Ph.D. Dissertation

Advancement in Hymenoptera Research through Anatomy Ontology Development

Katja C. Seltmann

Supervisors: Dr. Andrew R. Deans[1] and Dr. Zsolt Pénzes[2]
[1]Department of Entomology, Pennsylvania State University, 501 ASI Building, University Park, PA, USA
[2]Department 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
2012

Szeged

TABLE OF CONTENTS

ABBREVIATIONS

HAO - Hymenoptera Anatomy Ontology

NLP - Natural Language Processing

JHR - Journal of Hymenoptera Research

URI - Uniform Resource Identifier

LSID - Life Science Identifier

GUID - Globally Unique Identifier

OCR - Optical Character Recognition

MySQL - My Structured Query Language

OWL - Web Ontology Language

OBO - Open Biomedical Ontology

CLSM - Confocal Laser Scanning Microscopy

EST - Expressed Sequence Tag

XML -Extensible Markup Language

PURL - Persistent Uniform Resource Locator

BFO - Basic Formal Ontology

CARO - Common Anatomy Reference Ontology

RO - OBO Relationship Types or Relations Ontology

BSPO - Spatial Ontology

UBERON - UBER Anatomy Ontology

GO - Cellular Component Ontology

PLoS - Public Library of Science

SVG - Scalable Vector Graphics

FIGURES and TABLES

1. INTRODUCTION

1.1. Hymenoptera Phylogeny

Bees, wasps, ants, gall wasps, and sawflies all belong to an extraordinarily diverse lineage of insects, now referred to as Hymenoptera. Hymenoptera is considered to be the sister group to the remaining Holometabola (Savard et al., 2006; Wiegmann et al., 2009) with strongly supported monophyly, and with eighteen known autapomorphies (Sharkey, 2007; Vilhelmsen, 2001, 2007). The order has traditionally been divided into a series of informal groupings, Symphyta and Apocrita (Aculeata + Parasitica). Informal groups are not necessarily monophyletic, but the terminology is often useful and is commonly used for the purpose of discussion, due to their members' similar ecology and morphology. The basal lineages of Hymenoptera are informally known as the Symphyta. Apocrita are distinguished from these basal lineages by their possession of the "wasp-waist" (Mason & Huber, 1993). This narrow constriction between the propodeum and second abdominal segment allows for great range of motion for the metasoma, which is thought to be the primary adaptation responsible for the parasitic and stinging lifestyle. Comprising the Apocrita, the stinging Hymenoptera continue to be informally placed in a grouping known as the Aculeata (=Vespomorpha) and the non-stinging forms into the Parasitica (=Terebrantia) (Sharkey, 2007).

Rasnitsyn (1988, 1969, 1980) and Königsmann (1977a, 1977b, 1978a, 1978b) first attempted to outline an evolutionary hypothesis for Hymenoptera. Rasnitsyn's landmark 1988 paper remained the foundation for all subsequent analysis, until Ronquist, Rasnitsyn, Roy, Eriksson, and Lindgren (1999) published a reanalysis of the Rasnitsyn paper. Sharkey and Roy (2002) critically reanalyzed and recoded the wing characters from the Ronquist et al. paper, and showed how the original phylogenetic hypothesis was dependent on coded wing characters.

While results from these early morphological papers consistently support relationships in the Parasitica, including the monophyly of the Ichneumonidae + Braconidae, Ibaliidae + Figitidae + Cynipidae, Megaspilidae + Ceraphronidae and Chalcidoidea + Mymarommatidae, these works also illustrate how little is known about apocritan relationships at the family and superfamily level. Potentially polyphyletic superfamilies listed in these papers are the Evanioidea and Proctotrupoidea, and of the hymenopteran families recognized, 20 were questionably monophyletic, eight of which belong to the Chalcidoidea. These include the Agaonidae, Andrenidae, Aphelinidae,

Argidae, Aulacidae, Colletidae, Diapriidae, Megaspilidae, Melittidae, Pergidae, Perilampidae, Proctotrupidae, Pteromalidae, Roproniidae, Scelionidae, Tenthredinidae, Tetracampidae, Torymidae, and Tanaostigmatidae (Carpenter, Engel, Sharkey, & Heraty, 2003). Generally, among the aculeates, Chrysidoidea, Apoidea and Vespoidea superfamily relationships are well resolved. The family limits within the Chrysidoidea are supported. The Pompilidae are generally considered to belong to the Vespoidea; however, their placement is not resolved within the superfamily. Morphological indications even suggest that they may belong to the Apoidea (Brothers & Carpenter, 1993), although the monophyly of Pompilidae itself is well supported. As a general rule the family-level relationships in the Apoidea are largely unresolved. Sternotritidae and Melittidae appear in multiple places within the tree (Brothers, 1975), and the family-level relationships among the specoid wasps are debated. In 1997, Melo recognized five families of Apoidea: Ampulicidae, Apidae, Crabronidae, Specidae and Heterogynidae. This resolved the superfamily by placing all bees into the family Apidae. Michener (2007), considered to be the leading authority on bees, split the Apoidea into ten families: Sternotritidae, Colletidae, Andrenidae, Halictidae, Melittidae, Megachilidae, Apidae, Sphecidae, Crabronidae and Ampulicidae. Symphyta, although an informal grouping basal to the Apocrita is recognized as a paraphyletic grade, however the phylogeny at the family level is mostly resolved, through research by Vilhelmsen (2001) and Schulmeister (2003a, 2003b). The sister group to the Apocrita remains to be identified, although it is commonly concluded to be the Orussidae, which is the only extant symphytan with a parasitic lifestyle.

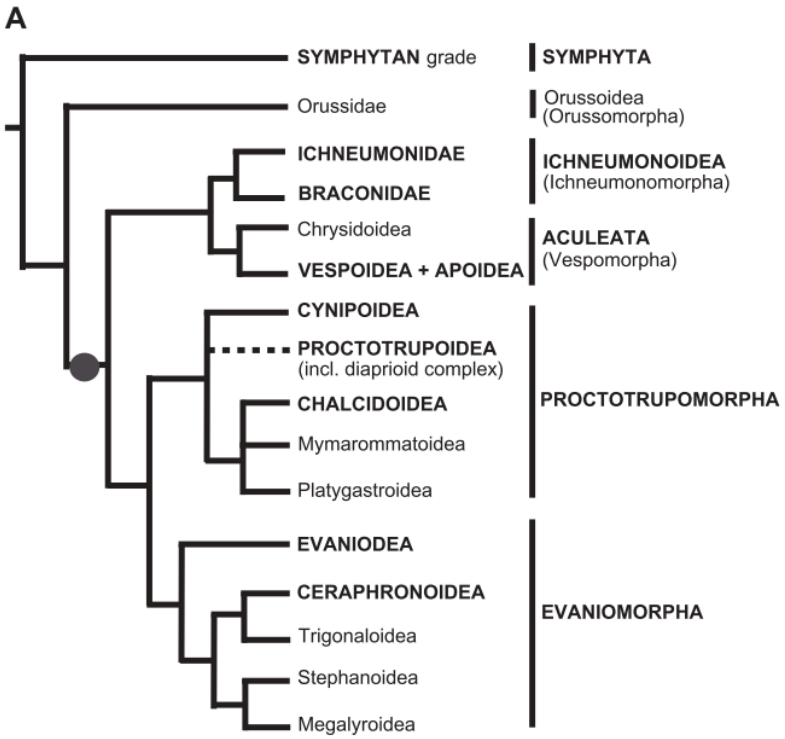

Figure 1: Summation tree of early studies from Sharkey (2007).

Early molecular studies were restricted to ribosomal and mitochondrial DNA markers, with considerable taxon sampling, and without robust resolution (Carpenter & Wheeler, 1999; Castro & Dowton, 1994, 1999, 2006, 2007; Dowton & Austin, 2001). Dowton and Austin indicated that the superfamily Proctotrupoidea was paraphaletic, with Roproniidae + Pelecinidae as sister to the Cynipoidea and Evanioidea (as Aulacidae, Gasteruptiidae, and Evaniidae) was polyphyletic. Sharkey (2007) is critical of the results, pointing out that Orussidae was recovered as a highly derived Ichneumonoid and the Aculeate + Ichneumonoidea sister group relationship was not recovered, however, both placement of Orussidae as sister to the Apocrita and the sister group relationship of Aculeate + Ichneumonoidea has long been an expected result in analysis, although not unequivocally supported. Castro and Dowton (2006) continued the work with 16S, 28S, 18S, and CO1. Again, Proctotrupoidea was recovered as paraphaletic and Evaniomorpha included Ceraphronoidea (figure 2). The sister group relationship of Aculuata and Ichneumonoidea was recovered, but with weak support, continuing to put the hypothesized relationship in doubt.

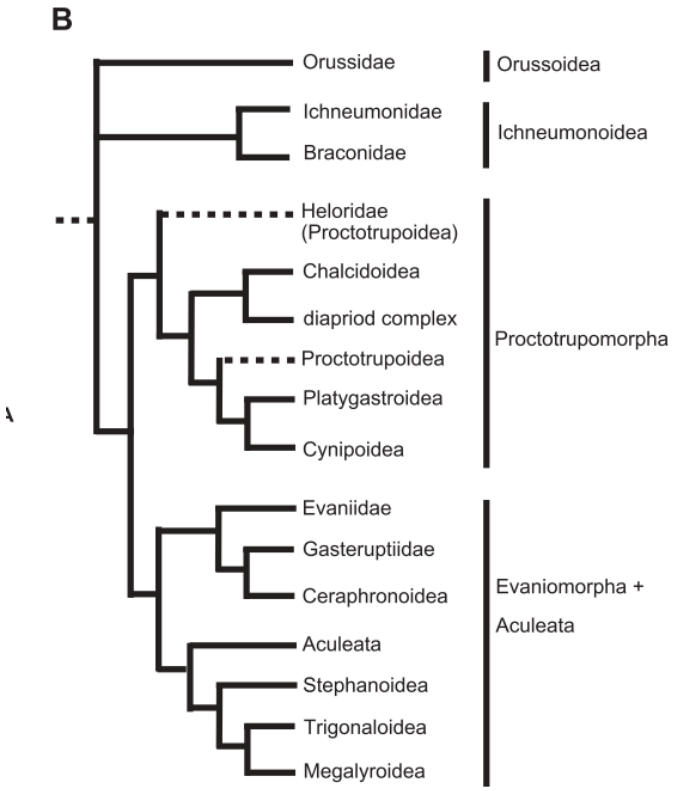

Figure 2: Tree from Castro and Dowton (2006), as summarized by Sharanowski et al. (2010). Dashed lines indicate paraphyly.

In 2003, the funding of the project "Building the Hymenopteran Tree of Life: Large Scale Phylogenetic Analysis of the Hymenoptera" (EF-0337220), resulted in a number of papers that attempted to resolve the issue of higher-level Hymenoptera relationships. Sharanowski et al. (2010) found Proctotrupomorpha (Cynipoidea + Proctotrupidea, minus Chalcidoidea) to be the sister to the Aculeata. Proctotrupoida itself has been found to be paraphyletic, and Evaniodiea falls within Ceraphronoidea (figure 3). Although the study was limited by taxon sampling, the potential for expressed sequence tags (ESTs) and genomic information to resolve these contentious relationships in the Hymenoptera was demonstrated, and followed the state of the art methodology of utilizing a high number of genes toward a phylogenomic approach (Rokas, Williams, King, & Carroll, 2003).

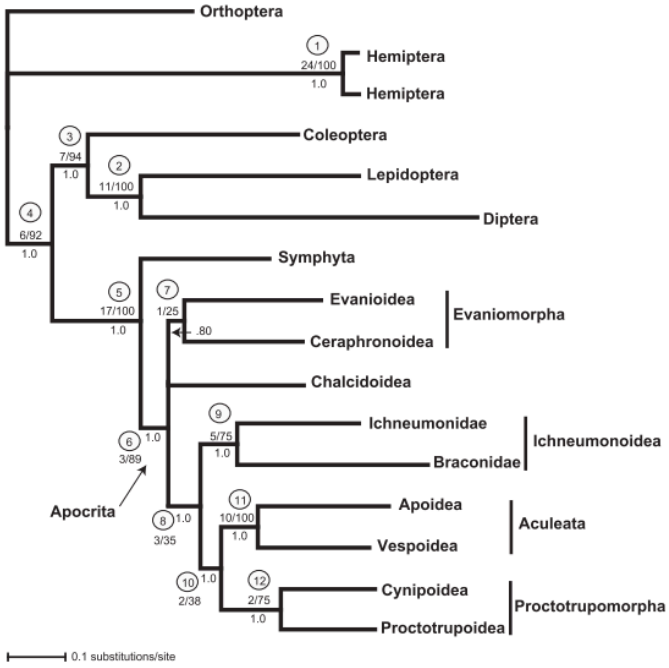

Figure 3: Tree from Sharanowski et al. (2010).

Vilhelmsen, Mikó, and Krogmann (2010) revisited morphology with a comprehensive study of the mesosoma in apocritan wasps, revealing a paraphyletic Proctotrupoidea and Evanioidea with inclusion of Trigonaloidea. Sharkey et al. (2012), included more complete higher-level taxon sampling that before, scored for 392 morphological characters, with only one new gene (EF-1α). All currently recognized superfamilies were recovered as monophyletic and the diapriid subfamily Ismarinae was raised to family status. The Aculeata + Ichneumonoidea relationship was not supported.

A few publications have attempted to create a 'best-guess' summary of the known relationships, deriving their data from either SuperTree methods (Davis, Baldauf, & Mayhew, 2010), or hypotheses based on non-statistical evaluations of the literature and personal experience (Sharkey, 2007), however, these studies preceded the recent publications of Sharkey et al. (2012) and Munro et al. (2011). Figure 1 summerizes the Sharkey (2007) hypothesis.

In summary, results examining family and superfamily level relationships tend to be highly variable, and lend support to a view that relationships are relatively weak and are dependent upon the analysis used to infer them. This lack of stability among superfamily and family level relationships, across independent datasets, leaves ample doubt for many of the well-known assertions regarding higher level hymenoptera

relationships (Sharanowski et al., 2010) and the summary of our present understanding of Hymenoptera is not easy to visualize. The monophyly of some families is still called into question, particularly the Andrenidae, Aphelinidae, Chalcididae, Colletidae, Eucharitidae, Eupelmidae, Eurytomidae, Megaspilidae, Melittidae, Perilampidae, Proctotrupidae, Pteromalidae, Tenthredinidae, Tetracampidae, and Torymidae (Munro et al., 2011; Sharkey, 2007). Although the present understanding of superfamilies was recovered in Sharkey et al, 2012, support for the following remains tenuous: Proctotrupoidea (as Austroniidae, Heloridae, Pelicinidae, Perandeniidae, Proctotrupidae, Proctorenyxidae, Roproniidae, Vanhorniidae), Evanioidea (as Evaniidae, Gasterupiidae, and Aulacidae), and Vespoidea (as Bradynobaenidae, Formicidae, Mutillidae, Pompilidae, Rhopalosomatidae, Sapygidae, Scoliidae, Sierolomorphidae, Tiphiidae, and Vespidae). Diaprioidea (Diapriidae, Monomachidae, and Maamingidae) is considered a separate superfamily and no longer part of Proctotrupoidea, and Platygastroidea now only comprises one family, Platygastridae (Sharkey, 2007).

1.2. The Study of Hymenoptera Utilizing the Semantic Web

Many open questions still exist in Hymenoptera studies, despite considerable recent progress. More than 140,000 species of Hymenoptera have been described (Mason & Huber, 1993) by generations of morphologists, primarily using prosaic and natural language. The guidelines for how these descriptions were created evolved via mentorship between students and their teachers, and the requirements of journals in which they were published. Contained within the descriptions is a body of Hymenoptera anatomical information that is important for the understanding of present day hymenopteran phylogenetic hypothesis, and for modern morphological research in general. As the number of new descriptions continues to expand, it becomes increasingly for researchers to incorporate these articles, and the descriptive statements contained in them. Discovering the meaning behind terms used in prosaic descriptions can be elusive, because new terminology is often inadequately illustrated, and comprehensive study requires a good deal of exploration for new students to fully realize the meaning that the describing author intended to convey. In addition, there is no way to synthesize prosaic descriptions easily as there is not a uniform process through which such descriptive statements are made, and, at present, no automated utility exists 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, it would be necessary to review all of the literature on braconids. This would involve locating the relevant articles based on taxon, using Google Scholar, Web of Science, Scopus, or other search engines, then reading or text searching each article for suitable keywords, such as 'ocelli' or 'eye'. At the present time, a single search does not exist that would retrieve this information (i.e., articles and all references to eye) in a semi-automated way. Thus, such simple question may require months of research to answer, and the search may leave many important articles overlooked. In addition, this method of literature search tends to result in searches that are confined to the specific taxon of interest, to the neglect of articles on other insect groups that may also be relevant to the researchers' morphological interests.

Hymenoptera studies is in an interesting position, with an estimated half a million species remaining to be described (Gaston, 1991), and many relationships among families still tenuous (see chapter *1.1. Hymenoptera Phylogeny*). The body of descriptive work already accumulated is minimal compared to the number of articles that may yet be produced. At the same time, the semantic Web, or a Web of Internet data using http protocols, and interconnected through the use of data identifiers and controlled vocabularies, is already a reality, and is in active use in contemporary research. An example of a research article making use of the semantic Web is a publication that contains active links to GenBank sequence data. Each such link would refer to the GenBank Web interface, where information about the DNA sequence and the author of the publication may be retrieved, along with links to information regarding the taxonomy of the organism. The latter may be expected to provide further links to the primary literature documenting the organism. Thus, through multiple, linked sources of information, the original snippet of data highlighted in the publication are semantically enhanced through multiple resolvable Web links, and specialized databases of information.

Hymenopterists now have the potential to improve on past methods of describing species by incorporating similar semantic, repeatable, and machine understandable techniques into our research methods. The work documented in this dissertation has advanced that state of affairs by contributing to the creation of a structured, controlled vocabulary of hymenoptera terminology, which is the Hymenoptera Anatomy Ontology (HAO).

Stated simply, an ontology is a set of concepts (definitions) used to model a formalized domain (in the case of interest, that domain is Hymenoptera anatomy), and the logical relationships between concepts. The goal of ontology creation is to enable

computer-based reasoning about morphological concepts (linked to terms) 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 as: **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**, one can conclude 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 just terms that represent concepts. A concept is the 'real life' physical structure on the wasp that one wants 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 these terms as all synonyms of the concept, increasing the power of the ontology to decipher descriptive language. Further explanation of anatomy ontology and Hymenoptera Anatomy Ontology construction can be found in chapter *3.1 Methods of Hymenoptera Anatomy Ontology (HAO) Construction*, below.

The increased clarification that is gained through the adoption of precisely defined anatomical terminology is only one benefit of the use of ontologies; other benefits exist that also suggest going beyond a simple glossary. As described in detail in Deans et al. (2012), the incorporation of concepts from an anatomy ontology into species descriptions increases their use for the greater scientific community and creates a corpus of semantic statements about biodiversity that can be used in computer reasoning. The incorporation of ontology is not new to biological sciences and 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). Simultaneous with these efforts, new open access publications are emerging, in which intelligent semantic markup is 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 an object in an explicit fashion. Public Library of Science (PLoS) is a nonprofit organization committed to making scientific literature available online published under a Creative Commons License, including highly competitive biology journals. 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 supplies it with a GUID. All of the PLoS and

ZooKeys articles are available freely online as Optically Recognized PDFs. The Biodiversity Heritage Library 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 loosening the restriction that new organism descriptions are required to be printed in paper copy journal (International Commission on Zoological Nomenclature, 2012).

Making relevant literature freely available (on the Web and without subscription) is the first step toward increasing the discoverability of descriptive data. The PLAZI (Agosti & Egloff, 2009) project is working to retroactively annotate the historical literature in biological sciences by developing applications such as the Golden Gate Editor (Agosti & Egloff, 2009; Sautter, Böhm, & Agosti, 2007), which makes old descriptions searchable. GoldenGate transforms the text of entire free-text articles into an Extensible Markup Language (XML) structured documents, following TaxonX schema. Marking up old literature in this manner allows their entire descriptive text to be searched and retrieved, through queries applied to article information based on taxon name or collecting events, or inclusion of GUIDs or LSIDs. The markup is considered to be semi-automated: it requires human input, but it is greatly aided by computer algorithms that can assist in discovering similarities found in sections of taxonomic literature. Articles marked up in this manner are then deposited in a custom database, the Search and Retrieval Server (SRS)(Plazi, 2012), where the enhanced searching capability of the marked-up articles can be utilized. Other software programs, TaxonFinder (Leary, Remsen, Norton, Patterson, & Sarkar, 2007) and NetiNeti (Akella, Norton, & Miller, 2012), utilize a well-known controlled vocabulary – taxon name lists – to discover relevant BHL articles for a user by utilizing this controlled vocabulary.

