders, Cyrtophorinae (Araneidae), and the group Theridiidae plus Nesticidae. *J. Arachnol.*, 17, 71–95.
- Coddington, J.A., 1990. Ontogeny and homology in the male palpus of orb-weaving spiders and their relatives, with comments on phylogeny (Araneoclad: Araneioidea, Deinopoidea). *Smithson. Contr. Zool.*, 496, 1–52.
- Coulet, A., Smail-Tabbone, M., Benlian, P., Napoli, A., Devignes, M.D., 2008. Ontology-guided data preparation for discovering genotype-phenotype relationships. *BMC Bioinformatics*, 9, S3.
- Courtout, M., Gibson, F., Lister, A.L., Malone, J., Schober, D., Brinkman, R.R., Ruttenberg, A., 2011. MIREOT: the minimum information to reference an external ontology term. *Appl. Ontol.*, 6, 23–33.
- Dacke, M., Nilsson, D.E., Warrant, E.J., Blest, A.D., Land, M.F., O’Carroll, D.C., 1999. Built-in polarizers form part of a compass organ in spiders. *Nature* 401, 470–473.
- Dahdul, W.M., Balhoff, J.P., Engeman, J., Grande, T., Hilton, E.J., Kothari, C., Lapp, H., Lundberg, J.G., Midford, P.E., Vision, T.J., Westerfield, M., Mabee, P.M., 2010a. Evolutionary characters, phenotypes and ontologies: curating data from the systematic biology literature. *PLoS ONE*, 5, e10708.
- 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., 2010b. The teleost anatomy ontology: anatomical representation for the genomics age. *Syst. Biol.*, 59, 369–383.
- Dahdul, W.M., Balhoff, J.P., Blackburn, D.C., Diehl, A.D., Haendel, M.A., Hall, B.K., Lapp, H., Lundberg, J.G., Mungall, C.J., Ringwald, M., Segerdell, E., Van Slyke, C.E., Vickaryous, M.K., Westerfield, M., Mabee, P.M., 2012. A unified anatomy ontology of the vertebrate skeletal system. *PLoS ONE*, 7, e51070.
- Day-Richter, J., Harris, M.A., Haendel, M., Lewis, S., 2007. OBO-Edit—an ontology editor for biologists. *Bioinformatics*, 23, 2198–2200.
- Dönitz, J., Grossmann, D., Schild, I., Schmitt-Engel, C., Bradler, S., Prpic, N.M., Bucher, G., 2013. TrOn: an anatomical ontology for the beetle *Tribolium castaneum*. *PLoS ONE*, 8, e70695.
- Donoghue, P.C., Purnell, M.A., 2005. Genome duplication, extinction and vertebrate evolution. *Trends Ecol. Evol.*, 20, 312–319.
- Drysdale, R., 2001. Phenotypic data in FlyBase. *Brief. Bioinform.*, 2, 68–80.

- Felsenstein, J., 1985. Phylogenies and the comparative method. *Am. Nat.* 125, 1–15.
- Forster, R.R., Platnick, N.I., 1985. A review of the austral spider family Orsolobidae (Arachnida, Araneae), with notes on the superfamily Dysderoidea. *Bull. Am. Mus. Nat. Hist.*, 181, 1–229.
- Forster, R.R., Platnick, N.I., Gray, M.R., 1987. A review of the spider superfamilies Hypochiloidea and Austrochiloidea (Araneae, Araneomorphae). *Bull. Am. Mus. Nat. Hist.*, 185, 1–116.
- Franz, N.M., 2013. Anatomy of a cladistic analysis. *Cladistics*. doi: 10.1111/cla.12042.
- Freeling, M., Thomas, B.C., 2006. Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Res.*, 16, 805–814.
- Gaudet, P., Livstone, M.S., Lewis, S.E., Thomas, P.D., 2011. Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. *Brief. Bioinform.*, 12, 449–462.
- Giannini, N.P., Goloboff, P.A., 2010. Delayed-response phylogenetic correlation: an optimization-based method to test covariation of continuous characters. *Evolution*, 64, 1885–1898.
- Goloboff, P.A., Farris, J.S., Nixon, K.C., 2008. TNT, a free program for phylogenetic analysis. *Cladistics* 24, 774–786.
