

Ph.D. Thesis

**Advancement in Hymenoptera Research through Anatomy
Ontology Development**

Katja C. Seltmann

Supervisors: Dr. Andrew R. Deans¹ and Dr. Zsolt Péntzes²

¹Department of Entomology, Pennsylvania State University, 501 ASI Building, University Park, PA, USA ²Department of Ecology, University of Szeged, 6726 Szeged, Közép fasor 52. Hungary

Doctoral School of Biology
Department of Ecology
Faculty of Science and Informatics
University of Szeged

Szeged

2012

INTRODUCTION

Bees, wasps, ants, gall wasps, and sawflies all belong to an extraordinarily diverse lineage of insects. Since 1758 and Linnaeus first brought these organisms together under the name Hymenoptera, more than 140,000 species (Mason & Huber, 1993) have been described by generations of morphologists, primarily using prosaic natural language. The guidelines for how these descriptions were created have evolved through mentorship between students and their advisors, and through the requirements of the journals in which they were published. Contained within these descriptions is a corpus of anatomical information that is important for understanding the present day hymenopteran phylogenetic hypothesis (see Munro et al., 2011; Sharanowski et al., 2010; Sharkey, 2007; Sharkey et al., 2012 for a contemporary review). As the number of new descriptions continues to expand, the challenge faced by researchers, who must incorporate these articles and the descriptive statements contained within them, is increasing. The meaning behind terms used in descriptions can be elusive, as new terminology is often inadequately defined and illustrated, and comprehensive study requires significant exploration in order to fully determine the intended meaning, particularly for new students to the field. In addition, there is currently no straightforward method of synthesizing these descriptive statements because there is no uniform process through which descriptive statements are constructed. The community also lacks an automated utility for searching across morphological literature. For example, if one wanted to discover all of the species in the family Braconidae (>20,000) that do not possess ocelli, light detecting structures usually found on the head, it would be necessary to review all of the literature on braconids. This would first require that one identify the relevant articles based on taxon, using Google Scholar or other search engines, and then read or text-search each article for keywords such as 'ocelli' or 'eye'. There is currently no single search capable of retrieving this information (i.e., all relevant articles and all references to eye) in even a semi-automated way. Thus, a simple question may require months of research to answer, and many important articles may still be

overlooked. In addition, this method of literature search tends to confine authors to exploring only the specific taxon of interest, potentially neglecting any articles for other insect groups that may be pertinent to the researcher's morphological interests.

Hymenoptera studies is in an interesting position: with an estimated million species remaining to be described (Gaston, 1991), the body of descriptive work already accumulated is minimal compared to the number of articles yet to be produced. We thus have the potential to modify and improve our methodologies, by incorporating semantic, repeatable, and machine understandable techniques into our descriptions. Thus, moving forward, the power of the statements we use to describe organisms could be greatly enhanced, and those statements made much more available for biological research. The first step toward this goal is the creation of a structured, controlled vocabulary of Hymenoptera terminology, the Hymenoptera Anatomy Ontology (HAO). This ontology (HAO) is the primary tool utilized and augmented in this thesis.

Hymenoptera and the Semantic Web

Stated simply, an ontology is a set of concepts used to model a formalized domain and the logical relationships between concepts. In this case, the domain is Hymenoptera anatomy, and the concepts are definitions of morphological structures that follow specific rules of logic (for examples see Yoder, Mikó, Seltmann, Bertone, & Deans, 2010). The goal of ontology creation is to enable computer-based reasoning about morphological concepts (linked to terms, or words in publications) that are defined based on structural similarity. For example, consider the three morphological concepts in Hymenoptera represented by the terms: **radicle**, **scape**, and **antennal segment**. These concepts are related in the HAO in the following way: **radicle part_of scape** and **scape is_a antennal segment**. If a **scape is_a antennal segment**, and the **radicle** is *part_of* a **scape**, we can deduce that if an insect does not have a **scape** than the insect will also not have a **radicle**. However, the words **radicle**, **scape** and **antennal segment** are simply terms that represent concepts. A concept is the 'real life' physical structure on the wasp

we want to define. For example the concept for **radicle** is "The area that is located proximally on the scape, is limited distally by a constriction and bears proximally the basal knob". However, this concept may also be termed **antennal condyle**, **articulatory bulb** or **radicula**. The ontology recognizes all of these terms to be synonyms of the same concept, again increasing the power of the ontology to decipher descriptive language.