Finding relevant literature is only one potential for semantic (=linked) technologies. There is also the potential for new scientific discoveries based on well-defined descriptive statements. The Phenoscape project (Balhoff et al., 2010; Dahdul 2010a; Dahdul et al., 2010b; Mabee et al., 2007) has advanced the use of anatomy ontology to extract Entity-Quality statements from legacy literature using Phenex software. These statements are then used to apply machine reasoning across the literature, inferring potential genetic pathways for phenotype characteristics. CharParser (Cui, 2012) software also annotates legacy descriptive statements, but specifically as a semi-automated system, without the necessity of a prior ontology, and with less emphasis on domain expertise. CharParser develops an independent glossary during the text mining and training process, and attempts to reduce the amount of user

input is necessary to infer meaning from the statements. In Hymenoptera, as in much of Insecta, the majority of the descriptive work still remains to be completed. This creates the opportunity for new tools to be developed, and for the incorporation of descriptions that are a priori semantically annotated (Deans, Yoder, & Balhoff, 2012; Mullins, Kawada, Balhoff, & Deans, 2012).

In general, all of these efforts indicate a general trend toward highly accessible, semantic, and discoverable forms of publication, where researchers can find relevant articles and information easily, coupled with anatomical ontologies, to make sense of all the terminology contained within those texts. Applying ontology to taxonomy takes advantage of these new trends, by making descriptive statements that are more relevant and are better utilizable by different scientific disciplines, whether by new prospective students of Hymenoptera, by researchers engaged in genomic discovery, or in many purposes not yet conceptualized.

1.3. Hymenoptera Anatomy Ontology Project, Participants, and History

There have been numerous attempts to clarify relevant anatomy of the Hymenoptera, spanning from comprehensive anatomical treatments of character systems across Hymenoptera (e.g., Oeser (1961) for the ovipositor system, Gibson (1985) and Vilhelmsen et al. (2010) for thoracic structures, Vilhelmsen (1996) for preoral cavity in lower Hymenoptera, and Schulmeister (2001) for male genitalia), to more focused taxonomic treatments that cover anatomy at a relatively small scale (e.g., Sharkey and Wharton (1997) for Braconidae, Gibson (1997) and Gibson, Read, and Fairchild (1998) for Chalcidoidea, Bolton (1994) for Formicidae, and Michener (2000) for Apoidea). There has been proportionally little effort, with few exceptions (Richards, 1997; Vilhelmsen et al., 2010) to unify our collective knowledge of hymenopteran anatomy in ways that are both deeply anatomical and broadly taxonomic.

In 2006, the Hymenoptera Anatomy Ontology (HAO; Deans & F Ronquist, 2006) was proposed at the 6th International Congress of Hymenopterists meeting in South Africa. Afterward, Dr. Andrew R. Deans and Dr. Matthew J. Yoder developed a prototype Web-based, collaborative ontology-editing interface and preliminary terms and concepts were collected from well-known Hymenoptera glossaries and resources. In 2008, Dr. Andrew R. Deans and Dr. Matthew J. Yoder used these initial efforts to secure funding from the United States National Science Foundation's Advances in Biological

Informatics program (grant #DBI-0850223). The funding of the HAO assembled an expert group of morphologists, bioinformatists, and students who significantly contributed to all aspects of the project including: Dr. Matthew J. Yoder (co-PI), Dr. István Mikó, Dr. Andrew R. Deans (PI), Dr. Matthew Bertone, Jim Balholf, Andrew Ernst, Patricia Mullins, and the author of this dissertation.

1.4. Other Arthropod Anatomy Ontologies and Alignment

Presently, anatomy ontologies for five other arthropod groups exist on the OBO Foundry (Smith et al., 2007). These are: spiders [Arachnida: Araneae; SPD; (Ramírez et al., 2007)], ticks [Arachnida: Ixodida; TADS; (Topalis et al., 2008)], mosquitoes [Insecta: Diptera: Culicidae; TGMA; (Topalis et al., 2008)], and *Drosophila melanogaster* [Insecta: Diptera: Drosophilidae; FBbt; (Drysdale, 2001)]. The size, or number of concepts in the ontology, varies from 552 (SPD) to 6,884 (FBbt). The variability and scope of each depends on its intended purpose and audience. For example, the HAO, was developed to aid in standardizing the meaning of anatomical concepts used by taxonomists to describe Hymenoptera, while also providing a way to reason across descriptive text. Thus, a highly granular level of detail is needed, including terms related to muscle attachments. The remaining arthropod ontologies have other purposes, including annotating vector genomes (TGMA and TADS) and classifying images for specific phylogenetic characters (SPD). The premise exists that basic phenotypic information can be shared across taxa through alignment of their corresponding anatomy ontologies. The resulting linkages can facilitate the transfer of knowledge between these domains, allowing for genetic-based phenotype hypothesis to develop. However, the great deal of variability between existing ontologies demands that a mapping between concepts of disparate ontologies be created, which is a challenge. Ontology matching, or creating links between varying existing ontologies, is generally accomplished through semi-automatic means (Mungall, Torniai, Gkoutos, Lewis, & Haendel, 2012). However, Bertone, Mikó, Yoder, Seltmann, and Deans (2012), using the HAO as the principle model to align against, demonstrated the effectiveness and increased accuracy of alignment by domain experts, supporting the notion that domain expertise is essential to the process of anatomy ontology development.

2. AIMS

The aims of the work contained in this dissertation are to aid in the development of the Hymenoptera Anatomy Ontology (HAO) and to demonstrate its utility in guiding modern morphological and taxonomic research. Specifically, the primary 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: *Utilizing the HAO for Literature Analysis*).

2. To analyze the descriptive terminology in Hymenoptera literature using Natural Language Processing (NLP) clustering methods and compare the results to our present understanding of Hymenoptera phylogenic relationships (subheading: *Utilizing the HAO for 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: *Utilizing the HAO for Literature Annotation*).

3. MATERIALS AND METHODS

3.1. Methods of Hymenoptera Anatomy Ontology (HAO) Construction

An ontology is a formalized system of concepts and their logical relationships. When the author and her collaborators developed the HAO, they followed the Open Biomedical Ontology (OBO) format. The specifications for the OBO file format were to use with OBO Edit software (N. Harris, Day-Richter, Mungall, Abdulla, & Deegan, n.d.), the preferred software for many anatomy ontology projects. We implemented OBO specifications for ontology construction in the database software mx (Mx, 2012), in which the HAO continues to be updated and maintained. The final versions of OBO files are released to the public via the OBO Foundry (Smith et al., 2007) only after further vetting (in the OBO Edit software), which is required because the OBO-EDIT software contains many error and logic checking stages that have not yet been implemented in mx. The use of the OBO file format proved worthwhile because it allowed us to take advantage of the Web for collaborative development – a novel approach to ontology development. OBO Edit, the software typically employed for OBO format ontology development, is a desktop application, and collaboration is traditionally achieved through Internet discussion lists.

The major components of the Hymenoptera Anatomy Ontology are labels (=terms; words used to refer to anatomical structures in the context of the ontology), concepts (= classes; the anatomical structures themselves, or the real 'thing'), identifiers (the unique number that refers to the concept), the name or preferred term for the concept, and relationships (how concepts are related with each other).

3.1.1. Properties of Relationships

An ontology can have any number of kinds of relationships. These relationships are included in a separate ontology, or the Relationship Ontology (RO), which helps maintain standard relationships across all ontologies, an important factor when we try to merge, or align different ontologies. We designed the HAO to incorporate the following relationships: is_a, part_of, integral_part_of, attached_to (used for muscles), and is_obsolete.

1. **is_a**: This is a relationship of inheritance; if A is_a B, and B is_a C, then A is_a C (A has all the properties of B and C; B has the properties of C, but not A). For

example, HAO:0000939 is the concept for label 'sitopore', defined as "The sclerite that is located in the proximal part of the hypopharygeal wall delimited distally by the functional mouth and proximally by the proximal boundary of the cibarium." Since we know 'sitopore' is_a sclerite (HAO:0000909), which is_a 'area', we can infer that the sitapore is_a 'area' (HAO:0000146). Logically the 'sitopore' must satisfy the definition of 'area' as well as 'sclerite' for the is_a relationship to be valid. The OBO formatted HAO entries for sitapore and area are below in figure 4.

2. **part_of**: C part_of C1 defines a relational property of permanent part-hood for Cs. It tells us that Cs, whenever they exist, exist as parts of C1s. This relation satisfies at least the following standard axioms of mereology: reflexivity (for all p, p part_of p); anti-symmetry (for all p, p1, if p part_of p1 and p1 part_of p then p and p1 are identical); and transitivity (for all p, p1, p2, if p part_of p1 and p1 part_of p2, then p part_of p2) (Malpas & Zalta, 2012). Analogous axioms hold also for part-hood as a relation between spatial regions. Thus, since the 'sitapore' is part of the hypopharyx, we can conclude that it is also part of the mouthparts, as the hypopharyx is inclusively part_of mouthparts. The OBO formatted HAO entries are given below, in figure 4.

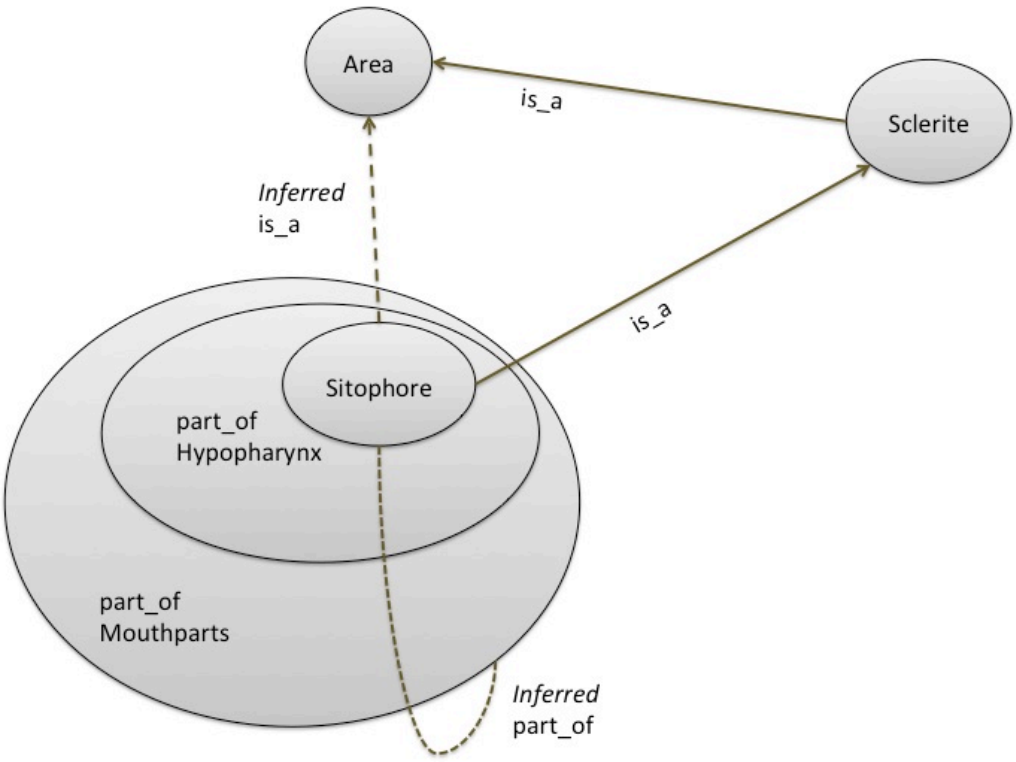

Figure 4: Inference from is_a and part_of relationships from the HAO.

3. **integral_part_of**: This relationship is used to describe situations where the child class is always part of the parent class (i.e., for the parent class to exist it must possess the constituent part). At the present time the only example in the HAO for this class is flagellum, although this will likely be extended for other segmentation definitions. The OBO formatted HAO entries are presented below, in figure 5.
4. **attached_to**: This relationship is, so far, unique to the HAO and is used to describe where muscles attach. For example, the posterior cranio-mandibular muscle (HAO:0000745) is attached_to the cranium [HAO:0000234] and is attached_to the mandible (HAO:0000506). This relationship does not specify site of muscle origin or insertion. The OBO formatted HAO entries are listed below, in figure 5.
5. **is_obsolete**: This relationship is reserved for HAO concepts that were given an identifier, and subsequently edited. Since concepts with identifiers may be included in publications, they need to be persistent. HAO concepts, once designated to be 'official' (i.e. given an identifier such as HAO:0000939), they will never be deleted from the ontology.

```
[Term]
id: HAO:0000146
name: area
def: "The anatomical structure that is delimited by material or
immaterial anatomical entities." [http://api.hymao.org/api/ref/67791]
is_a: HAO:0000003 ! anatomical structure

[Term]
id: HAO:0000348
name: fore coxa
comment: Synonymous concept.
is_obsolete: true

[Term]
id: HAO:0000361
name: functional mouth
def: "The anatomical space that is delimited posteriorly by the
distal part of the sitophore and anteriorly by the tormae."
[http://api.hymao.org/api/ref/78244]
synonym: "mouth" [http://api.hymao.org/api/ref/68619,
http://api.hymao.org/api/ref/36875]
is_a: HAO:0000005 ! anatomical space
relationship: OBO_REL:part_of HAO:0000809 ! preoral cavity

[Term]
id: HAO:0000342
name: flagellomere
def: "The annulus that is located distally of the pedicel."
[http://api.hymao.org/api/ref/67791]
```

```
synonym: "flagellar segment" RELATED
[http://canacoll.org/Hym/Staff/Gibson/apss/chalglos.htm]
synonym: "flagellar subdivision" RELATED
[http://api.hymao.org/api/ref/36927]
is_a: HAO:0000096 ! annulus
relationship: OBO_REL:0000004 HAO:0000343 ! integral_part_of
flagellum

[Term]
id: HAO:0000409
name: hypopharynx
def: "The area that is delimited proximally by the proximal margin of
the sitophore, distally by the salivarial orifice."
[http://api.hymao.org/api/ref/78244]
synonym: "hypopharyngeal wall" RELATED
[http://api.hymao.org/api/ref/78243]
is_a: HAO:0000146 ! area
relationship: BFO:0000050 HAO:0000639 ! part_of mouthparts

[Term]
id: HAO:0000745
name: posterior cranio-mandibular muscle
def: "The mandibular muscle that arises posterodorsally from the
cranium and inserts on the tendon attached anteroproximally on the
mandible." [http://api.hymao.org/api/ref/78258] synonym: "M.
craniomandibularis internus" [DOI:10.1016/j.ode.2006.06.003] synonym:
"adductor of the mandible" [http://api.hymao.org/api/ref/36927]
synonym: "anterior cranio-mandibular muscle"
[http://api.hymao.org/api/ref/78243]
is_a: HAO:0001779 ! mandibular muscle relationship: HAO:attached_to
HAO:0000234 ! cranium
relationship: HAO:attached_to HAO:0000506 ! mandible

[Term]
id: HAO:0000909
name: sclerite
def: "The area of the integument where the cuticle is well
sclerotised with thick exocuticle and is surrounded by conjunctivae."
[http://api.hymao.org/api/ref/67791]
synonym: "plate" RELATED [http://api.hymao.org/api/ref/68619]
synonym: "sclerome" RELATED [http://api.hymao.org/api/ref/78428]
is_a: HAO:0000146 ! area
relationship: BFO:0000050 HAO:0000421 ! part_of integument


[Term]
id: HAO:0000939
name: sitophore
def: "The sclerite that is located in the proximal part of the
hypopharygeal wall delimited distally by the functional mouth and
proximally by the proximal boundary of the cibarium. The tentorio-
hypopharyngeal muscle inserts on the proximal margin of the
sitophore." [http://api.hymao.org/api/ref/78243]
is_a: HAO:0000909 ! sclerite
relationship: BFO:0000050 HAO:0000409 ! part_of hypopharynx


[Term]
id: HAO:0001122
name: procoxa
```

```
def: "The coxa that is located on the fore leg."
[http://api.hymao.org/api/ref/67791]
synonym: "fore coxa" RELATED [http://api.hymao.org/api/ref/67791]
synonym: "forecoxa" RELATED [http://api.hymao.org/api/ref/43472,
http://api.hymao.org/api/ref/67791]
is_a: HAO:0000228 ! coxa
relationship: BFO:0000050 HAO:0000349 ! part_of fore leg
```

Figure 5: Example OBO format file from HAO.

3.1.2. Genus Differentia

Genus Differentia is a logically constructed type of definition designed to first describe an inclusive class of concepts (genus) and subsequently to describe the characteristics differentiating (differentia) it from other children of that concept. Definitions in this format typically follow the pattern '**B** is an **A** that **X'** (B Smith, 2005). Each component of a Genus Differentia definition is defined where (**B**) is the **definiendum** (the label being defined)**,** (**A**) is the **genus** (the broader category for that label or the difiniendum's parent), and (**X**) is the **differentia** (how that label differs from the genus's other children). For example, in the definition for hypopharynx, "A hypopharynx is the area that is delimited proximally by the proximal margin of the sitophore, distally by the salivary orifice", "hypopharynx" is the definiendum, "area" is the genus, and "proximally by the proximal margin of the sitophore, distally by the salivarial orifice" is the differentia; see figure 6 (below).

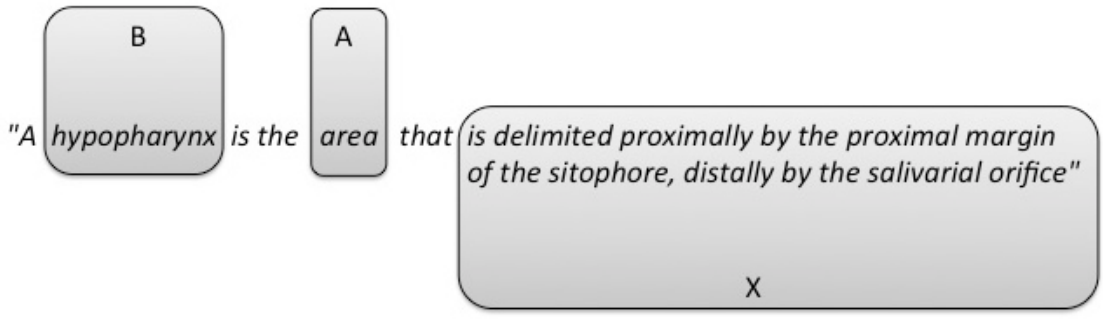

Figure 6: Genus Differentia definition for 'hypopharynx'.

Definitions structured in this manner inherit the definitions of all the concepts contained within the definition and the concepts it is related to. This assists authors, who themselves cannot easily explicitly define all the terms used in a publication, as their definitions will inevitably include labels that must also be defined. To precisely

define one concept requires an entire ontology and one that is internally consistent (i.e. only uses labels that are also defined in the ontology). For example in the definition of 'hypopharynx', the labels 'sitophore', 'salivarial orifice' and 'area' also need to be defined. However, in the definitions of each of these labels are labels that again need defining, creating a cascading effect from the interconnectedness of the labels contained within definitions (figure 7).