- Haendel, M.A., Neuhaus, F., Osumi-Sutherland, D., Mabee, P.M., Mejino Jr, J.L., Mungall, C.J., Smith, B., 2008. CARO—the common anatomy reference ontology. In: Burger, A., Davidson, D., Baldock, R., (Eds.), *Anatomy Ontologies for Bioinformatics*. Springer, London, pp. 327–349.
- Harris, M.A., Lock, A., Bähler, J., Oliver, S.G., Wood, V., 2013. FYPO: the fission yeast phenotype ontology. *Bioinformatics*, 29, 1671–1678.
- Ilic, K., Kellogg, E.A., Jaiswal, P., Zapata, F., Stevens, P.F., Vincent, L.P., Avraham, S., Reiser, L., Pujar, A., Sachs, M.M., Whitman, N.T., McCouch, S.R., Schaeffer, M.L., Ware, D.H., Stein, L.D., Rhee, S.Y., 2007. The plant structure ontology, a unified vocabulary of anatomy and morphology of a flowering plant. *Plant Physiol.*, 143, 587–599.
- Kim, K.M., Caetano-Anollés, G., 2011. The proteomic complexity and rise of the primordial ancestor of diversified life. *BMC Evol. Biol.*, 11, 140.
- Klingenberg, C.P., 2008. Morphological integration and developmental modularity. *Ann. Rev. Ecol. Evol. Syst.*, 39, 115–132.
- Kuntner, M., Coddington, J.A., Schneider, J.M., 2009. Intersexual arms race? Genital coevolution in nephilid spiders (Araneae, Nephilidae). *Evolution*, 63, 1451–1463.
- Mabee, P., Ashburner, M., Cronk, Q., Gkoutos, G., Haendel, M., Segerdell, E., Mungall, C., Westerfield, M., 2007a. Phenotype ontologies: the bridge between genomics and evolution. *Trends Ecol. Evol.*, 22, 345–350.
- Mabee, P.M., Arratia, G., Coburn, M., Haendel, M., Hilton, E.J., Lundberg, J.G., Mayden, R.L., Rios, N., Westerfield, M., 2007b. Connecting evolutionary morphology to genomics using ontologies: a case study from Cypriniformes including zebrafish. *J. Exp. Zool. B Mol. Dev. Evol.*, 308B, 655–668.
- Mabee, P., Balhoff, J.P., Dahdul, W.M., Lapp, H., Midford, P.E., Vision, T.J., Westerfield, M., 2012. 500,000 fish phenotypes: the new informatics landscape for evolutionary and developmental biology of the vertebrate skeleton. *J. Appl. Ichthyol.*, 28, 300–305.
- Maglia, A.M., Leopold, J.L., Pugener, L.A., Gauch, S., 2006. An anatomical ontology for amphibians. *Pac. Symp. Biocomput.* 2007, 367–378.
- McElhinny, C., Gibbons, P., Brack, C., Bauhus, J., 2005. Forest and woodland stand structural complexity: its definition and measurement. *For. Ecol. Manage.*, 218, 1–24.
- McLellan, T., Endler, J.A., 1998. The relative success of some methods for measuring and describing the shape of complex objects. *Syst. Biol.* 47, 264–281.
- McShea, D.W., 1991. Complexity and evolution: what everybody knows. *Biol. Philos.*, 6, 303–324.
- McShea, D.W., 1996. Metazoan complexity and evolution: is there a trend? *Evolution*, 50, 477–492.
- Michalik, P., Piacentini, L., Lipke, E., Ramirez, M.J., 2013. The enigmatic Otway odd-clawed spider (*Progradungula otwayensis* Milledge, 1997, Gradungulidae, Araneae): natural history, first description of the female and micro-computed tomography of the male palpal organ. *ZooKeys*, 335, 101–112.
- Moczek, A.P., 2008. On the origins of novelty in development and evolution. *BioEssays*, 30, 432–447.
- Müller, G.B. 2010. Epigenetic innovation. In: Pigliucci, M., Müller, G.B. (Eds.), *Evolution: The Extended Synthesis*. MIT Press, Cambridge, MA, pp. 307–328.