Although the improvements in clarification that can be achieved through clearly defined anatomical terminology are of clear benefit to research, there are other significant advantages to using an ontology over a simple glossary. As discussed in detail in Deans et al. (2012), incorporating concepts from an anatomy ontology into our descriptions increases the utility of the latter for the broader scientific community. Furthermore, this process results in a corpus of semantic statements about biodiversity that can be mined using computer-based reasoning. The incorporation of ontology is not new in the biological sciences, and as such is gaining increased recognition in the model organism community, particularly for tracking phenotypes of mutant organisms (Balhoff et al., 2010; Dahdul et al., 2010a; Dahdul et al., 2010b). At the same time new, open-access publications are emerging in which intelligent semantic markup is strongly encouraged. The journal *ZooKeys*, created using TaxPub XML markup, incorporates many semantic web links using Uniform Resource Identifiers (URIs), Life Science Identifiers (LSIDs), Globally Unique IDs (GUIDs), and other technologies to identify objects in explicit ways. Public Library of Science (PLOS) is a nonprofit organization committed to making scientific literature available online. PLOS journals are published under a Creative Commons License and include a highly competitive biology journal (*PLoS Biology*). PLOS is leading a charge toward semantic publication by enforcing the submission of new botanical names to the International Plant Names Index (IPNI), which stores the name in a database and applies a GUID. All of the PLOS and *ZooKeys* articles are available freely online as Optically Recognized PDFs. The PLAZI project (Plazi, 2012) is working toward retroactively annotating the historical literature in biological sciences through the development of applications such as the Golden Gate Software (Sautter, Böhm, & Agosti, 2007), which aims at making

old descriptions searchable. The Biodiversity Heritage Library (Biodiversity Heritage Library, 2011) is scanning biological literature that is out of copyright and is making it available on the internet. Perhaps most significant of all is the recent modification to the International Code of Zoological Nomenclature (International Commission on Zoological Nomenclature, 2012), which has recently loosened requirements that new organism descriptions be printed in paper copy journals (ICZN, 2012).

All of these efforts indicate a general trend toward highly accessible, semantic, and discoverable forms of scientific dissemination. This will enable researchers to find relevant articles and information easily, and, when coupled with anatomical ontologies, to make sense of all the terminology contained within those texts. The application of ontology to taxonomy is synergetic with these new trends, and holds the potential to make descriptive statements relevant and utilizable in diverse scientific disciplines, for new or prospective students of Hymenoptera, in genomic discovery, and, undoubtedly, in many purposes that have not yet been conceived.

AIMS

The aims of the work and contributions contained in this thesis are to contribute to the development of the Hymenoptera Anatomy Ontology (HAO) and to demonstrate its utility in guiding modern morphological and taxonomic research. Specifically, these aims are:

1. To accumulate terms and concepts for the Hymenoptera Anatomy Ontology by extracting terminology from text-based species descriptions and morphological texts (subheading: *Literature Analysis*).
2. To analyze the descriptive terminology in Hymenoptera literature using Natural Language Processing (NLP) clustering methods and to compare the results to our present understanding of Hymenoptera

phylogenetic relationships (subheading: *Literature Analysis*).

3. To promote the development of a methodology for linking taxonomic publications to Hymenoptera Anatomy Ontology concepts using Uniform Resource Identifiers (URIs), and to elucidate the benefits of ontology to the Hymenoptera community (subheading: *Ontology Utility and URI Development*).

MATERIALS AND METHODS

Building the HAO with Domain Expertise

The creation and implementation of an anatomy ontology requires significant expertise of the domain being described. The domain illustrated in the HAO is Hymenoptera anatomy; thus, skills in Hymenoptera morphology, taxonomy, and bioinformatics are essential for both guiding the augmentation of the ontology, and for communicating its benefits to the Hymenoptera research community. Terms and concepts were illustrated as part of this thesis, informing the HAO through exploration of mouthpart morphology characters and illustrations using brightfield (compound and Microptics®) and confocal laser imaging techniques. The essential nature of domain expertise is highlighted in all of the HAO publications, but its importance is emphasized in Bertone, Mikó, Yoder, Seltmann, and Deans, 2012; Mikó et al. (2012); Seltmann, Yoder, et al. (2012).