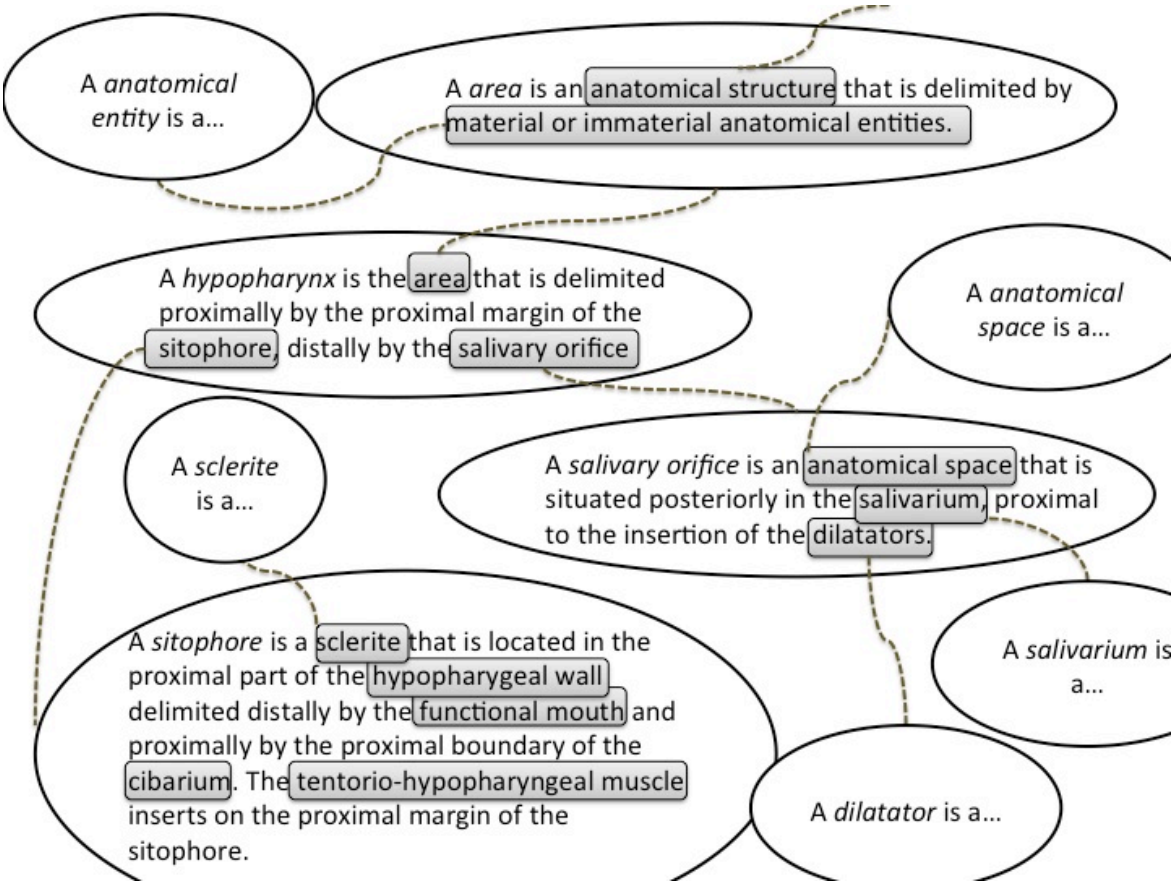

Figure 7: Definition inheritance in anatomy ontology.

3.1.3. Separation of Labels and Concepts

Labels and concepts are treated separately, and only concepts are given unique identifiers. Labels are considered to be simply strings of letters that are used to reference concepts by authors. In other words, concepts encapsulate meaningful anatomical observations, where labels are only words we use to represent that meaning. Label synonymy exists when different observers, looking at the same anatomical structure, have the same conceptualization of it but use different terms to represent their concept. In contrast, homonymy may exist if different observers use the same label

for different anatomical concepts (Yoder, Mikó, Seltmann, Bertone, & Deans, 2010).

The separation of concept from label is another important construct for understanding how the HAO is useful for clarification of terminology. In the context of the ontology itself, 'terms' are called 'labels'. The reason for the distinction is that, in publication, a term has greater inherent meaning, discussed in the context of the article by the author. This meaning often extends beyond the concept captured by the HAO, and written in the definition in which the 'label' is now associated. Now, in the context of the ontology, the complex term only exists only as a string of letters as representative of the ontology concept, not wholly representative of the authors' original intention (see discussion of homology and structurally equivalency below). `

Separating label from concept (=sensu system) in the ontology, allows for the identification of synonyms and homonyms from the intersection between 'label' (=term found in the ontology), 'reference', and 'concept'. Thus, in figure 8 the labels for A and B are synonyms, since two different labels (L1 & L3) describe the same class (C1), in different references (R2 & R1, respectively); A and C are also synonyms. The act of synonymy by an author can be seen in B and C, where two different labels (L2 & L3) were used for one class (C1) in the same reference (R1). B and D represent homonymy, where the same label (L3) is used to describe two different classes (C1 & C2) in different papers (R1 & R2).

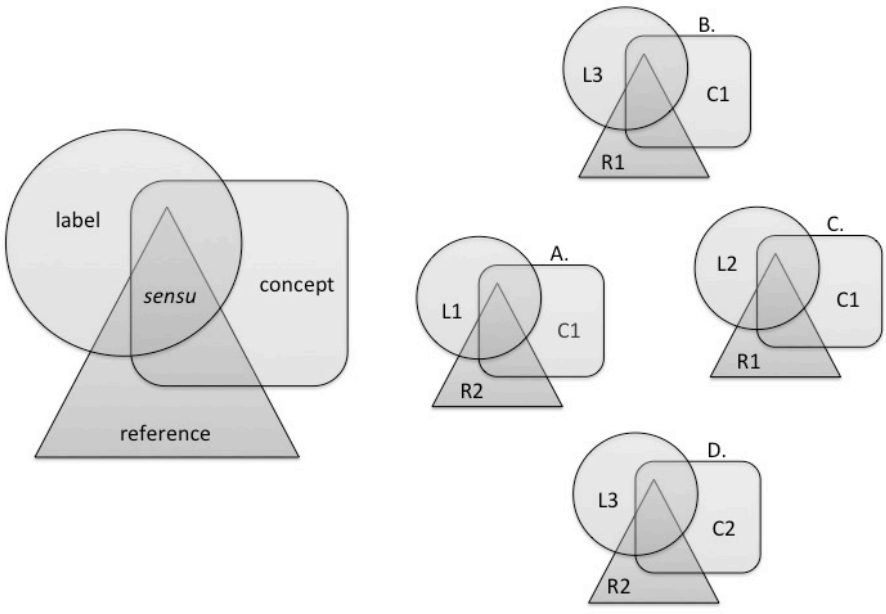

Figure 8: Sensu system; from Ontology Primer (Bertone, Yoder, Seltmann, Mikó, & Deans, 2010).

3.1.4. Preferred Terms

Because labels and concepts are separated in the HAO, it is not necessary to come to a general community agreement as to which word to use when representing a concept. Any term preference can be accommodated. It is important that the community does not have to reach consensus on terminology before moving forward with linking terms in manuscripts to the ontology. However, many hymenopterists, including one group at the 7th International Congress of Hymenopterists (Kőszeg, Hungary; June 2010), have called for the HAO Project to serve as an instrument for designating preferred labels for concepts. Simply put, preferred labels for concepts facilitate accurate communication, for future students of hymenoptera anatomy, or professionals attempting to learn from publications in disparate superfamilies. Additionally, a 'name' or designated preferred label must be included in the OBO formatted file. Within the HAO project, we have, as a result, established an email list and group for discussion with the Hymenoptera Community (via the International Society of Hymenopterists) to vote on preferred terms.

3.1.5. Ontology Representation

The archived representation of an ontology is an expressly formatted and defined text file (as pictured in part in figure 5, above). Typically, ontologies are represented in Web Ontology Language (OWL) or Open Biomedical Ontology (OBO) format. The OWL family of languages is a general-purpose means of representing knowledge on the Web, and the World Wide Web Consortium supports their development. OBO specifications are much more narrowly focused and technically simpler, their purpose being to serve the specific needs of biologists who use ontologies. The Hymenoptera Anatomy Ontology is constructed following OBO format guidelines (Day-Richter, 2012) and, with some exceptions, it is losslessly translated between OWL- and OBO format. OBO format includes (figure 9):

1. **Term**: Indicates a new concept is about to start in the OBO file, always separated by brackets (e.g. [Term]).
2. **id**: The unique identifier given to the concept.
3. **name**: The preferred label for the concept.
4. **def**: The definition of the concept, or the paramaters of the concept.
5. **synonym**: Other labels used to represent the same concept.
6. **is_a**: The is_a relationship is the only required relationship. It expresses the higher order in which the concept exists.

7. **relationship**: The other relationships applied to this concept. In the HAO these could be: is_obsolete, part_of, integral_part_of, and attached_to.

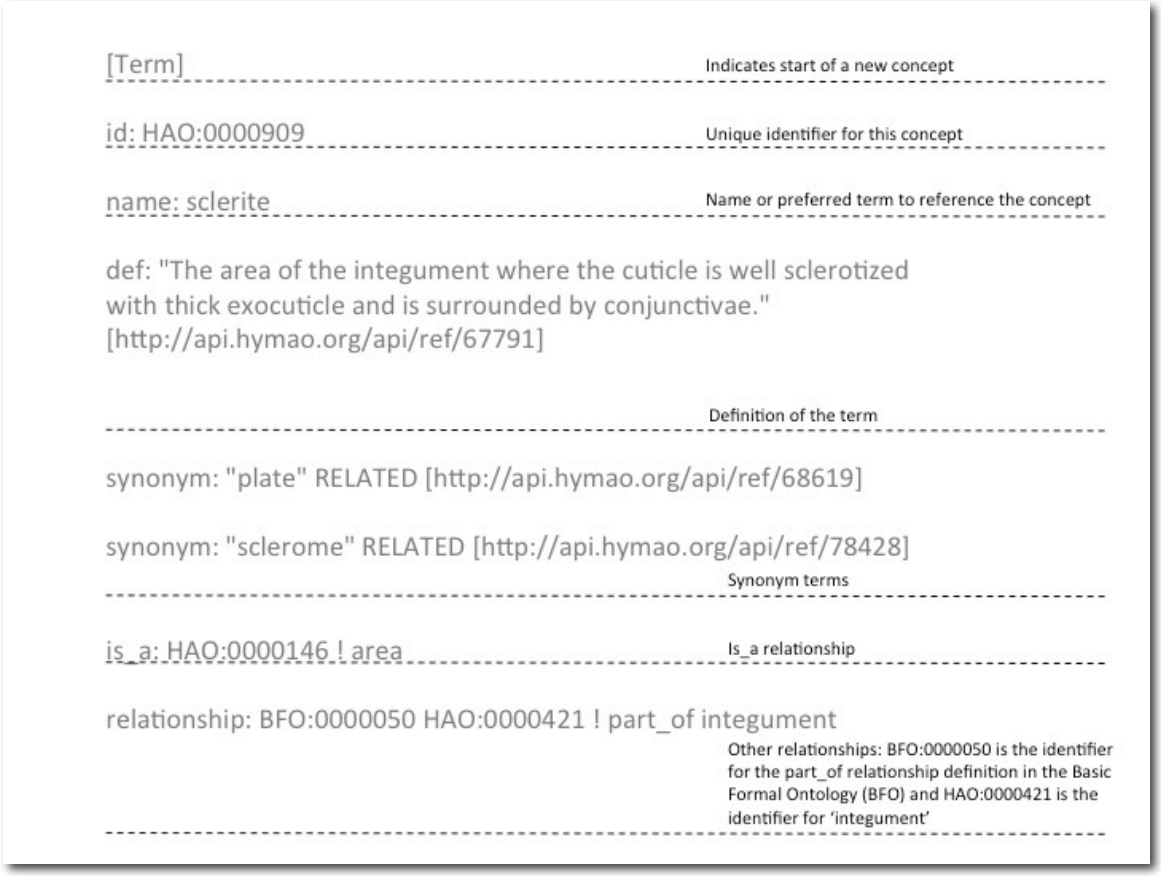

Figure 9: OBO format file definitions.

3.1.6. Homology and the HAO

Every ontology is underpinned by a design by its very nature. The specific philosophies that the author and her colleagues embraced during development of the HAO were critical in guiding the decision-making process governing both the selection of concepts to include in the HAO, and how those concepts are defined (see chapter *3.1.2. Genus Differentia*, above). Further, the criteria that were used are important to be aware, as they established the fundamental principles of the ontology, which must be understood and adhered to in order for it to be useful as a tool for referencing and aligning anatomical concepts by biologists. The difference between homology concepts as enabled in the HAO versus the biological concept of homology is discussed in Seltmann et al. (2012). Primarily, the HAO project rests on recognizing different instances of a topographically defined concept as 'the same' (e.g., the fore wing of taxon A is the same structure as the fore wing of taxon B), we refer to as 'structural

equivalents' or analogous structures. Homology in this context has been discussed extensively in the literature although referred to by a variety of names such as, topographical homology (Jardine, 1969; Rieppel, 1980), topographical correspondence (De Pinna, 1991; Rieppel, 1988), topographical identity (Brower & Schawaroch, 1996), and homology (Owens, 1843; Remane, 1952). In biology, however, homology is more explicit, referring to a more profound 'sameness', because it expresses a theory about structures sharing a common evolutionary ancestor (Brower & Schawaroch, 1996; de Pinna, 1991; Haszprunar, 1992; Lankester, 1870; Nixon & Carpenter, 2012; Patterson, 1982; Scotland, 2000; Wagner, 1989; Wiley & Lieberman, 2011). Homology in this evolutionary context is often dynamic, and may be controversial or involve conflicting and changing hypotheses. The dynamic nature of biological homology hypotheses conflicts with the goal of unambiguous definition of anatomical concepts, and, as such, overt reference to biological homology hypotheses are avoided in constructing definitions.

The HAO provides structure-based anatomical concepts, from which homology hypotheses can be developed and subsequently tested using phylogenetic methods. New concepts can be derived as necessary, but these concepts will be rooted in what can be observed, using positional placement of the anatomical feature in relation to other anatomical features, including attachment of muscles. Many HAO concepts are sufficient for the basis of evolutionary homology hypotheses; however this is not always the case. One example of how a structurally equivalent concept might not be biologically homologous pertains to processes (http://purl.obolibrary.org/obo/HAO_0000822). Many distantly related hymenopterans have similarly located (i.e. structurally equivalent) pronotal spines. These spines would correspond to the same concept in the HAO, but are not necessarily biologically homologous.

Defining concepts based on structural equivalency allows clustering of anatomical structures described in the literature. Because of the structural equivalency criterion, the match between ontology concept and author concept of a term does not need to be exact but rather the term in the paper must minimally meet the definition of the concept in the HAO. If an author intends to state that a particular HAO-defined anatomical concept was observed in some instance then that author must ensure that all of the components of the anatomical concept provided in the HAO were observed. For example, in many HAO concepts muscle attachments are used, and they are an important line of evidence in many definitions. An author must have observed these muscles and their points of attachment to state that the ontological concept in question

was present in their hymenopteran instance (i.e. specimen or taxon). Careful attention must also be made to the relationships a particular anatomical concept has to other concepts; the author must, ideally, also observe these relationships.

HAO concepts may still be referenced during the formulation or discussion of hypotheses, particularly character evolution is concerned, even without the prior observation of all of the components that define the anatomical concept. These loose references are useful for discussion when some structures, such as muscle attachments, are unavailable for observation. For example, a paleontologist studying two fossils may wish to discuss them in the context of anatomical concepts in the HAO, noting similarities the structures have to HAO concepts based on what is observable in the fossil. By doing so, the study provides an anchor for discussion of the evolution of an anatomical concept, something potentially useful to future researcher. However, in extant organisms, it is arguable that research should always include internal, not only external features (Deans, Mikó, Wipfler, & Friedrich, 2012), since assumptions concerning external structures in the absence of understanding their internal component has led to misunderstandings. For example, sclerites were often previously defined based entirely on 'sutures', or invaginations in the integument. Those invaginations, or apodemes, are typically attachment points for muscles and not really sutures at all; as a suture implies the joining together of two plates (Snodgrass, 1956).

When an author's concept of a term, based on observations pertaining to its structure, differs from that of the definition in the HAO, a new concept should be created for inclusion in the latter. It is best practice to add a new term to the HAO in this case rather than to try and force some association with a present HAO concept. However, new terms are not added to the HAO because of the part of the concept definition was not, or could not be observed.

3.1.7. Incorporation of Other Relevant Ontologies

It is beneficial to participate in the ongoing efforts of other relevant ontology initiatives during the creation of a novel ontology. All of the ontologies listed below are found on the OBO foundry website (http://www.obofoundry.org/) and were referenced or augmented while crating the HAO. These are the:

1. **Phenotypic Quality Ontology** (PATO - http://www.obofoundry.org/cgi-bin/detail.cgi?id=quality)
2. **Basic Formal Ontology** (BFO - http://www.obofoundry.org/cgi-bin/detail.cgi?id=bfo)

3. **Common Anatomy Reference Ontology** (CARO - http://www.obofoundry.org/cgi-bin/detail.cgi?id=caro)
4. **OBO Relationship Types** (OBO_REL - http://www.obofoundry.org/cgi-bin/detail.cgi?id=relationship)
5. **Relation Ontology** (RO - http://www.obofoundry.org/cgi-bin/detail.cgi?id=ro),
6. **Spatial Ontology** (BSPO - http://www.obofoundry.org/cgi-bin/detail.cgi?id=spatial)
7. **UBER Anatomy Ontology** (UBERON - http://www.obofoundry.org/cgi-bin/detail.cgi?id=uberon)

PATO is intended entirely for quantitative and qualitative terms (i.e. 'shiny', 'brown', 'posterior') that are useful in annotation of genes or descriptive characters regardless of domain. Originally PATO was created to interact with the Cellular Component (GO) ontology as a means of annotating gene sequences. Later its value for annotating descriptive statements from the literature was realized, requiring PATO to significantly expand. During the process of literature analysis we provided terms for expansion of PATO that are commonly used in Hymenoptera descriptive texts, although many are still needed. PATO is not directly referenced through identifiers in the HAO, however it is apparent that augmenting PATO will be useful for future work of incorporating ontology into Hymenoptera descriptive statements. In addition, incorporation of new terms into any ontology requires significant curation. The majority of the ontologies to have a term incorporated one must first add the term, reference, and proposed definition to a list (i.e. SourceForge tracker for PATO - http://sourceforge.net/tracker/?group_id=76834&atid=595654) where it is evaluated for possible inclusion in the ontology. This process can be time consuming and difficult if the terms are needed for immediate consumption in a project or experiment, thus it is beneficial to identify them well ahead of time. One method for handling disparate needs from multiple users is creation of a 'Slim'. Slims are subsets of terms that users create to fit their specific needs. An arthropod specific Slim for PATO is being evaluated, as many of the terms necessary for arthropod descriptions (i.e. surface sculpture) are not typically used in gene annotation, particularly not for vertebrates. By creating a Slim, arthropod domain terminology would not be over-crowding the most commonly used terms for annotating zebra fish. Also, Slims are equally important because for very inclusive ontologies, like PATO, the incorporation of terms is one alternative to later having to align disparate qualitative ontologies. Ontologies are not as beneficial alone as

they are aligned with other ontologies, so evaluations occur across groups of organisms may occur.

The Spatial Ontology (BSPO) includes positional references, commonly included in anatomical definitions (i.e. anterior, posterior). BSPO is not directly referenced (through the use of BSPO identifiers) in the HAO. Common Anatomy Reference Ontology (CARO) is considered a foundational ontology, or the basic framework on which other anatomy ontologies can be built. By referencing the same foundational ontology, alignment for anatomy ontology is made fundamentally easier (Bertone et al., 2012; Mungall et al., 2012). CARO terms included in HAO are: anatomical entity [CARO:0000000], anatomical structure [CARO:0000003], portion of organism substance [CARO:0000004], anatomical space [CARO:0000005], material anatomical entity [CARO:0000006], immaterial anatomical entity [CARO:0000007], anatomical line [CARO:0000008], anatomical point [CARO:0000009], anatomical surface [CARO:0000010], anatomical system [CARO:0000011], multi-celled organism [CARO:0000012], cell [CARO:0000013], cell component [CARO:0000014], compound organ component [CARO:0000019], simple organ [CARO:0000021], compound organ [CARO:0000024], male organism [CARO:0000027], female organism [CARO:0000028], hermaphroditic organism [CARO:0000029], asexual organism [CARO:0000030], organism subdivision [CARO:0000032], acellular anatomical structure [CARO:0000040], anatomical cluster [CARO:0000041], extraembryonic structure [CARO:0000042], portion of tissue [CARO:0000043], sequential hermaphroditic organism [CARO:0000045], gonochoristic organism [CARO:0000048], protandrous hermaphroditic organism [CARO:0000049], protogynous hermaphroditic organism [CARO:0000050], sequential hermaphroditic organism [CARO:0000046], anatomical group [CARO:0000054], multi-tissue structure [CARO:0000055], cell space [CARO:0000062], portion of cell substance [CARO:0000063], basial lamina [CARO:0000065]. For examination of the utility of these included terms see chapter *4.1. Contributions to the Hymenoptera Anatomy Ontology (HAO).*

OBO Relationship Types Ontology (OBO_REL), now the Relation Ontology (RO), standardizes the relationships used in the HAO and we reference the ids for those relations in the OBO_REL. (ex. 'OBO_REL:part_of'). Although OBO Relationship Types Ontology is considered to be legacy, and is superseded by the Relation Ontology, it will remain in perpetuity, maintaining persistent identifiers for its concepts. At the end of the HAO file itself are those relations defined, including relations not yet in RO, such as HAO:attached_to (figure 10).

```
[Typedef]
id: HAO:attached_to
name: attached_to

[Typedef] id: OBO_REL:0000004
name: integral_part_of
is_transitive: true
is_reflexive: true
is_anti_symmetric: true

[Typedef] id: OBO_REL:part_of
name: part_of
is_transitive: true
is_reflexive: true
is_anti_symmetric: true
```

Figure 10: Typedef section of the Hymenoptera Anatomy Ontology OBO file.