- Müller, G.B., Wagner, G.P., 1991. Novelty in evolution: restructuring the concept. *Ann. Rev. Ecol. Syst.*, 22, 9–256.
- Newman, S.A., Müller, G.B., 2000. Epigenetic mechanisms of character origination. *J. Exp. Zool.*, 288, 304–317.
- Niknejad, A., Comte, A., Parmentier, G., Roux, J., Bastian, F.B., Robinson-Rechavi, M., 2012. vHOG, a multispecies vertebrate ontology of homologous organs groups. *Bioinformatics*, 28, 1017–1020.
- Noy, N.F., Shah, N.H., Whetzel, P.L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D.L., Storey, M.A., Chute, C.G., Musen, M.A., 2009. BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Res.*, 37, W170–W173.
- Parmentier, G., Bastian, F.B., Robinson-Rechavi, M., 2010. Homolonto: generating homology relationships by pairwise alignment of ontologies and application to vertebrate anatomy. *Bioinformatics*, 26, 1766–1771.
- Peterson, T., Müller, G.B., 2013. What is evolutionary novelty? Process versus character based definitions. *J. Exp. Zool. B Mol. Dev. Evol.*, 320B, 345–350.
- Platnick, N.I., Gertsch, W.J., 1976. The suborders of spiders: a cladistic analysis (Arachnida, Araneae). *Am. Mus. Novit.*, 2607, 1–15.
- Platnick, N.I., Abraham, N., Alvarez-Padilla, F., Andriamalala, D., Baehr, B., Baert, L., Bonaldo, A.B., Brescovit, A.D., Chousou-Polydouri, N., Dupérré, N., Eichenberger, B., Fannes, W., Gaubomme, E., Gillespie, R.G., Grismado, C.J., Griswold, C.E., Harvey, M., Henrard, A., Hormiga, G., Izquierdo, M.A., Jocqué, R., Kranz-Baltensperger, Y., Kropf, C., Ott, R., Ramírez, M.J., Raven, R.J., Rheims, C.A., Ruiz, G.R., Santos, A.J., Saucedo, A., Sierwald, P., Szuts, T., Ubick, D., Wang, X.P., 2012. Tarsal organ morphology and the phylogeny of goblin spiders (Araneae, Oonopidae), with notes on basal genera. *Am. Mus. Novit.*, 3736, 1–52.
- Prosdocimi, F., Chisham, B., Pontelli, E., Thompson, J.D., Stoltzfus, A., 2009. Initial implementation of a comparative data analysis ontology. *Evol. Bioinform. Online*, 5, 47.
- Purcell, W.F., 1909. Development and origin of respiratory organs in Araneae. *Quart. J. Microsc. Sci. (N.S.)*, 54, 1–110.
- 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. Linking of digital images to phylogenetic data matrices using a morphological ontology. *Syst. Biol.*, 56, 283–294.
- Ramírez, M.J., 2000. Respiratory system morphology and the phylogeny of haplogyne spiders (Araneae, Araneomorphae). *J. Arachnol.*, 28, 149–157.
- Ramírez, M.J. in press. The morphology and phylogeny of dionychan spiders (Araneae: Araneomorphae). *Bull. Am. Mus. Nat. Hist.*
- Ramírez, M.J., Michalik, P., 2014. Data from: calculating structural complexity in phylogenies using ancestral ontologies. Dryad Digit. Repository, doi:10.5061/dryad.5f0h8.
- Richter, S., Loesel, R., Purschke, G., Schmidt-Rhaesa, A., Scholtz, G., Stach, T., Vogt, L., Wanninger, A., Brenneis, G., Döring, C., Faller, S., Fritsch, M., Grobe, P., Heuer, C.M., Kaul, S., Möller, O.S., Müller, C.H.G., Rieger, V., Rothe, B.H., Stegner, M.E.J., Harzsch, S., 2010. Invertebrate neurophylogeny: suggested terms and definitions for a neuroanatomical glossary. *Frontiers Zool.*, 7, 29.
- Rosse, C., Mejino, J.L. Jr, 2003. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J. Biomed. Inform.*, 36, 478–500.