Constructing the HAO: Software

The research described here depended on both the design and implementation of the HAO. The primary ontology development software for the HAO is mx (Mx, 2012), a Ruby on Rails, MySQL-based open source content management system for descriptive taxonomy primarily coded by Dr. Matthew J. Yoder. Since 2008, the author of this thesis has coded several public portals for mx, and has extended mx for the descriptive term-based cluster analysis that is one of the key contributions of this thesis. Instances of

mx now manage all aspects of descriptive taxonomy, including: specimens, collecting events, extracts, sequencing progress, primer design, images, descriptions, diagnostic keys, literature, matrices, phylogenetic characters, ontology, public Web portals for data, and phylogenetic trees. The potential for interconnectivity via a database containing research driven taxonomic data (name catalogs, matrices, and so forth), with anatomy ontology development software, has guided our commitment to integrating anatomy ontology as part of the normal taxonomic revisionary process, and has presented us with a good workbench for demonstrating its benefits for Hymenoptera taxonomy. Examples illustrating the potential utility of an inclusive system for dissemination of taxonomic information, including interactive keys linked to HAO concepts, have appeared in the literature, with contributions by Seltmann (Sharkey et al., 2009).

From the standpoint of software implementation, the Ruby on Rails framework allows for rapid application development and has provided a productive environment for experimentation, including the studies presented in this thesis. The software is versioned using the content management repository SourceForge (SourceForge, 2012), thereby maintaining a transparent methodology. The repository manages versions of the software code base, and provides the flexibility to allow multiple developers to code on the project. The software for HAO development is outlined in Yoder et al. (2010) for the HAO in general, Seltmann et al. (2012) for the Proofer tool, and Seltmann, Péntzes, Yoder, Bertone, and Deans, 2012 for the Analyzer tool.

Methods for Aim 1&2: Literature Analysis

One major aim of this thesis is the exploration of descriptive terms accumulated as a product of building the HAO, including terms from legacy literature. Fundamentally, the development of the Hymenoptera Anatomy Ontology was experimental in its utilization of internet-based (mx) software and in the methodology it applied for extracting terms from legacy literature. In its early stages, the bulk of the terms and concepts were gathered via expert human inspection of the known primary literature, including important morphological publications and online glossaries. This method of

term accumulation was not expected to reveal the entire Hymenoptera lexicon, due to the hypothesized specificity of terminology based on higher-level classification groupings (ex. Chalcidoidea, Ichneumonoidea, Aculeata). In order to facilitate obscure term discovery, an active learning, dictionary-based, natural language recognition tool was implemented for examining text. This tool, referred to as the 'Proofer', is described in one major publication of the author and her colleagues. This tool constitutes part of an iterative approach to developing phenotype-relevant ontologies, and enables discovery of obscure descriptive terminology. As further documented in this thesis, Part I of this experiment was to sample the online *Journal of Hymenoptera Research* taxonomic descriptions for terminology not yet included in the HAO using the Proofer. In Part II, the sampled articles were then analyzed for occurrences of terms, using a variety of clustering methods, and the results were subsequently compared to our present understanding of Hymenoptera lineages. The general course of Proofer development and term analysis is discussed in Seltmann et al. (in press), and reviewed as the *Literature Analysis* section of this thesis.

Methods for Aim 3: Ontology Utility and URI Development

Term discovery from literature via text mining is one way that ontology may enhance document analysis and taxonomic description association. However, *post hoc* term discovery cannot directly illuminate the exact definition or associated ontological concept of a term as used in the source text. To remedy this, we developed a relatively simple and easily comprehensible methodology for linking terminology in descriptive texts with the Hymenoptera Anatomy Ontology via dedicated Uniform Resource Identifier (URI) tables, which could be included in a manuscript. The tables are created automatically, using the 'Analyzer' mx-based software tool, thereby linking specific words in the manuscript with individual, defined concepts in the HAO. Once published, the links reference Hymenoptera Portal Webpages, where definitions and illustrations of those concepts may be retrieved. A general discussion of the utility of this URI-based methodology for the Hymenoptera community is given in Seltmann et al. (2012) and

further presented in this thesis.