The Basic Formal Ontology (BFO) is a very high level ontology with only around 34 included classes. In the HAO BFO concepts are referenced for part_of relationships (figure 5, above), although identifiers for BFO are found in the Relation Ontology (RO) and not in the BFO flat file itself. Again, BFO is a domain neutral ontology whose purpose is to provide a framework for other ontologies, thereby increasing interoperation and support consistent annotation across domains (http://www.ifomis.org/bfo). For construction of the HAO we only needed to be certain to remain consistent with the Relations Ontology, and to realize the separation of Entity and Quality concepts. For example, this distinction can become subtle in the case of surface sculpture. If a sclerite is 'rugulose' for example, it is a wrinkled area, and is a Quality (as defined by BFO "A specifically dependent continuant that is exhibited if it inheres in an entity or entities at all (a categorical property)." However, one component of a rugulose surface looked at individually may be defined, as "The apodeme that is elongate" [http://purl.obolibrary.org/obo/HAO_0000899], and labeled 'ridge', which is an Entity, because it exists as itself and is not dependant (as in the definition for Quality). One then may ask; when do repeating Entities become a Quality? This subtle distinction remains open for debate, however. Since anatomy ontologies are intended to describe 'real things' for practical purposes (Merrill, 2006), the most important factor becomes how the term is used in author's text, not necessarily the philosophical distinction.

3.1.8. Software

The primary ontology development software for the HAO is mx, 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 dissertation has coded public portals for mx, and extended mx for the term cluster analysis aspect of this dissertation. Presently, there are multiple instances of mx on two dedicated servers (North Carolina State University & Texas A&M) to manage all aspects of descriptive taxonomy including: specimens, collecting events, extracts, sequencing progress, primer design, images, descriptions, diagnostic keys, literature, matrices, phylogenetic characters, ontology, published Web portals for data, and phylogenetic trees. The interconnectivity of a database containing research driven taxonomic data (name catalogs, matrices, etc), with anatomy ontology development software, guided our commitment to integrating anatomy ontology as part of the normal taxonomic revisionary process, and presented us with a good workbench for demonstrating its benefits for Hymenoptera taxonomy. Examples of potential utility of an inclusive system for dissemination of taxonomic information were published including interactive keys linked to HAO concepts (Sharkey et al., 2009).

The mx on Rails framework allows for rapid application development and a productive environment for experimentation. The software is version controlled using the content management repository SourceForge (SourceForge, 2012), thereby maintaining a transparent methodology, versions of the code base, and 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) regarding the Analyzer tool, and (Seltmann, Pénzes, Yoder, Bertone, & Deans, in press) for the Proofer tool.

3.2. Utilizing the HAO for Literature Analysis

One major aim of this dissertation is the exploration of descriptive terms accumulated as a product of the construction of the Hymenoptera Anatomy Ontology. Fundamentally, the development of the HAO was experimental in its use of Web based mx software and our methodology for extracting terms from legacy literature. In its early stages, the bulk of the terms and concepts were gathered for the HAO via expert examination of the known primary literature, including important morphological publications and online glossaries. This method of term accumulation was not expected to completely 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 the discovery of obscure terms, the author implemented an active learning, dictionary-based, natural language recognition tool for analyzing the text of relevant publications. This tool, referred to as the 'Proofer', is part of an iterative approach to developing phenotype-relevant ontologies, and enables discovery of obscure descriptive terminology.

In order to validate the utility of this approach, the author undertook an experiment based on sampling taxonomic descriptions from the online Journal of Hymenoptera Research (JHR) for terminology not yet included in the HAO, using the Proofer software tool. The sampled articles were then analyzed for the occurrence of relevant terms using a variety of data clustering methods, and were subsequently compared to our present understanding of Hymenoptera lineages. The general course of Proofer development and term analysis is presented in (Seltmann et al., in press), and is described in further detail below.

3.2.1. Term Collection

One of the most important and growing online resources in the biological sciences is the Biodiversity Heritage Library (BHL)(Biodiversity Heritage Library, 2011), a clearinghouse for legacy literature, all of which is optically scanned for character recognition (OCR). The International Society of Hymenopterists (ISH) (International Society of Hymenopterists, 2012) archives the Journal of Hymenoptera Research (JHR), which it publishes, in the BHL. We extracted OCR text for JHR (years: 1993-2007, the latest year available at the time of data collection) from the BHL and manually partitioned the 353 articles for upload into the mx database (Mx, 2012). When this exercise was conducted, the BHL Application Programming Interface (API) did not return OCR for specific articles, only for entire issues of the journal. Processing of the BHL OCR thus required manual cutting and pasting of the text into the database. We made no attempt to correct the OCR output. Relevant metadata, including reference citation information, was associated with each article. Article bibliographic information was collected from Google Scholar (Scholar, 2010), aggregated in Zotero (Zotero, 2010), and downloaded in Endnote (Endnote, 2010) format. The collected Endnote citations were subsequently uploaded into the mx database using a custom Endnote importation tool that we developed.

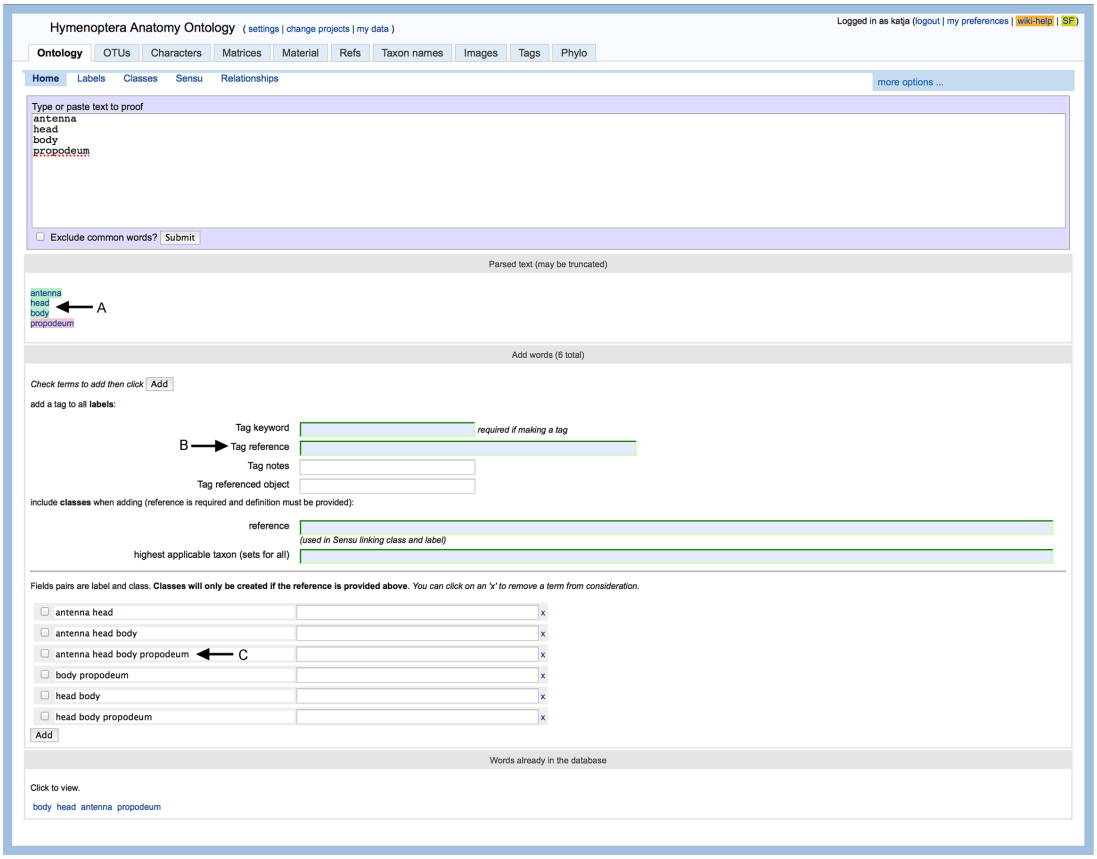

Figure 11: Screenshot of the mx 'Proofer' interface for string matching terms in the database with OCR text.

Once the articles were in the database, a simple dictionary-based, entity recognition tool was developed in mx to match terms captured for the HAO within blocks of text. The tool, or 'Proofer', uses string matching, allowing for commonly found exceptions and special cases, thus reducing the impact of malformed OCR, as was commonly found in the BHL-delivered JHR text. The Proofer displays a list of matches to terms in the ontology for the expert user. These terms are highlighted and linked to the display page for that term (figure 11: A). It also presents a proposed list of terms that could be added to the database if the user chooses. In order to create this list, sentences are first broken down into phrases by splitting sentences at small words (1-3 characters long), removing those small words, and splitting at punctuation (period, comma, semi-colon, etc). These phrases are then displayed to the user in a list format starting with a single unmatched word, or term not already in the database, and 1-5 flanking words expanded from left to right (figure 11: C). Users then browse the list of proposed unmatched terms and select those that should be added to the database; as a result, user (human) input is necessary during the final addition of terms to the database. Adding

flanking words reveals more complex anatomical labels such as ‘propleural arm muscle’ where ‘propleural arm’ may already be a label in the database but ‘propleural arm muscle’ may not. All terms added in this manner were annotated (‘tagged’) as JHR-BHL entered objects to facilitate future analyses of the terms collected during this exercise (tag field illustrated, figure 11: B). Also, in order to reduce the number of potential terms presented to the reviewer, active learning (Day, Aberdeen, Hirschman, Kozierok, Robinson, & Vilain, 1997) was employed based on a feedback mechanism between application and user.

Words that are presented to the user for possible inclusion into the database and that are not selected by the user are added to a ‘stop words’ table. If a word is rejected by the user 10 times (i.e. from ten separate articles) that word is added to the final stop words list and no longer presented in subsequent articles, thus reducing the total number of words presented to the user for evaluation.

3.2.2. Analysis of Collected Terms

For each of the 353 articles small amounts of metadata (as ‘tags’) were captured in the database to facilitate creating lists of terms specific for analysis. First, the articles were reviewed and placed into one of two categories: 'description of new taxon' or 'non-description'. Articles were deemed descriptive based on the use of the words 'description of' in the article title or if taxonomic treatments were contained within the body of the article. Additionally for each article, the name of the taxon being described was captured in the database at the family level. Finally, terms representing morphological (i.e. anatomical) concepts and those representing qualitative concepts were differentiated.

The resulting data were then used to produce text files useful in R (R Development Core Team, 2010)(version 2.11.1), creating an occurrence (presence/absence) matrix using anatomical terms as characters and articles as terminals, with each article tied to a taxon as described within the article. Terms designated as characters were limited to morphological terms and totaled 816. Qualitative terminology (ie. 'shiny', 'brown', 'rugulose') were not included in the dataset. The terms 'cell', 'area' and 'costa' were removed from the character list as these terms are commonly used in other disciplines besides descriptive biology and often had non-morphological meaning in descriptions. 179 articles were used as terminals, representing 35 families and 10 superfamilies.

Synonyms and plural terms were summed in the analysis and terms were

analyzed as they were recorded in the database. The characters were scored in a binary matrix according to the presence (1) or absence (0) of the term occurrence within the text of a given article. Four permutations of the matrix were created based on the occurrence of a term. Analyses were performed that included terms that occurred 2, 10, 50 and 100 times in at least one article. Choosing articles for analysis based solely on occurrence of a term in all articles did not retrieve discrete article sets at higher frequencies, as common terms (figure 19 in chapter *4.3. Results for Utilizing the HAO for Literature Annotation*) are ubiquitous. Restricting the terms included in analysis limited the number of included terms to: 796, 500, 123 and 40 respectively (figure 12).

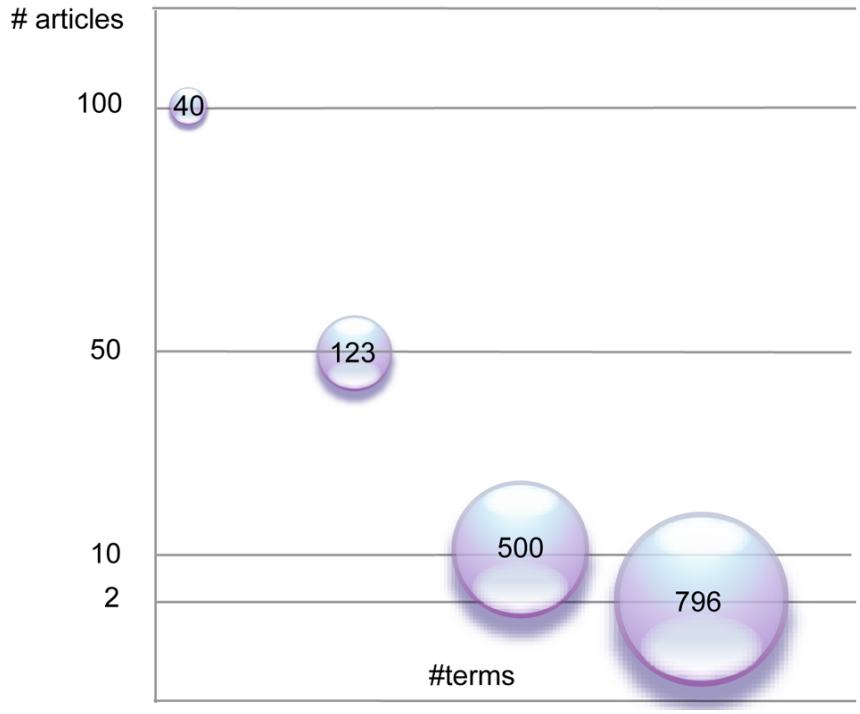

Figure 12: The number of characters (terms) present in at least 2, 10, 50, and 100 articles.

In order to assess the quality of the clusters that were produced among the articles by the respective algorithms and procedure, it is useful to have a ground truth model of the 'best' result, indicating families should normally cluster together. However, as outlined in chapter *1.1.1. Hymenoptera Phylogeny*, many of these relationships are, in fact, tenuous. Table 1 represents the author's synthesized understanding as to the membership of families to superfamilies, and superfamiles to suborder based on the present day literature. The inclusion of genera to families at the time the article was submitted to JHR was not called into question, and they were included based on the assumptions made in the article at the time the article was published. The table also

summarizes the families described in the BHL articles from those years that were included in analysis.

Family	Superfamily	Suborder
Ampulicidae	Apoidea	Apocrita
Andrenidae	Apoidea	Apocrita
Apidae	Apoidea	Apocrita
Colletidae	Apoidea	Apocrita
Crabronidae	Apoidea	Apocrita
Halictidae	Apoidea	Apocrita
Sphecidae	Apoidea	Apocrita
Aphelinidae	Chalcidoidea	Apocrita
Chalcididae	Chalcidoidea	Apocrita
Encyrtidae	Chalcidoidea	Apocrita
Eulophidae	Chalcidoidea	Apocrita
Eupelmidae	Chalcidoidea	Apocrita
Leucospidae	Chalcidoidea	Apocrita
Mymaridae	Chalcidoidea	Apocrita
Ormyridae	Chalcidoidea	Apocrita
Pteromalidae	Chalcidoidea	Apocrita
Torymidae	Chalcidoidea	Apocrita
Trichogrammatidae	Chalcidoidea	Apocrita
Bethylidae	Chrysidoidea	Apocrita
Scolebythidae	Chrysidoidea	Apocrita
Charipidae	Cynipoidea	Apocrita
Cynipidae	Cynipoidea	Apocrita
Ibaliidae	Cynipoidea	Apocrita
Braconidae	Ichneumonoidea	Apocrita
Ichneumonidae	Ichneumonoidea	Apocrita
Mymarommatidae	Mymarommatoidea	Apocrita
Stephanidae	Stephanoidea	Apocrita
Formicidae	Vespoidea	Apocrita
Mutillidae	Vespoidea	Apocrita
Pompilidae	Vespoidea	Apocrita
Scoliidae	Vespoidea	Apocrita
Tiphiidae	Vespoidea	Apocrita
Vespidae	Vespoidea	Apocrita
Orussidae	Orussoidea	Symphyta
Argidae	Tenthredinoidea	Symphyta
Pergidae	Tenthredinoidea	Symphyta
Tenthredinidae	Tenthredinoidea	Symphyta

Table 1: Expected groupings for Hymenoptera family and subfamilies.

In order to assess the occurrence of terms contained within articles, matrices were investigated using agglomerative hierarchical clustering methods performed in the software R (R Development Core Team, 2010)(version 2.11.1) using packages 'stats' (R

Development Core Team, 2010), 'simba' (Jurasinski & Retzer, 2012), 'vegan' (Oksanen et al., 2010), and 'ape' (Paradis, Claude, & Strimmer, 2004). The range of recovered groups (clusters) on the trees varied from 59-160, depending on which analysis method was used (see figure 13, below). Groups were revealed by trimming trees after analysis, and were evaluated based on two criteria. First, family and superfamily membership was assigned to each terminal, based on the taxa described in the article. A family or superfamily is a group of organisms based on shared characteristics, associated together under the auspices of the classification hierarchical system, under which other groupings (tribe, subfamily, genus, species) are clustered. Superfamilies are groups that contain multiple families.

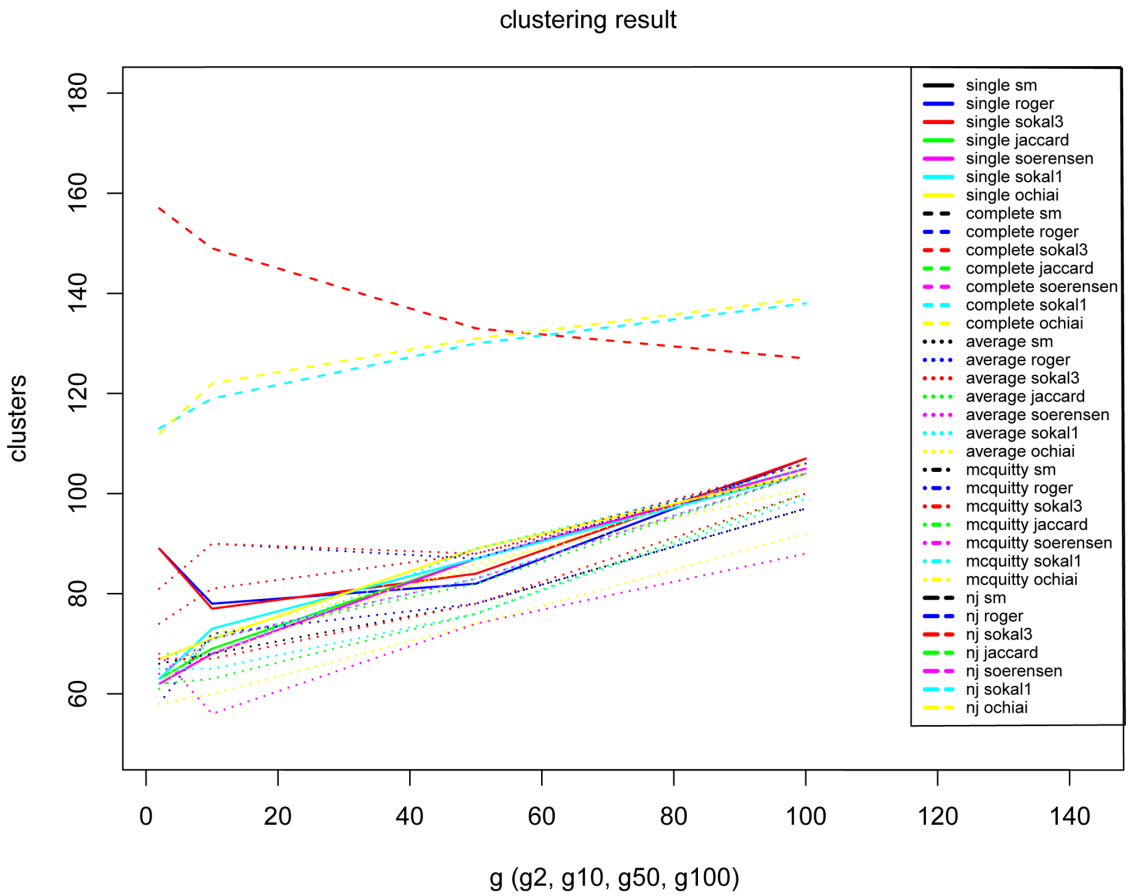

Figure 13: Variation of number of returned clusters based on clustering method and term occurrence in articles.