- Rowe, L., Arnqvist, G., 2012. Sexual selection and the evolution of genital shape and complexity in water striders. *Evolution*, 66, 40–54.
- Seegerdell, E., Bowes, J.B., Pollet, N., Vize, P.D., 2008. An ontology for *Xenopus* anatomy and development. *BMC Dev. Biol.*, 8, 92.
- Sereno, P.C., 2007. Logical basis for morphological characters in phylogenetics. *Cladistics*, 23, 565–587.
- Smith, B., Ceusters, W., Klagges, B., Köhler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A.L., Rosse, C., 2005. Relations in biomedical ontologies. *Genome Biol.* 6, R46.
- Sprague, J., Bayraktaroglu, L., Clements, D., Conlin, T., Haendel, M., Howe, D.G., Mani, P., Ramachandran, S., Schaper, K., Seegerdell, E., Song, P., Sprunger, B., Taylor, S., Van Slyke, C.E., Westerfield, M., 2006. The Zebrafish Information Network: the zebrafish model organism database. *Nucleic Acids Res.* 34, D581–D585.
- Tatarnic, N.J., Cassis, G., 2010. Sexual coevolution in the traumatically inseminating plant bug genus *Coridromius*. *J. Evol. Biol.*, 23, 1321–1326.
- The Gene Ontology Consortium, 2012. The gene ontology: enhancements for 2011. *Nucleic Acids Res.*, 40, D559–D564.
- Valentine, J.W., Collins, A.G., Meyer, C.P., 1994. Morphological complexity increase in metazoans. *Paleobiology*, 13, 1–142.
- Vogel, C., Chothia, C., 2006. Protein family expansions and biological complexity. *PLoS Comput. Biol.*, 2, e48.
- Vogt, L., 2009. The future role of bio-ontologies for developing a general data standard in biology: chance and challenge for zoomorphology. *Zoomorphol.* 128, 201–217.
- Vogt, L., Bartolomaeus, T., Giribet, G., 2010. The linguistic problem of morphology: structure versus homology and the standardization of morphological data. *Cladistics*, 26, 301–325.
- Vogt, L., Nickel, M., Jenner, R.A., Deans, A.R., 2013. The need for data standards in zoomorphology. *J. Morphol.*, 7, 793–808.
- Walls, R.L., Athreya, B., Cooper, L., Elser, J., Gandolfo, M.A., Jaiswal, P., Mungall, C.J., Preece, J., Rensing, S., Smith, B., Stevenson, D.W., 2012. Ontologies as integrative tools for plant science. *Am. J. Bot.*, 99, 1263–1275.
- Washington, N.L., Haendel, M.A., Mungall, C.J., Ashburner, M., Westerfield, M., Lewis, S.E., 2009. Linking human diseases to animal models using ontology-based phenotype annotation. *PLoS Biol.*, 7, e1000247.
- Wills, M.A., Briggs, D.E.G., Fortey, R.A. 1998. Evolutionary correlates of arthropod tagmosis: scrambled legs. In: Fortey, R.A., Thomas, R.H. (Eds.), *Arthropod Relationships*. Chapman & Hall, London, pp. 57–66.
- Wolff, J.O., Nentwig, W., Gorb, S.N., 2013. The great silk alternative: multiple co-evolution of web loss and sticky hairs in spiders. *PLoS ONE* 8, e62682.
- Wolstencroft, K., Stevens, R., Haarslev, V. 2007. Applying OWL reasoning to genomic data. In: Baker, C.J.O., Cheung, H. (Eds.), *Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences*. Springer, New York, pp. 225–248.
- Xiang, Z., Courtout, M., Brinkman, R.R., Rutenberg, A., He, Y., 2010. OntoFox: web-based support for ontology reuse. *BMC Res. Notes*, 3, 175.
- Yoder, M.J., Miko, I., Seltmann, K.C., Bertone, M.A., Deans, A.R., 2010. A gross anatomy ontology for Hymenoptera. *PLoS ONE*, 5, e15991.
- Zhang, H., Li, Y.F., Tan, H.B.K., 2010. Measuring design complexity of semantic web ontologies. *J. Syst. Softw.*, 83, 803–814.