RESULTS

Access to the Hymenoptera Anatomy Ontology

The Hymenoptera Anatomy Ontology continues to be developed by the author and her colleagues, and can be accessed from multiple electronic resources. It is expected that most hymenopterists will access the HAO through the HAO Portal Website (<http://portal.hymao.org/>) in an online dictionary-style format, but the text versions of the ontology are also accessible through several widely used biomedical databases. The HAO project's association with the greater biomedical ontology community is maintained via National Center for Biomedical Ontology (<http://www.bioontology.org/>), ensuring that the HAO will be archived for long-term sustainability and distributed for broad potential use in other domains. The archived representation of the HAO are maintained in a standard type, based on either the Web Ontology Language (OWL) or Open Biomedical Ontology (OBO) format, and can be downloaded in either (OWL: <http://bit.ly/UnICTE> or OBO: <http://bit.ly/Tm1n6U>).

The Open Biomedical Ontology (OBO) Foundry (Smith et al., 2007) supports archiving and development of OWL and OBO formats as part of an effort to maintain and promote the use of biological ontologies across biological and medical domains. The OBO Foundry also facilitates ontology dissemination and use, as ontologies archived there are automatically made available through other portals such as BioPortal and Ontobee (BioPortal: <http://bit.ly/XVHdro> and Ontobee: <http://bit.ly/U2WQKa>). Further information regarding the decimation and impact of the Hymenoptera Anatomy Ontology, as a product and a revolution to descriptive taxonomy, can be reviewed in the Yoder et al. (2010) publication.

Results of Aim 1&2: Literature Analysis

One main set of results contributing to this thesis was obtained from a computer-assisted analysis of prior literature that was undertaken using the

HAO-based 'Proofer' text-mining tool. The analysis was based on a collection of 353 articles from the literature. Based on this set 1189 new morphological terms used by Hymenoptera taxonomists were collected. The broad conclusion that could be obtained through this analysis was that taxonomists use domain-specific terminology that follows taxonomic specialization, particularly at superfamily and family level groupings. Additionally, several different term-based cluster analyses were applied, and the exhibited a great deal of variability depending on the cluster algorithm used. The variability appears to reflect the fact that that a great deal of noise exists in this dataset, which can be assumed to be due to variations in terminology across the literature.

A significant augmentation of the development of the Hymenoptera Anatomy Ontology was achieved based on the new terms that were collected through this analysis and that were subsequently added to the mx database. The Proofer tool established its value in improving the efficiency of term extraction from legacy literature by reducing the number of terms presented to the user for review. To quantify this reduction, a systematic comparison of the number of terms presented to the user was performed, with and without the Proofer's full functionality implemented, for 25 randomly selected articles. This comparison demonstrated that the Proofer reduced the number of terms displayed to the user by 1/3 of the total actual word count of the article, which constituted an 80% reduction in the number of combinations of words displayed to a user by the Proofer. 180 of the 353 articles were identified to contain descriptions of new taxa, wholly or in part. The shortest tree returned from analysis was from the 'Sorensen-Average' cluster analysis, including characters that were coded for 2 or more terminals, and pruned to superfamily level. This tree resulted in 63 distinct groupings when the tree was pruned, with observable large clusters of Ichneumonoidea, Chalcidoidea, Symphyta, and Aculeata.

Furthermore, the analysis resulted in a compilation of the most common terms (morphological and qualitative) used by hymenopterists. The HAO itself only handles anatomical terminology, not qualitative terms (e.g. shiny, large). However separate ontology, The Phenotype and Trait Ontology

(PATO), exists to provide these qualitative concepts for phenotype annotation. We examined the applicability of PATO to Hymenoptera phenotype descriptions (intersection of HAO concept and PATO concept), and suggested potential terms for inclusion in PATO, as they are those most commonly used in Hymenoptera species descriptions. Results from development of the HAO through term discovery in the literature are presented in a recently accepted publication of the author (Seltmann et al., in press) and further documented in this thesis.