These families and superfamilies are generally listed in the analyzed journal article; if not, the taxon was placed according to our present understanding of Hymenoptera relationships. Once terminals (taxa) were assigned to a family/superfamily the trees were pruned according to these groups. For example, if two terminals belonged to the

same family, and reached the next internal node, they were considered to belong to the same group. Ideally, 10 groups of superfamilies and 37 of families were expected, as this is the number of Hymenoptera families/superfamilies published in the JHR articles used in the analysis.

The grouping of terminals on the basis of binary characters (Sokal & Michener, 1958) was extensively investigated via agglomerative hierarchical clustering using the linkage methods (single, complete, average (UPGMA) and McQuitty (WPGMA)) and neighbor-joining (NJ) (Saitou & Nei, 1987). Figure 13 outlines the range of outcomes based on which distance - clustering method was chosen. The former methods result in ultrametric trees (dendrograms, which are 'rooted'), while the result of the NJ method approaches an additive tree (unrooted) that is based on optimization of the distance on the whole tree (Saitou & Nei, 1987). Seven different metric distances were selected, 3 of which were symmetric (incorporate absence matching) and 4 of which were asymmetric (absence matching is ignored), as follows Kaufman and Rousseeuw (1990) and Legendre and Legendre (1998). The Sorensen-Average results tree is included (figure 18) to visually illustrate the grouping results because it most accurately follows groupings expected for Hymenoptera. All other trees and analysis files are available in the supplementary documents for the Seltmann et al. (in press) paper.

3.3. Utilizing the HAO for Literature Annotation

Term discovery through text mining is one pathway through which ontology may aid document analysis and taxonomic description association. However, term discovery a posteriori still yields results that are vague with respect to the intended definition or associated ontological concept for a given term in the text. Aim 3 of this dissertation is to define a methodology for the association of terms and concepts, and to promote it as an integral component of the process of authoring a manuscript, based on a digital ontology-based link that is established prior to publication.

To facilitate this, the author and her colleagues developed a relatively simple and easily comprehensible methodology for linking terminology in descriptive texts with the Hymenoptera Anatomy Ontology through Uniform Resource Identifier (URI) tables, which could be included in a manuscript. The tables are created using a software tool that we named ‘Analyzer’, based on the mx software platform. The Analyzer software is based on linking specific words in the manuscript with uniquely defined

concepts in the HAO. Once published, the links so established refer to digital records at the Hymenoptera Portal Webpage (http://portal.hymao.org/, figure 14), where definitions and illustrations of those concepts may be retrieved. The general discussion of URI for the Hymenoptera community is derived from Seltmann et al. (2012), and is discussed further in the context of this dissertation below.

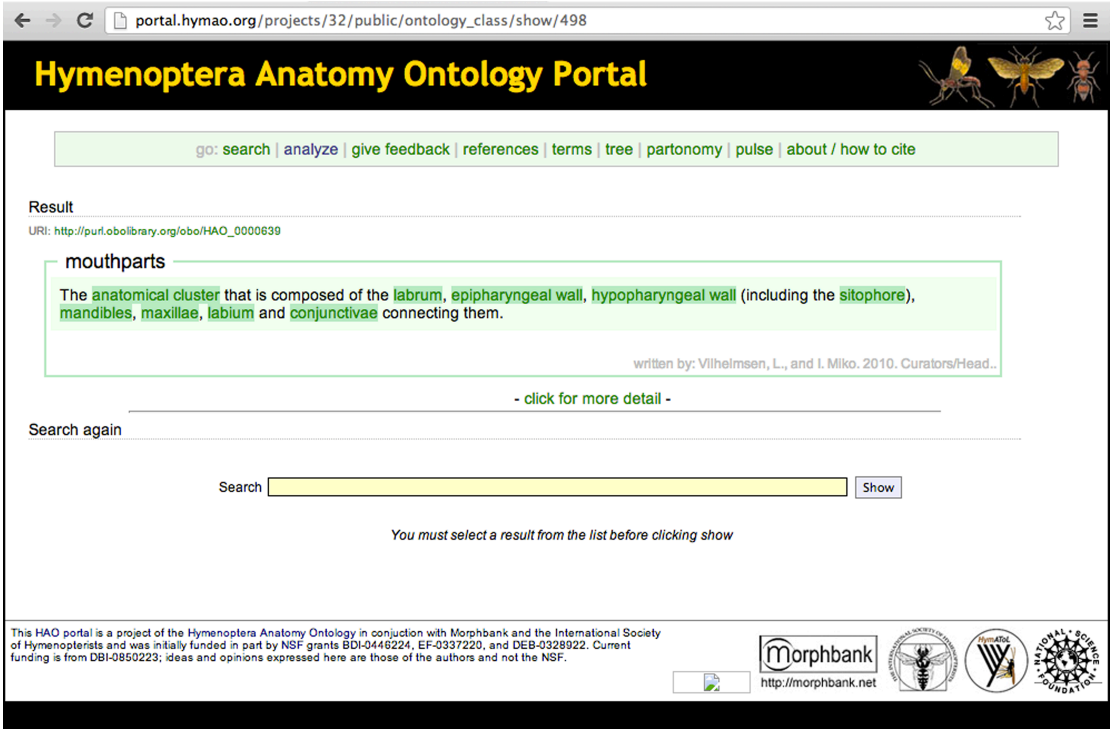

Figure 14: Example Hymenoptera Anatomy Ontology Portal WebPage resolving to the result page for concept HAO:0000639.

3.3.1. Unique Identifiers for Anatomical Concepts

Conceptually, creating URIs is analogous to submitting sequences to GenBank (http://www.ncbi.nlm.nih.gov/genbank/), the National Institute of Health genetic sequence database, and receiving unique identifiers (i.e., GenBank accession numbers) that serve as publishable and persistent references to the data. The HAO serves a similar function by providing unique, resolvable, identifiers for Hymenoptera anatomical concepts. The identifiers used with the HAO are Uniform Resource Identifiers (URIs), which consist of a Persistent Uniform Resource Locator (PURL: http://purl.obolibrary.org/obo/), plus an HAO identifier in OBO format (e.g. HAO:0000397). The latter is a combination of the namespace used in the OBO Foundry (e.g., 'HAO') and a unique seven-digit number (e.g. '0000397'). The ':' in the identifier is

replaced with a '_' in the URL form. Any given URI points to only a single concept within the HAO (e.g., http://purl.obolibrary.org/obo/HAO_0000397 = "The tagma that is located anterior to the thorax" = head). The seven-digit number is unique for that concept in the HAO, and the namespace is unique for ontologies registered to the OBO Foundry.

A persistent uniform resource locator (PURL) is a Uniform Resource Locator (URL) that is used to redirect to the location of a Web resource. PURLs are used to curate URL resolution; redirecting HTTP using HTTP status codes. Using a PURL allows for the actual Web address we wish the URI to resolve to change over time. Additionally, PURLs are important because they allow the HAO to be utilized in multiple contexts. In other words, the PURLs, in combination with Web-server configurations, allow for different responses based on who (or what) is making the request. For example, a request from a Web browser—a person clicking on a link in a journal article or Website—would return content that a human can interpret (e.g., a Webpage), whereas requests from computational sources would receive responses that the application can understand (leaving out the human-readable components).

Computationally, the redirection of URIs is, as discussed above, relatively simple. However, a greater challenge is involved in facilitating the production, use, and publication of URIs in morphological and descriptive publications. Materials and methods sections within papers that reference hymenopteran anatomy frequently point to concepts published elsewhere (e.g., "Morphological terminology largely follows Mikó et al. (2010)" (Talamas, Masner, & Johnson, 2011)). New or revised anatomical concepts are typically presented in a paragraph format, with terms highlighted and a definition or discussion following.

In Seltmann et al. (2012), we introduced a novel, table-based format that facilitates making reference to existing anatomical concepts in the HAO in a standardized form in published literature. An extensive example for how this is handled is given in chapter *4.5* of this dissertation *(Results from Morphological Exploration for HAO Augmentation)*. Although authors may include a variety of information in the URI table, or choose to link terms directly in the text, we propose a few column names below. The only necessary columns for a completely referenced table are Term and URI.

Possible URI Table Columns:

1. **Term**: The literal text (string of letters) used in the paper for the anatomical concept. It is important to reference anatomical terms used within the table

exactly throughout a paper, to eliminate ambiguity. Simplifying terms within a document, without including these simplifications in the table, reduces the effectiveness of a table intended to clarify terminology. For example, if "abdominal tergum 3" is in the table, the structure should be fully referenced throughout the document as "abdominal tergum 3" and not simplified to "the tergum" anywhere. Additionally, terms new to the publication should be sent to the HAO so they can be included as synonyms once the manuscript is published.

2. **URI**: The unique, resolvable, identifier for the anatomical concept. The URI includes the entire URL for the term, not only the identifier. These are then clickable in the table, resolving back to the Hymenoptera Portal WebPages.
3. **Definition**: A verbatim replication of the definition in the HAO, which are written as genus-differentia (see Smith, 2005).
4. **Definition Citation**: A citation from which the concept/definition was derived; within the HAO
5. **HAO Term**: The present preferred term for the concept denoted by the URI in the HAO. This term is provided only when it does not match the term the author uses in his or her publication. This option provides authors the freedom to reference anatomy in any manner he or she sees fit, and provides the term to search for the concept in the HAO. Preferred terms may change in the HAO, but synonyms are equally easy to search by.
6. **Comment**: Comments may pertain to any or all of the columns for a given concept. For example they can be used to clarify subtleties or provide taxon-specific discussion.
7. **Synonyms**: Synonyms for the label, this may include the term used to represent the concept in the text if it is not the HAO preferred term. This column is redundant with the inclusion of Term and HAO Term, however it is less specific and allows for the authors to include all known synonyms for a given Term.
8. **Abbreviation**: Abbreviation for the term in text and figures. These may or may not be included in the HAO.

Abbreviation	**HAO Term**	**URI**	**Term**
axc	axillular carina	http://purl.obolibrary.org/obo/HAO_0000161	
epc	epomial	http://purl.obolibrary.org/obo/HAO_0000307	vertical

	carina		epomial carina
	femoral groove	http://purl.obolibrary.org/obo/HAO_0000326	femoral depresssion
	gena	http://purl.obolibrary.org/obo/HAO_0000371	

Table 2: Example URI table from Talamas et al. (2011). HAO Term represents the label in the HAO associated with the concept. Term represents the authors' term for the concept as it is represented in the publication.

3.3.2. Creating a URI Table

URI tables are created prior to the submission of a manuscript, and are included at the end of the document, in the appendix. It is preferred that they are included in the text, and not in attached supplementary material, however, the latter is often dependent on the policies of the journal. The goal of building a URI table is to find an appropriate URI that represents the anatomical structures referenced in a paper. The highest priority candidates for a URI table are those structures that have historically been poorly defined, new concepts, or those that need clarification (e.g. they have synonymous and/or homonymous terms). Examples of URI tables are given above, in (table 2), and are used in the chapter *4.5. Results from Morphological Exploration for HAO Augmentation* (table 4) appearing later in this dissertation.

3.3.3. Using the HAO Analyzer Tool

The primary goal of the HAO software Analyzer tool is to facilitate the discovery of URIs and associated concepts in the HAO, and, by doing so, to speed up the creation of URI tables for authors. In addition, the Analyzer outlines those terms and URIs in a table format, to encourage inclusion in publications (journal articles, books, or digital applications such as Websites). The mechanism created to accomplish this is a simple software user interface (figure 16) facilitating the submission of an author's anatomy-related text. The text that is submitted is then broken down into individual and groups of words, and these are compared against the terms in the HAO. Terms are matched letter-for-letter, and no 'fuzzy' matching attempts are made. In other words, the software does not try to predict what anatomical concepts the author is referring to, and

matches entirely on the term, without textural context. However, synonyms are matched if they are included in the ontology, thus 'forewing' or 'fore wing' would both produce a return to concept (http://purl.obolibrary.org/obo/HAO_0000351). Terms, URI, Definitions, References, and Preferred Terms are returned in a table-like fashion. Additionally homonymous and synonymous terms are listed (and homonymous terms highlighted in pink) for further review. The software is able to recognize homonymous terms because they are associated with more than one concept in the HAO (see sensu discussion above, in chapter *3.1.1. Properties of Relationships*). The general workflow for creating a URI table begins during the manuscript process, where terminology is vetted against the HAO (figure 15). This part of the process is organic, and the HAO is utilized primarily as an illustrated glossary of terms and resource for further references regarding concepts. Also, the process allows the author to provide feedback for new terms to include in the HAO well before the manuscript is complete. Including new concepts in the HAO should not be a bottleneck to publication, and should be provided early. Secondly, the Analyzer tool is utilized to format the URI table for inclusion in the manuscript, or to examine entire paragraphs from the manuscript instead of single terms; again providing feedback to the HAO curator group if terms are not present in the HAO.

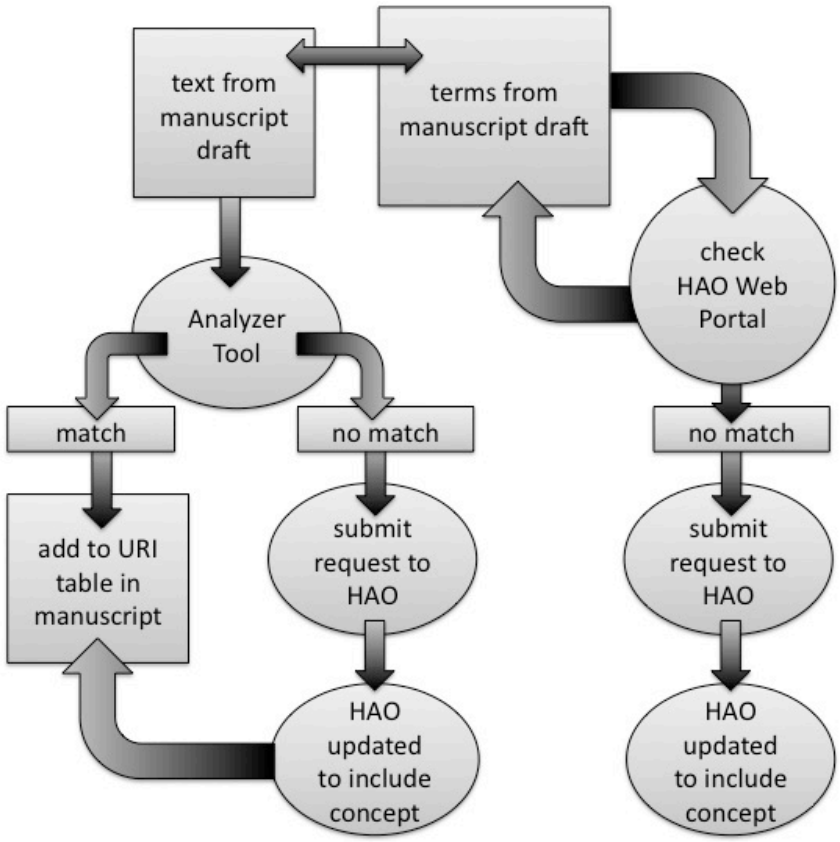

Figure 15: Analyzer workflow and diagram, modified from Seltmann et al. (2012).

To use the analyzer, one follows these steps:

1. **Paste Text**: Paste the diagnosis or description sections of a manuscript into the Analyzer window (figure 16: A, without extracting individual anatomical terms.
2. **Validate the Form**: Since the Analyzer is a public form we include a challenge-response test ('reCAPTCHA', figure 16: B), which must be filled out before submission, verifying that the operator is human is and not a software agent (Von Ahn, Maurer, McMillen, Abraham, & Blum, 2008)
3. **Submit**: Click Submit and wait. An Ajax enabled spinner appears to indicate wait time. The waiting time varies depending on the amount of text submitted. It is recommended that users search small chunks of text at a time, not an entire manuscript at once.

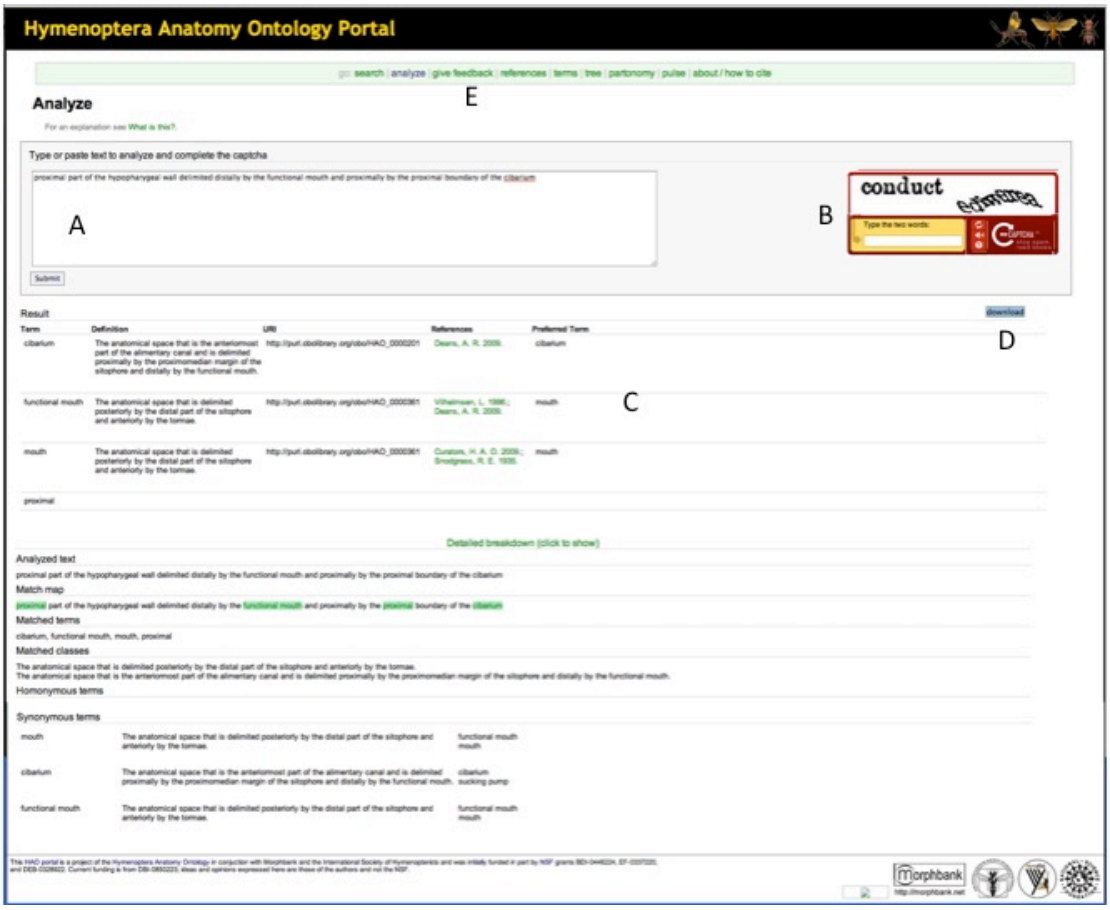

Figure 16: Analyzer tool at http://portal.hymao.org/.

4. **Examine Results**: Results are displayed in a table format suitable for inclusion in the manuscript (figure 16: C). If a term is not in the HAO, it will not appear in the results list. One uses the Match Map supplied in the Detailed Breakdown

Report to check if all of ones terms are included. If terms are missing that need to be included, one uses the feedback mechanism (figure 16: E) to have them added in the HAO. References represent citations wherein the term was used in conjunction with the definition/concept as interpreted by an HAO curator. Carefully reviewing each conceptual match is critical; users are not to assume that the result is using the term the user has provided in the manner intended.

5. **Download Results**: The user downloads the result by clicking the 'Download' button (figure 16: D). The downloaded file is a tab-delimited file that can be opened directly in a spreadsheet program (e.g., Microsoft Excel) or a text editor.
6. **Edit the Results**: If multiple concepts (definitions) are available for a given term, one chooses the one for ones table that matches ones concept.
7. **Check the Detailed Breakdown**: For additional details, one clicks on Detailed Breakdown. This report can be used to visually inspect for errors, mal-formatted terms, synonyms, and homonyms. The report includes:
    a. **Analyzed Text**: Confirms that ones text was submitted correctly. The analyzer may truncate ones text if it is too long, or if other problems are encountered.
    b. **Match Map**: The words that were matched and returned are highlighted in the context of the text here. Green highlights indicate a 1:1 mapping (i.e., there is only one concept present for the given term). Pink highlights indicate homonymy, and that there are multiple concepts for a given term.
    c. **Matched Terms**: A simple comma separated list of the terms that were found.
    d. **Matched Classes**: A list of the concepts that were matched.
    e. **Homonymous Terms**: Matched terms that are homonyms.
    f. **Synonymous Terms**: Matched terms that are synonyms.

3.4. Morphological Exploration for HAO Augmentation

Term extraction from the literature has the potential to be significantly more automated than is currently possible, through the use of natural language processing techniques such as are explored in aims 1&2 of this dissertation, *Utilizing the HAO for Literature Analysis.* However, aligning terms with concepts, or defining new concepts,

requires significant expertise in the domain it is describing. The domain illustrated in the HAO is Hymenoptera, requiring development of skills in Hymenoptera morphology, taxonomy, and bioinformatics in order to inform and participate in augmenting the ontology, as well as communicate its benefits to the Hymenoptera community. Here we provide a case study, examining the Mouthparts in Hymenoptera, and demonstrate the utility of the URI table as a mechanism for reporting constituent parts. The essential nature of domain expertise is highlighted in all of the HAO publications, but its importance is emphasized in Bertone et al. (2012) and Mikó et al. (2012).

3.4.1. Hymenotpera Mouthparts

Hymenoptera mouthparts are diverse throughout the order, accommodating many specialized lifestyles and feeding habits. The fundamental structures can widely vary, from shorter spongy glossa of evaniid wasps to the long-tongue siphoning mechanisms of the apiform bees. This portion of the dissertation is an exploration of variation in this anatomical cluster, aligning various structures using the structural equivalency criteria, and subsequently illustrating them through image and annotation, the internal and external structures of the mouthparts (extending to head, where applicable), in order to inform the HAO. Definitions included in the HAO focused on curators Lars Vilhelmsen, István Mikó and Gary Gibson. Term exploration for illustration primarily follows Beutel and Vilhelmsen (2007); Vilhelmsen (1996), (Snodgrass, 1910, 1942, 1956; Youssef, 1971) for comparison to highly derived *Apis meliffera*, and general structures in (Mason & Huber, 1993; Snodgrass, 1935). Additional references are included in the *Results: Mouthparts and Head Morphology* section of this dissertation, particularly in the URI table where, included with each definition, is the principle reference for that definition. Commonly, these include reference to the HAO curators who aligned concepts across publications or rewrote the definition in genus differentia.

Evaniidae (*Acanthinevania sp*. and *Evania appendigaster*) are the models of choice because they are readily available in large numbers, generally large in size (head up to 4-5mm across), and offer an aculeate variation on previously described work (listed above). Additionally, differences found by (Deans & Huben, 2003) in the mouthparts of Evaniidae demonstrate that this region has the potential to be a rich source of new characters, to be important to Evanioidea morphophylogeny, and to be extensible to hymenoptera systematics. *Acanthinevania* (Bradly, 1908) are found exclusively in Australia. Specimens for this study were collected in Victoria, Little Desert

National Park: Eastern Black McCabes Hut Track: 12.6 km SW Dimboola, on 6-22.11.2002 (-36.5275S, 141.916944E). *Evania appendigaster* (Linnaeus) are holarctic, common, and large wasps. The specimen dissected for illustration was collected at College Station, Texas on 6.9.2010 (30.6289N, -96.3311W).

3.4.2. Specimen Preparation and Imaging

Specimens from 75% EtoH were dissected in glycerin under a compound microscope. Dissections typically followed the protocol developed by István Mikó (personal communication, Mikó, Yoder, Seltmann, Bertone, & Deans, 2011). First the specimens are removed from glycerine and placed on a small amount of Blu-Tack©, an adhesive mounting putty, on a flat slide. The specimens then are sliced with a thin, new double-blade shaving razor, utilizing only one slice per edge of razor. The Blu-Tack maintains the position of the specimen during the process. Afterward, while looking under the microscope, the specimen internal components are examined. Fat bodies may or may not be removed using a minution or fine forcepts, depending on the intended purpose of the image. In general the goal is to capture the internal components in situ, with as little modification as is feasable. Ideally this will also increase the repeatablility of the results.

Concepts are illustrated using brightfield (compound microscope or Microptics©), and confocal laser imaging techniques. The utility of confocal techniques for study is demonstrated in (Deans, Mikó, Wipfler, & Friedrich, 2012; Michels & Gorb, 2012; Mikó et al., 2012). The Cellular and Molecular Imaging Facility at North Carolina State University, LSM 710 confocal workstation was utilized, typically with a 10x objective due to the large size of the subject. Additional initial start settings are found in Table 3. These initial settings were set as a start point, than nuanced for each image.

Scaling X	1.384 µm
Scaling Y	1.384 µm
Scaling Z	1.097 µm
Image size	x: 1024, y: 1024, z:215, channels: 3, 8 bit
Dimensions	x: 1415.60µm ,Y: 1415.60 ,z:234.76
Scan mode	stack
Zoom	0.6
Objective	C-Apochromat 10x/0.45 W M27
Pixel dwell	3.15 µs
Average	line 4

Master gain	Ch1: 395 Ch2: 469 ChD: 101
Digital gain	Ch1: 2.20 Ch2: 1.20 Ch3: 1.00
Digital offset	Ch1: 2.20 Ch2: 1.20 ChD: 0.00
Pinhole	48 µm
Filters	CH1: 499 - 577 CH2: 581 - 699
Beam splitters	MBS : MBS 488
Lasers	488 nm : 70.0%

Table 3: Default settings used for laser confocal images.

Specimens were sandwiched between 2 large slide cover slips, only adhered using glycerin (for larger specimens) or clear agar (for smaller cuts). No additional staining was needed for the images as the sclerotised components of the anatomy fluoresced at a different wavelength than the soft tissues, creating the necessary distinction between the structures. Additionally, it was always critical to make sure the Argon laser was utilized in imaging, as it greatly increased the level of clarity of images. Laser confocal image stacks were rendered in ZEN software or in the Zeiss LSM Image Browser software, version 4.2.0.121.

Brightfield images were created using image focus stacking software, either using the software Helicon Focus for images shot with Microptics system, or the software Combine Z for those created using compound microscopy. The majority of the later brightfield images were imaged in glycerin, which caused the specimens to move slightly during imaging. To minimize this effect fewer image stacks were taken, and the time between images was lengthened to allow the glycerin to settle before the next shot.

Subsequently, images were minimally manipulated in Adobe Photoshop© cropping, annotating and correcting for lighting issues and uploaded to the mx database for annotation. Scalable Vector Graphic (SVG) overlay annotations for the HAO were constructed in Inkscape©. Text annotations are added directly to the mx database as 'Tags'; see Yoder et al. (2010) for more detail about methods of annotation in the HAO.

4. RESULTS

4.1. Contributions to the Hymenoptera Anatomy Ontology (HAO).

The author's dissertation research contributing to, and based on, the development of the HAO has been extensively documented in a series of published articles that she has coauthored during the past two years (Seltmann et al., 2012; Mikó et al., 2012; Bertone et al., 2012; Seltmann et al., in press; Yoder et al., 2010), and through the illustrations and data visualized through the HAO Portal Website (http://portal.hymao.org/). Each of these publications has addressed a separate aspect of research associated with the HAO, as further summarized below and in the sections that follow.

The general design, developmental principles, and structure of the HAO were first presented by the author and her colleagues in Yoder et al. (2010). This dissertation further explained key benefits and issues associated with the use of anatomy ontology for morphology. These include many well-known terminology issues such as homonymy (i.e., the use of the same term for different structures) and synonymy (i.e., the use of different terms for the same structure). In the same article, we introduced a third issues related to defined terminology, 'concept drift', which refers to the application of a term to an increasingly diverse set of structures over time.

Subsequently, the author and her colleagues contributed the paper Seltmann et al. (2012), in order to provide a hymenopterist-centric introduction to the HAO. This article also focused on describing the new tools that were made available for hymenoptera research, and on clarifying some key concerns for researchers, such as how issues of homology are treated in the ontology. The topics contained within the Seltmann paper were motivated by prevailing points of confusion about ontology that were identified as key road blocks to be overcome in order for the research community to best understand the design principles of the HAO, and the most important advantages it holds over a simpler tool (e.g., a simple standardized glossary of terms). In the same paper, we attempted to address ontology-related jargon through our own glossary rather than avoiding the terminology altogether.

Among the other HAO related publications, the article Bertone et al. (2012) is especially notable, as it has demonstrated that the HAO, due to its highly granular nature, is ideal for alignment with other anatomy ontologies, expanding significantly its power and scope.

4.1.1. Further Developments and Implications for the Research Community

Broader outcomes from the HAO project are numerous. Since 2008, the HAO has since grown into the largest, best illustrated, and most documented multi-species arthropod anatomy ontology in existence. At present (September, 2012) this includes 2055 anatomical concepts and 3622 labels for these concepts, 2880 images, 8598 test annotations, and 269 references. General ontological issues, specific anatomical challenges and potential applications of the HAO have been discussed at several domain oriented meetings (2009–2011 Entomological Society of America meetings, 7th International Congress of Hymenopterists) and at workshops hosted at North Carolina State University (2010), and the Swedish Malaise Trap Project's 3rd Hymenoptera Workshop in Öland, Sweden (2011).

The Hymenoptera Anatomy Ontology continues to be developed and is made available from multiple resources. Hymenopterists will most likely access the HAO through the HAO Portal Website in an online dictionary-style format, but the text versions of the ontology are accessible through several widely used biomedical databases. The project is associated with the greater biomedical ontology community, via the National Center for Biomedical Ontology (http://www.bioontology.org/), which ensures that the HAO will be archived for long-term sustainability and distributed for broad use in other domains. The HAO can be downloaded in either (OWL: http://bit.ly/UnICTE or OBO: http://bit.ly/Tm1n6U) format. 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).

Concepts are continually being added to the HAO at an ever-finer level of granularity. Certain components of Hymenoptera morphology are more difficult to describe under the constraints of structural equivalency, and these have yet to be included in the HAO. In particular wing venation has manifested as difficult because many wing veins have secondarily been lost, thus cannot be referenced under purely structural equivalency criteria without taking into account biological homology (see chapter *3.1.6. Homology and the HAO*, above). Another important component to Hymenoptera descriptive texts are descriptions of surface sculpture, particularly those

referenced by R. Harris (1979). Upon examination of these terms, ideally they should, among others, be included in the Phenotype Quality Ontology (PATO).

The author determined that a number of these terms do occur at a low frequency in Biodiversity Heritage Library species descriptions, with the highest being 'Punctate' as defined by Harris as "Set with fine, impressed points or punctures appearing as pin pricks", found in 25 percent of articles. A number of the top qualitative terms found in Hymenoptera texts are not eligible for inclusion in PATO. These include 'relative terms' or adverbs ending in -ly such as: nearly, somewhat, slightly, entirely, weakly, partly, broadly, and primarily (among others). These terms, as relative terms, are not considered repeatable, or quantitative in a strict sense, and should be avoided in descriptions. Another broad set of terms has potential replacement terms that are suitable. These include: small, short (=PATO:0002364, 'shortened'), similar (=PATO:0000632 symmetrical), and single (= multiple possibilities depending on what is being referenced as single, number 1). The majority of the simple and commonly used Harris terms (reticulate, rugulose, coriaceous, foveolate, crenulate, costate, punctate) do not exist in PATO. However, areolate and foveate do exist in PATO as of May 2011. Their inclusion is likely due to the revision of *Evaniscus* using Entity-Quality format (Mullins, Kawada, Balhoff, & Deans, 2012). Other terms for possible inclusion are: metallic, petiolate, shiny, contiguous, sclerotised, golden, and hyliane.

'Shiny' is presently in PATO but should be a homonym, as the definition that shiny is related in PATO does not match its function in Hymenoptera descriptions. PATO references shiny as having higher saturation of color [PATO:0001229], whereas should be a synonym of glistening [PATO:0001373]. Spatial terms (posterior, anterior) are for the most part already included in PATO. In total, 919 qualitative terms were extracted from the Journal of Hymenoptera Literature using the Proofer tool out of 353 articles. The most commonly used Qualitative terms identified in this author's analysis are listed in figure 17, and similar to the morphological terms (see chapter *4.2. Results for Utilizing the HAO for Literature Analysis*, below); tend to be very general, lacking detail associated with uniqueness for a particular taxon.

Common Anatomy Reference Ontology (CARO) Terminology was incorporated early in the HAO development, however CARO term usage varies substantially. Of the (number) terms incorporated as a foundational ontology in the HAO anatomical space [CARO:0000005] is referenced in definitions 21 times, anatomical line [CARO:0000008] is referenced 14 times, anatomical cluster [CARO:0000041] is referenced 70 times, and Anatomical space [CARO:0000005] is reference 22 times. Area is considered an

anatomical structure [CARO:0000003], thus anatomical structure is indirectly referenced 242 times in the HAO making it one of the most important foundational terms to the HAO. The term 'area' itself was not included in CARO and was defined in the HAO as "The anatomical structure that is delimited by material or immaterial anatomical entities" (http://api.hymao.org/api/ref/67791), despite inclusion already in PATO:0001323 as "A 2-D extent quality inhering in a bearer by virtue of the bearer's two dimensional extent", rendering the HAO 'area' as a necessary homonym of PATO 'area'.

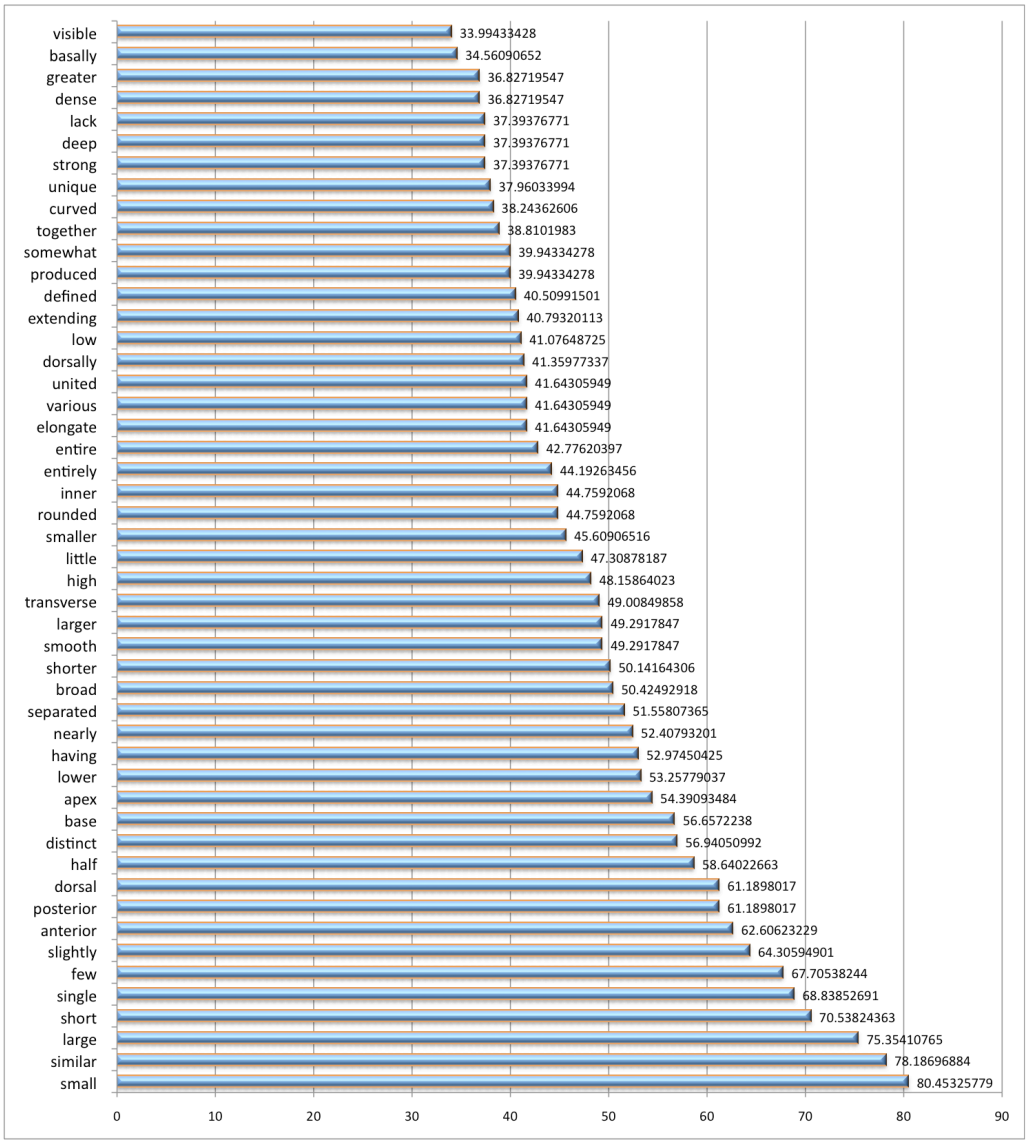

Figure 17: Author's determination of the most commonly used qualitative terms in the Hymenoptera literature. Terms in this figure are ranked based on frequency of occurrence among all articles (i.e., the number of articles in which a term occurred). Data of the author.

4.2. Results for Utilizing the HAO for Literature Analysis

The HAO-based analysis of literature that was conducted within this dissertation shed light on a number of issues related to terminology usage in the Hymenoptera literature and fulfilled aims 1&2 of this dissertation. Our results from this literature analysis indicate that taxonomists use domain-specific terminology that follows taxonomic specialization, particularly at superfamily and family level groupings. Additionally, our results demonstrate that there is a great deal of variability of cluster analysis results depending on the cluster algorithm used. The variability emphasizes that a great deal of noise, as expected, appears in the dataset. From the 353 articles we collected 1189 new morphological terms used by Hymenoptera taxonomists. These terms were added to the mx database, augmenting the development of the Hymenoptera Anatomy Ontology. The most frequently used anatomical terms utilized by hymenopterists in publication are summarized in figure 19 in chapter *4.5. Results from Morphological Exploration for HAO Augmentation*.

The Proofer tool, which the author developed to assist analysis of these articles, significantly improved the efficiency of term extraction from legacy literature by reducing the number of terms presented to the user for review. The author conducted a comparison of the number of terms presented to the user, with and without the Proofer stop words list, for 25 randomly selected articles. This comparison demonstrated that the Proofer stop word list reduced the number of terms displayed to the user by 1/3 of the total actual word count of the article, which was, most notably, an 80% reduction in the number of combinations of words displayed to a user by the Proofer tool. 180 of the 353 articles were identified as containing descriptions of new taxa, wholly or in part. The shortest tree returned from analysis was based on the Sorensen distance with average based cluster analysis algorithm, 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. These clusters represent the entirety of the major superfamily or higher-level groupings present in analysis, as not all superfamilies were included, as they were not represented in the Journal of Hymenoptera Research.