Results of Aim 3: Ontology Utility, Outreach and URI Development

One measure of success for any bioinformatics tool is its level of adoption by its target community. Outreach with the Hymenoptera community has been of utmost importance to the HAO project, and these discussions have already lead to original research, including a paper (Seltmann et al., 2012) of the which was a direct response to questions raised during several HAO project workshops. Since publication of the URI table concept was initiated, seven morphology publications (Buffington & Van Noort, 2012; Johnson & Musetti, 2011; Krogmann & Nel, 2012; Mikó et al., 2012; Sharkey & Stoelb, 2012; Talamas, Masner, & Johnson, 2011; Wharton, Ward, & Mikó, 2012) have adopted HAO terminology and the URI/Analyzer methodology. The general level of awareness and understanding of these tools amongst Hymenoptera morphologists has also increased during this period.

DISCUSSION

The work in this thesis was focused on the further development and application of the Hymenoptera Anatomy Ontology, a resource based on a foundation of explicitly defined anatomical concepts and a straightforward mechanism for referencing these concepts (URIs). Seltmann and colleagues demonstrated the necessity for such a unified resource for Hymenoptera terminology, as it was shown that hymenopterists use terminology specific to

superfamily or family they are describing. The implications of these developments are several-fold. In addition to increasing the repeatability of research on Hymenoptera, references to well-defined and illustrated anatomical concepts will open up their interpretation and use for a much broader array of biologists than the core group of highly specialized taxonomists that would obviously also benefit. The HAO, like other biological ontology efforts, is rapidly evolving, both in its underlying data and its application. Additionally, novel functionality for constructing anatomy ontologies, regardless of domain, was demonstrated by Seltsmann and colleagues. Facilitated construction of further arthropod anatomy ontologies will benefit the entire anatomy ontology community, and potentially impact many aspects of our science, as publications become semantically available.

Fundamentally, beneficial impacts may be anticipated for all areas of biological science that may depend on the correct interpretation of hymenoptera anatomical structures. This may include: biodiversity, host-parasite biology, collection digitization, genomics, ecology, evolutionary developmental biology (evo-devo), invasive species evaluation, agro-ecosystem management, and biological control.

REFERENCES

- Balhoff, J. P., Dahdul, W. M., Kothari, C. R., Lapp, H., Lundberg, J. G., Mabee, P., Midford, P. E., et al. (2010). Phenex: ontological annotation of phenotypic diversity. *PLoS ONE*, 5(5), e10500. doi:10.1371/journal.pone.0010500
- Bertone, M. A., Mikó, I., Yoder, M. J., Seltmann, K. C., & Deans, A. R. (2012). Matching arthropod anatomy ontologies to the Hymenoptera Anatomy Ontology: Results from a manual alignment. Database. *Database : the journal of biological databases and curation*.
- Biodiversity Heritage Library. (2011). Biodiversity Heritage Library. Retrieved February 1, 2011, from <http://www.biodiversitylibrary.org>
- Buffington, M. L., & Van Noort, S. (2012). Revision of the afro-tropical oberthuerellinae (cynipoidea, liopteridae). *ZooKeys*, (202), 1–154. doi:10.3897/zookeys.202.2136
- Dahdul, W. M., Balhoff, J. P., Engeman, J., Grande, T., Hilton, E. J., Kothari, C., Lapp, H., et al. (2010). Evolutionary characters, phenotypes and ontologies: curating data from the systematic biology literature. (J. Kelso, Ed.) *PLoS ONE*, 5(5), e10708. doi:10.1371/journal.pone.0010708
- Dahdul, W. M., Lundberg, J. G., Midford, P. E., Balhoff, J. P., Lapp, H., Vision, T. J., Haendel, M. A., et al. (2010). The teleost anatomy ontology: anatomical representation for the genomics age. *Systematic biology*, 59(4), 369–83. doi:10.1093/sysbio/syq013
- Deans, A. R., Yoder, M. J., & Balhoff, J. P. (2012). Time to change how we describe biodiversity. *Trends in ecology & evolution*, 27(2), 78–84. doi:10.1016/j.tree.2011.11.007
- Gaston, K. J. (1991). The Magnitude of Global Insect Species Richness. *Conservation Biology*, 5(3), 283–296. doi:10.1111/j.1523-1739.1991.tb00140.x
- ICZN. (2012). International Code of Zoological Nomenclature. Retrieved October 8, 2012, from <http://iczn.org/code>
- International Commission on Zoological Nomenclature. (2012). Amendment of Articles 8, 9, 10, 21 and 78 of the International Code of Zoological Nomenclature to expand and refine methods of publication. *ZooKeys*, 219, 1–10. doi:10.3897/zookeys.219.3944
- Johnson, N. F., & Musetti, L. (2011). Redescription and revision of the Neotropical genus *Pseudoheptascelio* Szabó (Hymenoptera, Platygasteridae, Scelioninae), parasitoids of eggs of short-horned grasshoppers (Orthoptera, Acrididae). *ZooKeys*, (136), 93–112. doi:10.3897/zookeys.136.1580
- Krogmann, L., & Nel, A. (2012). On the edge of parasitoidism: a new Lower Cretaceous woodwasp forming the putative sister group of Xiphydriidae +