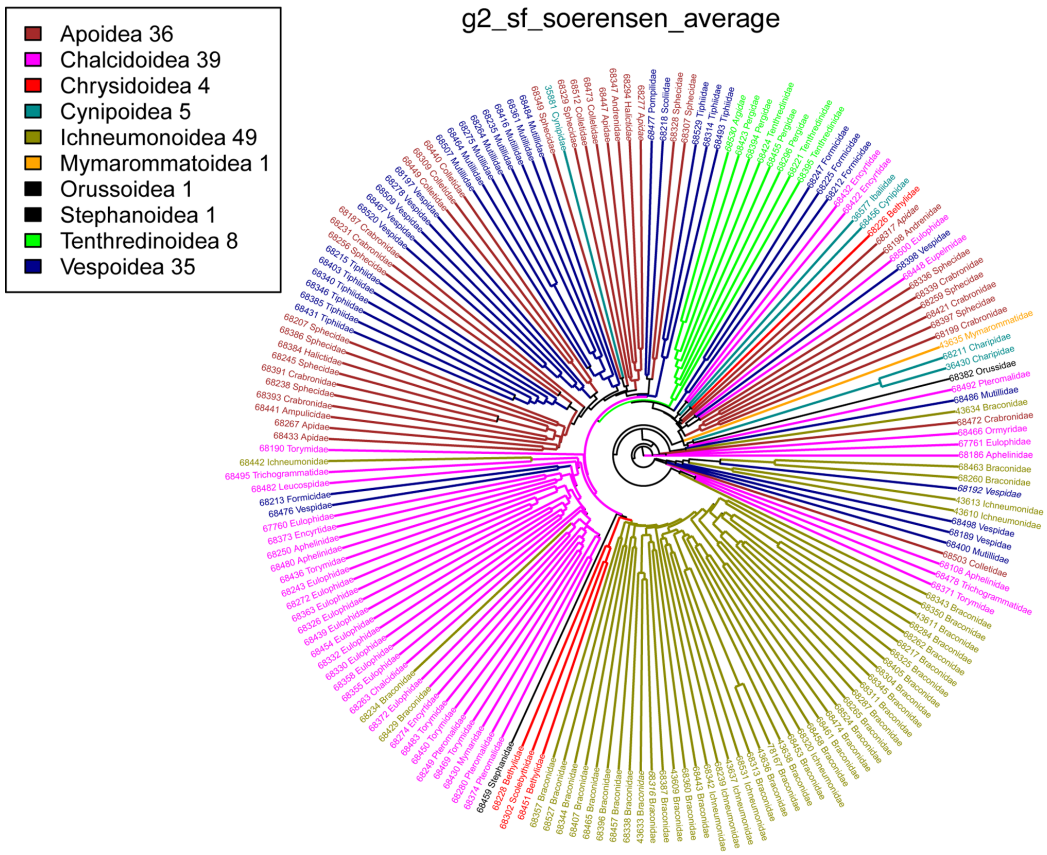

Figure 18: Sorensen-Average untrimmed tree with superfamily name, and number of groupings calculated to superfamily level. The tree represented is the entire, untrimmed tree and the number after the superfamily is the number of groupings retrieved when the tree is trimmed.

4.3. Results for Utilizing the HAO for Literature Annotation

Aim 3 of this dissertation was to develop a methodology for annotating literature by establishing links to HAO concepts prior to publication, and to subsequent promote that methodology to the community. The Seltmann et al. (2012) paper was a direct culmination of discussion with the research community, in particular based on discussion and questions raised during several HAO project workshops. From the time the author and her colleagues first published the URI table concept, a further seven morphology publications (Buffington & Van Noort, 2012; Johnson & Musetti, 2011; Krogmann & Nel, 2012; Mikó et al., 2012; Sharkey & Stoelb, 2012; Talamas et al., 2011; Wharton, Ward, & Mikó, 2012) have adopted HAO terminology by using the URI/Analyzer methodology as pioneered by the author and her colleagues, and reviewed above. In addition, the general level of understanding amongst Hymenoptera

morphologists about anatomy ontology has greatly increased, providing evidence that the project has met one key measurement of success for any bioinformatics tool: its level of adoption by its target community.

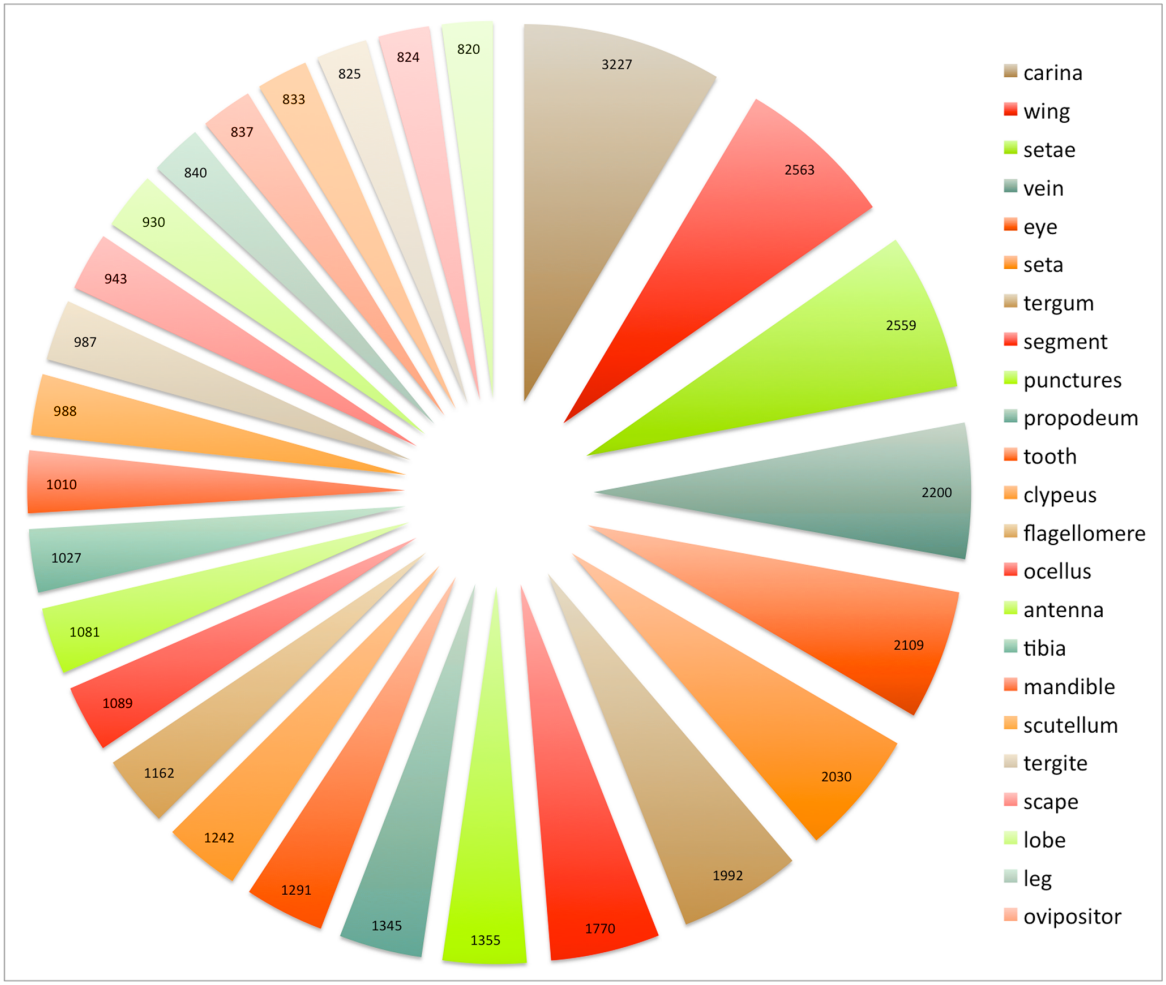

Figure 19: The author's analysis of the most commonly used anatomical terms in the Hymenoptera literature. Terms in this figure are ranked based on frequency of occurrence among all articles (the number of articles in which a term occurred). Number on chart and size of pie represents the number of total times the term occurred in all articles.

4.4. Results from Morphological Exploration for HAO Augmentation

A further primary outcome of this dissertation consisted of a demonstration the specific applicability of the HAO in the field of morphology, specifically to the alignment of hymenoptera mouthpart anatomy to the HAO.

As presented above, the most common morphological terminology as elucidated from the examination of Biodiversity Heritage Library (figure 17) material reveals a preoccupation particularly with external sclerites. Multiple head and mouthpart terms are included in the list including: 'tooth', 'mandible', and 'clypeus'. The inclusion of a morphology component to this dissertation is directly indicative of the evaluation that ontology construction should be conducted, at least in part, by domain experts, and that cross-discipline skills were necessary to successfully contribute to anatomy ontology construction. Below is the description of one model for aculeate mouthpart alignment for the HAO, *Acanthinevania sp.* and the resulting URI table from the effort.

The results indicated here are representative of an ongoing effort to define, finalize, and illustrate the 85 constituent parts identified and included in the HAO under 'Mouthparts'. Constituent parts are those that have the ontological part_of relationship with the 'Mouthparts', or "The anatomical cluster that is composed of the labrum, epipharyngeal wall, hypopharyngeal wall (including the sitophore), mandibles, maxillae, labium and conjunctivae connecting them." [HAO:0000639]. In situ observations of the external sclerites of the maxillo-labial complex in *Acanthinevania* sp. are below, with a single illustration of the infrabuccal pouch as observed in *Evania appendigaster*. Definitions of all terms, abbreviation, HAO URI, and references are included in a URI table format. The URIs found in the table, link back to the Hymenoptera Anatomy Ontology WebPortal pages, where additional information is available.

The maxillo-labial complex consists of the labium, or "the anatomical cluster that is composed of the unpaired glossa, paraglossae, unpaired prementum, unpaired postmentum, labial palps. The labium is situated between the maxillae and continuous anterodorsally with the hypopharyngeal wall" [HAO:0000453] and the maxilla or "the anatomical cluster that consists of cardo, stipes, galea, lacinia and maxillary palps, situated lateral to the labium" [HAO:0000513]. External sclerites are highlighted and internal muscles indicated on the illustrations prepared by the author (results are not yet published).

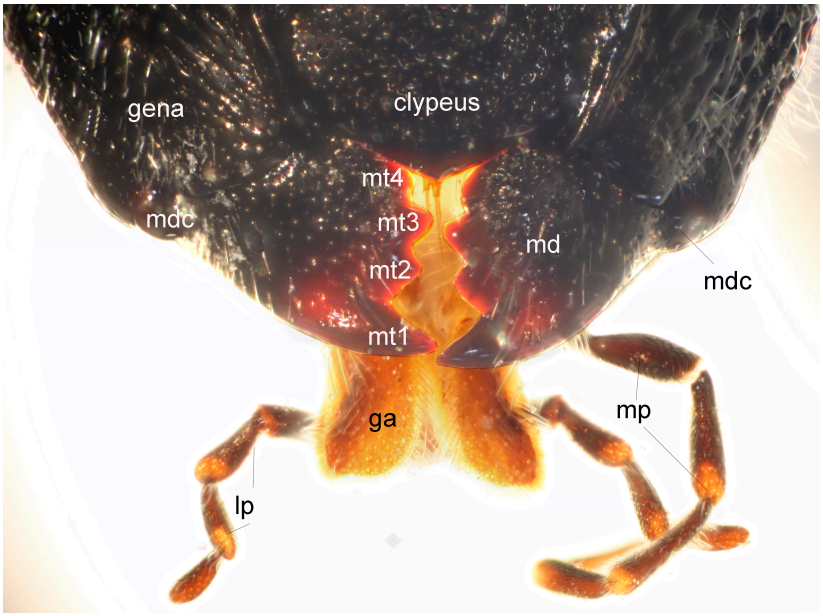

Figure 20: Anterior head of *Acanthinevania sp.* showing an in situ view of the mouthparts. Labial palps (lp), Maxilla palps (mp), mandibular condyle (mdc), exposed galea (ga), mandible (md), and mandibular teeth (mt1-mt4). Author's image and labeling.

4.4.1. Oral Cavity and the Head

Features in this discussion regarding the oral cavity and head of *Acanthinevania sp.* are principally illustrated by figures 20, 21, and 24.

1. **Head and Integument**: In general, the head of the *Acathevania* are very robust, with thick, heavily sculptured integument (figure 20).
2. **Hypostoma (h)**: Site of attachment of the conjunctiva of the maxilla via the cardo, and the cranium. The hypostoma is heavily sclerotised and striated.
3. **Proboscis fossa (pf)**: Triangular, as pictured in figure 21.
4. **Mandible (md)**: With 4 apparent teeth (mt1-4), and enlarged apical tooth (mt1). The mandibles articulate with the gena by the mandibular condyle (mdc) and are very strongly attached through mandibular muscles (ms).
5. **Tentorium (ten)**: Heavily sclerotised, hollow. This major internal apodeme of the head functions as the attachment point for the illustrated **tentorio-antennal muscles (tam**), **dorsal premental adductor (adpr), ventral premental adductor (avpr)**, and **tentorio-stipial muscle (tsm)**.
6. The **clypeo-epipharyngeal muscle (cem)** as pictured in figures 21 and 24, is interwoven with the longitudinal muscles of the sucking pump, which extends parallel to the epistipes.

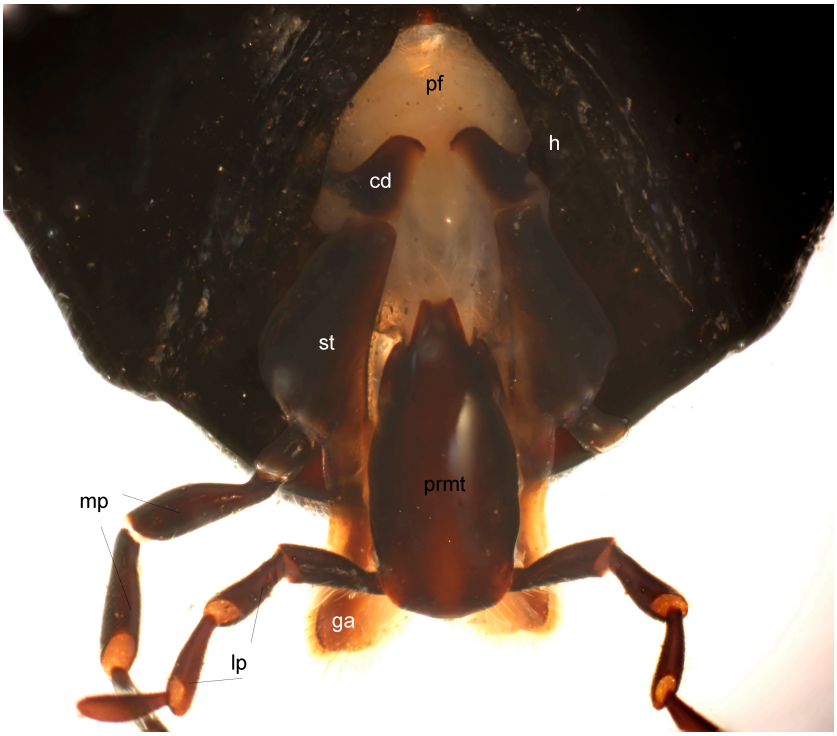

Figure 21: In situ mouthparts of the posterior head of *Acanthinevania sp.* Proboscis fossa (pf), prementum (prmt), galea (ga), maxilla palp (mp), labial palp (lp), stipes (st), cardo (cd), and hypostoma (h). Author's image and labeling.

4.4.2. Maxilla

Features in this discussion regarding the maxilla of *Acanthinevania sp.* are principally illustrated by figures 22.

1. **Galea (ga)**: **Apical lobe of the galea (alg)** strongly fringed with a galeal brush apically with a **basal lobe of the galea (blg)** lateral to the **maxilla palps (mp) -** number of maxilla palps is 6. When at rest the galea extends below the prementum dorsally.
1. **Lacinia (lc)**: Large and bilobed, lateral to the stipes and about the same length.
2. **Epistipes (epi)**: The point of **articulation with the lateral arm (lapmt art)** of the **prementum (prmt)** is greatly reduced, visible only as a sclerotised band at the base of the lacinia, which extends to the stipes. The articulation between labium and maxilla is strong, consisting primarily of soft tissue.
3. **Stipes (st)**: Heavily sclerotised, somewhat triangle shaped dorsally and smooth. The major sclerite of the maxilla articulates with the glossa, cardo, and paraglossa. The **tentorio-antennal muscle (tam)** attaches proximal to the

lateral arm of the prementum articulation, and basal to the cardo.

4. **Cardo (cd)**: thin and articulating with the **occipital foramen (of)**, serving as a hinge-like mechanism for the maxillo-labial complex. The **cranial articulation of the cardo (cac)** is basal to the **cardo (cd)**, articulating with the **hypostoma (h)**, and C-shaped.

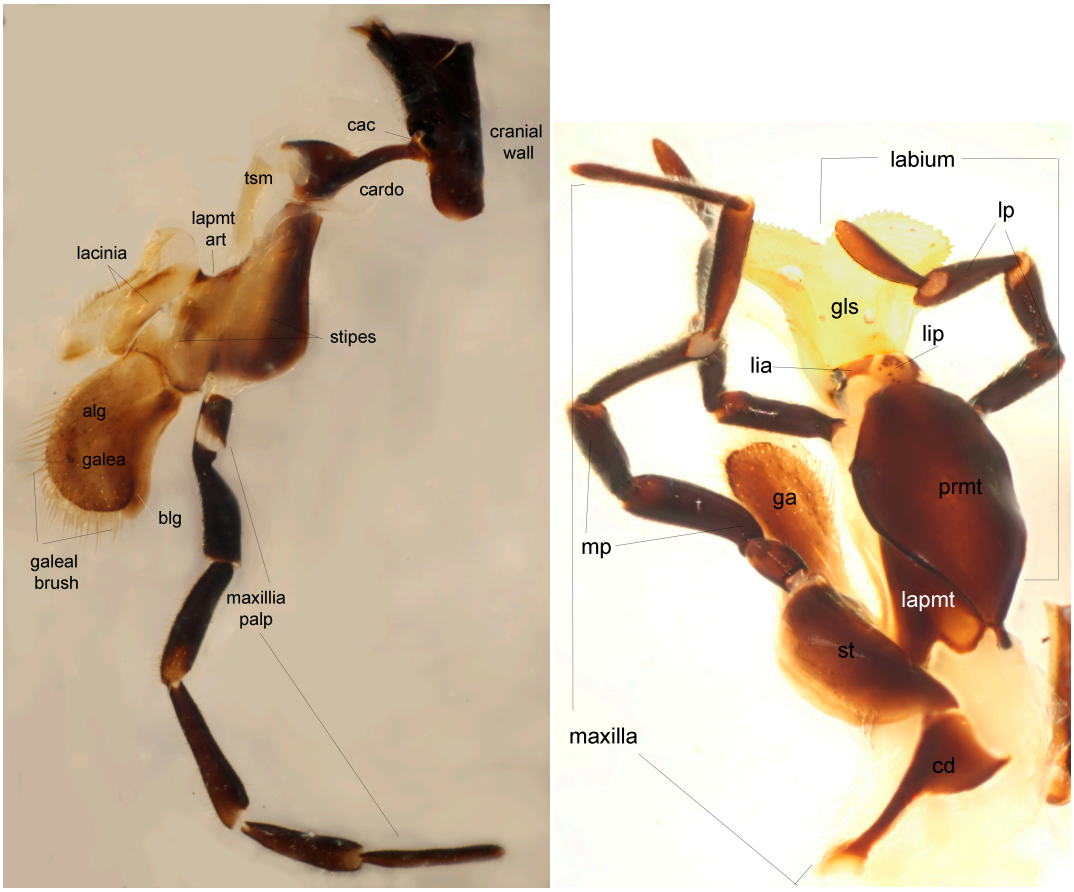

Figure 22: Sclerites of the maxilla (left) and posterior labial complex, viewed dorsally (right), of *Acanthinevania* sp. Cranial articulation of the cardo (cac), tentorio-stipital muscle (tsm), articulation of the lateral arm of the prementum (lapmt art), anterior lobe of the galea (alg), basal lobe of the galea (blg), glossa (gls), ligular plate (lip), ligular arms (lia), galea (ga), stipes (st), prementum (prmt), lateral arms of the prementum (lapmt), cardo (cd), maxilla palp (mp) and labial palp (lp). Author's image and labeling.

4.4.3. Labium

Features in this discussion regarding the labium of *Acanthinevania sp.* are principally illustrated by figures 22, 23, 24, and 25.

1. **Prementum (prmt)**: Sclerite large, extending over half of the length of the labium, not subdivided ventrally. A small round shaped sclerite is contiguous

basally with the premental shield. This structure, or **ligular plate (lp)**, has attached arms extending laterally around the glossa (gl) in a ring-like fashion serving as the attachment for the **paraglossa (pg)**. Palpal fossa of the prementum is non-existent. Absent are submentum, or mentum sclerites.