- Euhymenoptera. *Systematic Entomology*, 37(1), 215–222.
doi:10.1111/j.1365-3113.2011.00608.x
- Mason, W., & Huber. (1993). Order Hymenoptera. *Hymenoptera of the world: An identification guide to families*. (pp. 4–12).
- Mikó, I., Friedrich, F., Yoder, M. J., Hines, H. M., Deitz, L. L., Bertone, M. a, Seltmann, K. C., et al. (2012). On dorsal prothoracic appendages in treehoppers (Hemiptera: Membracidae) and the nature of morphological evidence. *PLoS ONE*, 7(1), e30137. doi:10.1371/journal.pone.0030137
- Munro, J. B., Heraty, J. M., Burks, R. A., Hawks, D., Mottern, J., Cruaud, A., Rasplus, J.-Y., et al. (2011). A molecular phylogeny of the Chalcidoidea (Hymenoptera). *PloS one*, 6(11), e27023. doi:10.1371/journal.pone.0027023
- Mx. (2012). mx. Retrieved March 7, 2011, from <http://purl.oclc.org/NET/mx-database>
- Plazi. (2012). Plazi: Search and Retrieval Server. Retrieved December 10, 2012, from <http://plazi.org:8080/dspace/community-list>
- Sautter, G., Böhm, K., & Agosti, D. (2007). Semi-automated XML markup of biosystematic legacy literature with the GoldenGATE editor. *Pacific Symposium on Biocomputing*, 402, 391–402. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/17992751>
- Seltmann, K. C., Péntzes, Z., Yoder, M. J., Bertone, M. A., & Deans, A. R. (2012). Utilizing Descriptive Statements from the Biodiversity Heritage Library to Expand the Hymenoptera Anatomy Ontology. *PLoS ONE*.
- Seltmann, K. C., Yoder, M., Miko, I., Forshage, M., Bertone, M., Agosti, D., Austin, A., et al. (2012). A hymenopterists' guide to the Hymenoptera Anatomy Ontology: utility, clarification, and future directions. *Journal of Hymenoptera Research*, 27, 67. doi:10.3897/jhr.27.2961
- Sharanowski, B. J., Robbertse, B., Walker, J., Voss, S. R., Yoder, R., Spatafora, J., & Sharkey, M. J. (2010). Expressed sequence tags reveal Proctotrupomorpha (minus Chalcidoidea) as sister to Aculeata (Hymenoptera: Insecta). *Molecular phylogenetics and evolution*, 57(1), 101–112. doi:10.1016/j.ympev.2010.07.006
- Sharkey, M. J. (2007). Phylogeny and Classification of Hymenoptera. *Zootaxa*, 548, 521 – 548.
- Sharkey, M. J., Carpenter, J. M., Vilhelmsen, L., Heraty, J., Liljebblad, J., Dowling, A. P. G., Schulmeister, S., et al. (2012). Phylogenetic relationships among superfamilies of Hymenoptera. *Cladistics*, 28(1), 80–112. doi:10.1111/j.1096-0031.2011.00366.x
- Sharkey, M. J., & Stoelb, S. (2012). Revision of Zelodia (Hymenoptera, Braconidae, Agathidinae) from Thailand. *Journal of Hymenoptera Research*, 26, 31. doi:10.3897/jhr.26.2527
- Sharkey, M. J., Yu, D. S., Noort, S., Seltmann, K., Penev, L., & others. (2009). Revision of the Oriental genera of Agathidinae (Hymenoptera, Braconidae)

- with an emphasis on Thailand including interactive keys to genera published in three different formats. *ZooKeys*, (21), 19–54.
- Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L. J., et al. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology*, 25(11), 1251–5. doi:10.1038/nbt1346
- SourceForge. (n.d.). SourceForge. 2012. Retrieved from sourceforge.net
- Talamas, E. J., Masner, L., & Johnson, N. F. (2011). Revision of the *Paridris* nephtaspecies group (Hymenoptera, Platygastroidea, Platygastriidae). *ZooKeys*, (133), 49–94. doi:10.3897/zookeys.133.1613
- Wharton, R., Ward, L., & Miko, I. (2012). New neotropical species of *Opiinae* (Hymenoptera, Braconidae) reared from fruit-infesting and leaf-mining Tephritidae (Diptera) with comments on the *Diachasmimorpha mexicana* species group and the genera *Lorenzopius* and *Tubiformopius*. *ZooKeys*, 243, 27–82. doi:10.3897/zookeys.243.3990
- Yoder, M. J., Mikó, I., Seltmann, K. C., Bertone, M. A., & Deans, A. R. (2010). A gross anatomy ontology for hymenoptera. (C. S. Moreau, Ed.) *PLoS ONE*, 5(12), e15991. doi:10.1371/journal.pone.0015991

PUBLICATION LIST

Publications with significant contribution to the thesis

Seltmann KC, Péntzes Z, Yoder MJ, Bertone MA, Deans AR (2012) Utilizing Descriptive Statements from the Biodiversity Heritage Library to Expand the Hymenoptera Anatomy Ontology. PloS ONE. (*accepted*, December 29, 2012). [Impact factor: 4.092].

Seltmann KC, Yoder MJ, Mikó I, Forshage M, Bertone M, et al. (2012) A hymenopterists' guide to the Hymenoptera Anatomy Ontology: utility, clarification, and future directions. Journal of Hymenoptera Research 27: 67. [Impact factor: 0.531]

Yoder MJ, Mikó I, Seltmann KC, Bertone MA, Deans AR (2010) A gross anatomy ontology for hymenoptera. PloS ONE 5: e15991. [Impact factor: 4.411].

Other publications contributing to the thesis

Bertone M, Blinn R, Stanfield T, Dew K, Seltmann KC, et al. (2012) Results and insights from the NCSU Insect Museum GigaPan project. ZooKeys 209: 115–132. [Impact factor: 0.879].

Bertone MA, Mikó I, Yoder MJ, Seltmann KC, Deans AR (2012) Matching arthropod anatomy ontologies to the Hymenoptera Anatomy Ontology: Results from a manual alignment. Database. (*accepted*, November 27, 2012). [Impact factor: 3.179].

Mikó I, Friedrich F, Yoder MJ, Hines HM, Deitz LL, et al. (2012) On dorsal prothoracic appendages in treehoppers (Hemiptera: Membracidae) and the nature of morphological evidence. PloS ONE 7: e30137. [Impact factor: 4.092].

Mikó I, Yoder MJ, Bertone MA, Seltmann KC, Deans AR (2011) Multi-Species Anatomy Ontology Development Requires a Pluralist Approach to Label_Class

Mapping. Proceedings of the International Conference on Biomedical Ontology. Buffalo, NY, Vol. 0. pp. 8–10.

Penev L, Sharkey M, Erwin T, Noort S van, Buffington M, et al. (2009) Data publication and dissemination of interactive keys under the open access model. *ZooKeys* 21: 1–17. [Impact factor: 0.514].

Sharkey MJ, Yu DS, Noort S, Seltmann KC, Penev L, et al. (2009) Revision of the Oriental genera of Agathidinae (Hymenoptera, Braconidae) with an emphasis on Thailand including interactive keys to genera published in three different formats. *ZooKeys*: 19–54. [Impact factor: 0.514].