2. **Lateral arm of prementum (laprmt)**: Lateral extensions of the prementum (prmt), serve as attachment point for the labrum to the maxilla.
3. **Ligula**: Complex composed of the glossa, paraglossa, ligular plate and ligular arms.
4. **Ligular plate (lip)**: Round plate found at apical end of the premental shield (figure 23), attachment for the **paraglossa (pgl)** via the **ligular arms (lia)**.
5. **Glossa (gl)**: Triangular in shape, broader apically. Multiple rows of **glossal ridges (gr)**, comprising of transverse setiferous rows ventrally (figure 23).
6. **Paraglossa (pgl)**: Situated laterally to the **salivary orifice (so)**. Bilobed and sclerotised as an extension of the ligular arms on the lateral sides of the labium. A non-sclerotised portion of the paraglossa, or **paraglossal extension (pgex)**, extends in a triangular shape, parallel to the glossa in thin membrane.
7. **Labial palp (lp)**: 4 palpomeres without significant modification in any of the meres, at least on a gross level.
8. **Functional mouth (fm)**: Anterior, basal position on the labrum. Opening to the digestive system.
9. The **premento-paraglossal muscle (ppm)** inserts on the **posterior glossal plate (pgp).** The **ventral salivarial dilator (dvs)** and **premento-glossal muscle (pgm)** are difficult to discern in figure 24 as both arise from the ventral prementum anterior to the **premento-paraglossal muscle (ppm)** and lack of image clarity**.** The **ventral salivarial dilator (dvs)** and **dorsal salivarial dialator (dds)** are both clearly visible in figure 25, attached to the **salivary duct (sd).**

Figure 23: Posterior labial complex (left) and anterior labial complex (right) of Acanthinevania sp. Lateral arms of the prementum (lapmt), stipes (st), prementum (prmt), galea (ga), ligula (lg), glossa (gls), paraglossal extension (pgex), lacina (lc), epistipes (epi), paraglossa (pgl), and glossal ridges (gr). Author's image and labeling.

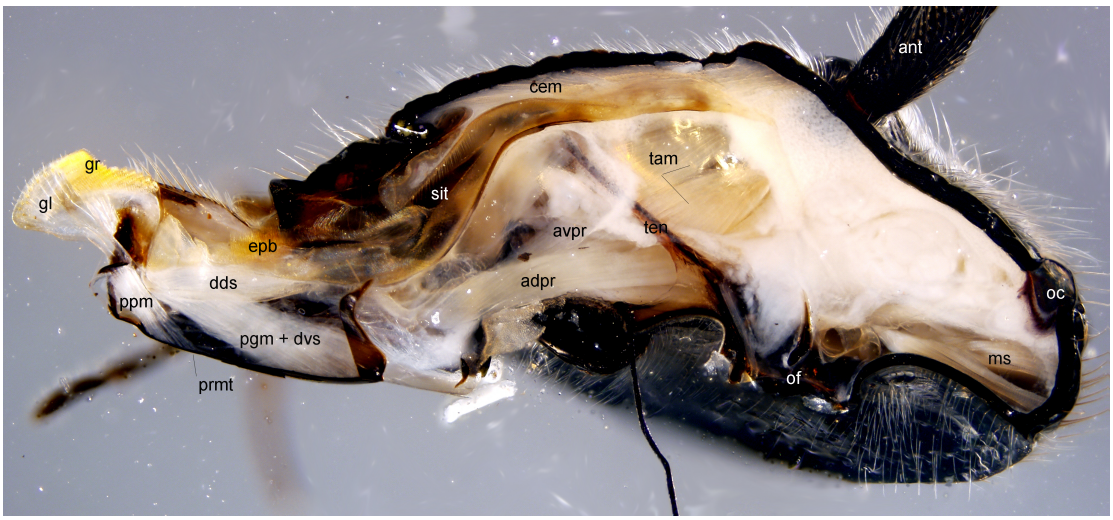

Figure 24: Brightfield median sagittal view of entire Acanthinevania head. Glossa (gl), glossal ridges (gr), premento-paraglossal muscle (ppm), prementum (prmt), dorsal salivarial dilator

(dds), epipharyngeal brush (epb), premento-glossal muscle (pgm), ventral salivarial dilator (dvs), dorsal premental adductors (adpr), ventral premental adductor (avpr), tentorium (ten), tentorio-antennal muscles (tam), sitopore (sit), clypeo-epipharyngeal muscle (cem), antenna (ant), mandibular muscles (ms), and ocellus (oc). Author's image and labeling.

Figure 25: Laser confocal image of median sagittal view of labium of *Acanthinevania*. Clypeo-epipharyngeal muscle (cem), sitopore (sit), dorsal premental adductors (adpr), epipharyngeal wall (ew), functional mouth (fm), galea (ga), lacinia (lc), salivary duct (sd), ventral salivarial dilator (dvs), premento-glossal muscle (pgm), prementum (prmt), dorsal salivarial dilator (dds), paraglossa (pgl), opening of the salivary duct (osd), glossal ridges (gr), glossa (gl), premento-paraglossal muscle (ppm), and posterior glossal plate (pgp). Author's image and labeling.

The muscles of hymenoptera, as other insects, differ from vertebrates in one

particularly interesting way. They consist of loosely associated fibers, gathered in bundles but without a perimysium, or sheath to hold the muscles in place. Therefore it is common to find muscles with fan-like attachments, or a single muscle has been split into two points of attachment (Snodgrass, 1942). It is obvious that there is a lot of variability in muscles, splits in strands and moving from one sclerotised attachment to anther seems fairly plastic. Snodgrass (1956) suggests that the attachment of muscles to the body is by cuticular ingrowths, or tendons and apodemes. He suggests that this is done by bundles of 'threads' referred to as 'tonofibrillae' (ton). It is possible this is true, as we observe that the leading portion of the muscle fluoresces at similar wavelength as the sclerotised cuticle, implying that it is constructed of similar material, although no epidermal cells amongst the muscle fiber were observed (figure 26).

The tentorium (ten) is a major apodeme in the hymenoptera head with its origins at the tentorial pits and its attachment at the gular sulci. The tentorium is an extension of the cuticle, and serves as the attachment site for many cranial and maxillo-labial complex muscles of the head. Interestingly, however we observed that the tentorium is hollow (figure 26), and when the specimen is fresh, can be easily bent without breaking.

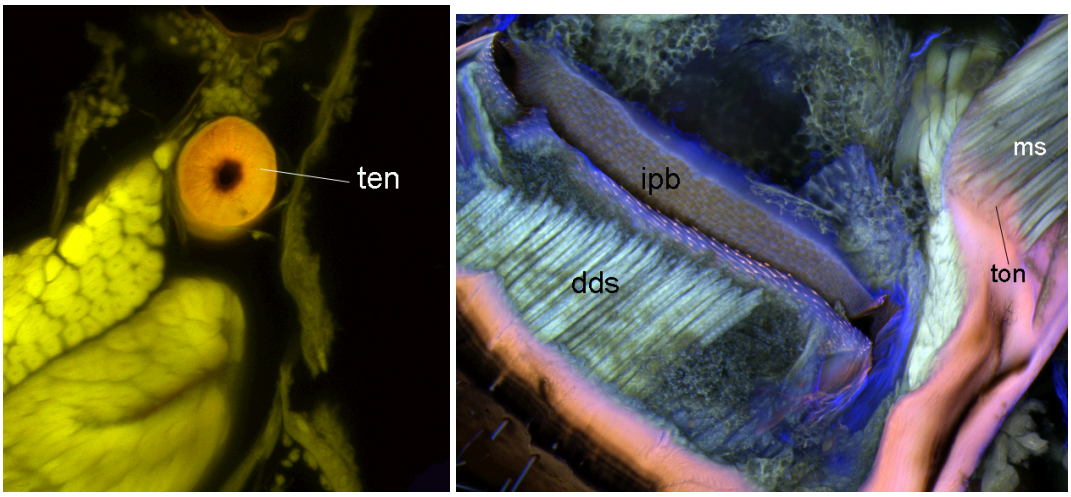

Figure 26: Cross section of one arm of the tentorium (left) and transverse, ventral view with possible 'tonofibrillae' indicated at the initiation of a mandibular muscle (right). Tentorium (ten). infrabuccal pouch (ipb), tonofibrillae (ton), mandibular muscle (ms), and dorsal salivarial dilator (dds). Author's image and labeling.

It is unknown exactly how Evanidae wasps feed. However, based on the illustration of food in the infrabuccal pouch (ibp) (figure 27), and the lack of reduction in the mouthparts it is clear that they do. The functional mouth is situated in the medial,

anterior labrum, just below the clypeus. If the glossa, whose ridges were intended to trap food substances, were to fold forward, the glossal ridges (gr) would come into contact with the epipharyngeal brush (epb). The dorsal salivarial dilators are attached to the infrabuccal pouch ventrally.

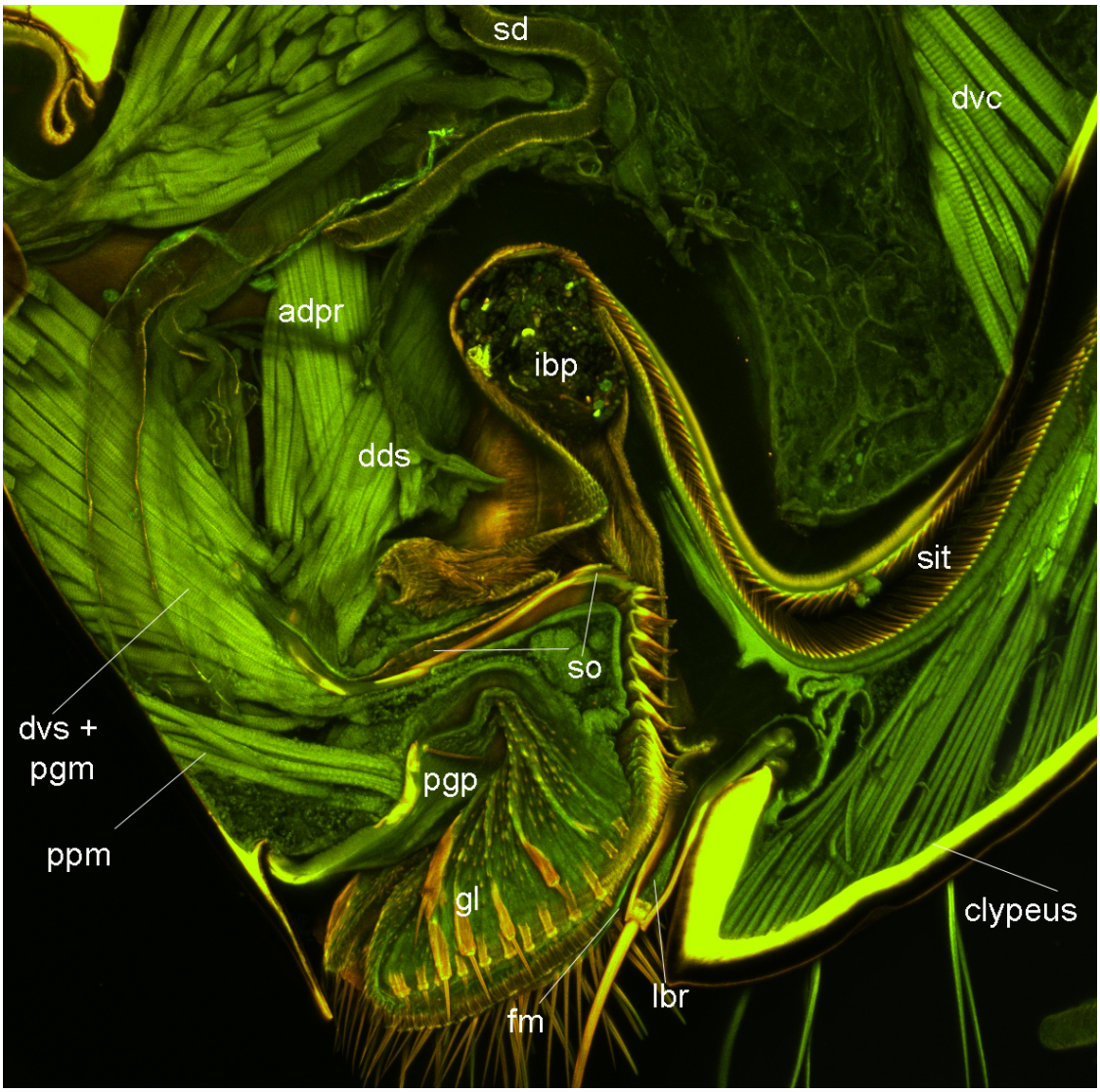

Figure 27: Sagittal view of *Evania appendigaster* mouthparts indicating a full infrabuccal pouch. Ventral salivarial dilator (dvs), premento-glossal muscle (pgm), premento-paraglossal muscle (ppm), posterior glossal plate (pgp), glossa (gl), functional mouth (fm), salivary orifice (so), dorsal salivarial dilator (dds), dorsal premental adductors (adpr), infrabuccal pouch (ibp), sitophore (sit), ventral cibarial dilatator (dvc), and labrum (lbr). Author's image and labeling.

4.5.4. URI Table for 'Mouthparts' Morphology

Term	abbreviation	Concept	URI	References
antenna	ant	The anatomical structure that is composed of ringlike sclerites and the anatomical structures encircled by these sclerites and that is articulated with the cranium.	http://purl.obolibrary.org/obo/HAO_0000101	(Mikó, 2009)
antennal muscle		The muscle that inserts on the antenna.	http://purl.obolibrary.org/obo/HAO_0001172	(Vilhelmsen & Mikó, 2010)
anterior glossal sclerite		The sclerite that is paired and is located on the proximal lateral margin of the glossa. The prementо-glossal muscle inserts on the sclerite.	http://purl.obolibrary.org/obo/HAO_0000112	(Deans, 2009; Vilhelmsen, 2010)
apical lobe of the galea	alg	The raised area that is located apically on the dorsal face of galea separated from the basal lobe by the transverse galeal furrow.		(Seltmann, 2010)
apical tooth	mt1	The tooth that is located apically on the external margin of the mandible.	http://purl.obolibrary.org/obo/HAO_0001972	(Seltmann, 2010)
apodeme		The process that is internal.	http://purl.obolibrary.org/obo/HAO_0000142	(Mikó, 2009)
articulation	art	The area that is located on the sclerite and that makes movable direct contact with another sclerite.	http://purl.obolibrary.org/obo/HAO_0001485	(Ronquist & Nordlander, 1989)
basal lobe of the galea	blg	The raised area that is located on the dorsal face of galea, distally from the transverse galeal furrow and apically from the lacinia.		(Seltmann, 2010)

cardo	cd	The area that is located proximally on the maxilla, proximal to the stipes.	http://purl.obolibrary.org/obo/HAO_0000187	(Karlsson & Ronquist, 2012; Mikó, 2009)
clypeo-epipharyngeal muscle	cem	The epipharyngeal muscle that arises from the clypeus and inserts on the epipharyngeal wall.	http://purl.obolibrary.org/obo/HAO_0000896	(Vilhelmsen & Mikó, 2010)
clypeus		The area that corresponds to the site of origin of the clypeo-epipharyngeal muscle.	http://purl.obolibrary.org/obo/HAO_0000212	(Karlsson & Ronquist, 2012; Ronquist & Nordlander, 1989)
condyle		The articular surface that is convex and is inserted into the fossa of an adjacent sclerite.	http://purl.obolibrary.org/obo/HAO_0000220	(Karlsson & Ronquist, 2012; Ronquist & Nordlander, 1989)
conjunctiva		The area of the integument that is weakly sclerotised, with thin exocuticle.	http://purl.obolibrary.org/obo/HAO_0000221	(Mikó, 2009; Ronquist & Nordlander, 1989)
cranial articulation of the cardo	cac	The sclerite of the cardo that articulates with the hypostoma.		(Seltmann, 2010)
cranium		The sclerite that is articulated with the cervical prominence, the scapes and the mandibles.	http://purl.obolibrary.org/obo/HAO_0000234	(Deans, 2009)
cuticle		The acellular anatomical structure that is the external layer of the integument (covers the entire body surface as well as lines ectodermal	http://purl.obolibrary.org/obo/HAO_0000240	(Mikó, 2009; Nichols, 1989)

		invaginations such as the stomodeum, proctodeum and tracheae) and produced by the epidermal cells.		
dorsal premental adductors	adpr	The tentorio-labial muscle that arises from the tentorial arm and inserts on the distal end of the hypopharyngeal rod or on the proximodorsal corner of the anterior glossal sclerite, lateral to the distal opening of the salivarium.	http://purl.obolibrary.org/obo/HAO_0000264	(Vilhelmsen, 1996, 2010)
dorsal salivarial dilator	dds	The salivarial muscle that arises from the proximal part of the hypopharyngeal rod or the distal end of the lateral arm of the prementum and inserts on the dorsal salivarial wall.	http://purl.obolibrary.org/obo/HAO_0000274	(Vilhelmsen, 2010)
epipharyngeal brush	epb	The setiferous patch that is unpaired, sublateral and situated distally on the epipharyngeal wall on the left torma.	http://purl.obolibrary.org/obo/HAO_0001765	(Vilhelmsen, 2010)
epipharyngeal wall	ew	The conjunctiva that extends between the distal margin of the labrum and the proximal boundary of the cibarium.	http://purl.obolibrary.org/obo/HAO_0000300	(Vilhelmsen, 1996)
epistipes	epi	The basal sclerotised part of lacinia that articulates between the juncture of the lateral arm of the prementum, lacinia and the stipes.		(Seltmann, 2010)
foramen		The anatomical space that is surrounded by sclerites and allows for the passage of haemolymph, nerves and tracheae.	http://purl.obolibrary.org/obo/HAO_0000345	(Mikó, 2009)
fossa		The patch that is impressed and corresponds to an apophysis.	http://purl.obolibrary.org/obo/HAO_0000718	(Gibson et al., 1998)

functional mouth	fm	The anatomical space that is delimited posteriorly by the distal part of the sitophore and anteriorly by the tormae.	http://purl.obolibrary.org/obo/HAO_0000361	(Deans, 2009; Vilhelmsen, 1996)
galea	ga	The lobe that is located on the maxilla distal to the stipes and lateral to the lacinia.	http://purl.obolibrary.org/obo/HAO_0000368	(Karlsson & Ronquist, 2012; Mikó, 2009)
gena		The area that is delimited by the intersection of the interorbital plane, the margin of the compound eye, the margin of the oral foramen, the occipital carina and the malar sulcus.	http://purl.obolibrary.org/obo/HAO_0000371	(Mason & Huber, 1993; Yoder, 2009)
glossa	gl	The lobe that is median, unpaired and is situated on the labium distally on the salivarial orifice. The glossa is the median labial endite.	http://purl.obolibrary.org/obo/HAO_0000376	(Deans, 2009; Karlsson & Ronquist, 2012; Snodgrass, 1935)
glossal ridges	gr	Multiple transverse rows of setae on the ventral side of the glossa.		(Seltmann, 2010)
gular sulcus		The sulcus that corresponds with the postgenal ridge.	http://purl.obolibrary.org/obo/HAO_0000780	(Mikó, 2009)
head		The tagma that is located anterior to the thorax.	http://purl.obolibrary.org/obo/HAO_0000397	(Deans, 2009; Snodgrass, 1935)

hypostoma	h	The area that extends on the posterior (ventral) margin of the oral foramen along the site of attachments of the conjunctiva connecting the cranium with the maxillae and is delimited laterally by the pleurostomal fossa.	http://purl.obolibrary.org/obo/HAO_0000411	(Mikó, 2009; Vilhelmsen, 1999).
infrabuccal pouch	ibp	The pouch that is situated on the hypopharyngeal wall distally of the sitophore.	http://purl.obolibrary.org/obo/HAO_0001563	(Vilhelmsen, 1996)
integument		The anatomical system that forms the covering layer of the animal, ectodermal in origin and composed of epidermal cells producing the cuticle.	http://purl.obolibrary.org/obo/HAO_0000421	(Mikó, 2009; Mikó, Yoder, Bertone, & Deans, 2009; Nichols, 1989)
intrinsic maxillary palp muscle	impm	The maxillar muscle that arises from a maxillary palpal segment and inserts on the palpal segment distal to it.	http://purl.obolibrary.org/obo/HAO_0001557	(Vilhelmsen & Mikó, 2010)
labial palp	lp	The palp that is situated on the labium articulating laterally on the prementum.	http://purl.obolibrary.org/obo/HAO_0000450	(Karlsson & Ronquist, 2012; Vilhelmsen, 1996)
labium		The anatomical cluster that is composed of the unpaired glossa, paraglossae, unpaired prementum, unpaired postmentum, labial palps. The labium is situated between the maxillae and continuous anterodorsally with the hypopharyngeal wall.	http://purl.obolibrary.org/obo/HAO_0000453	(Deans, 2009)

labrum	lbr	The sclerite that is situated along the distal margin of the clypeus and is connected along its proximal margin with the distal margin of the epipharyngeal wall.	http://purl.obolibrary.org/obo/HAO_0000456	(Karlsson & Ronquist, 2012; Snodgrass, 1935)
lacinia	lc	The lobe that is located on the maxilla, distal to the stipes and median to the galea.	http://purl.obolibrary.org/obo/HAO_0000457	(Mikó, 2009)
ligula		The area that is located distally on the labium and formed from the combined glossae and paraglossae.	http://purl.obolibrary.org/obo/HAO_0000496	(Mikó, 2009)
ligular arms		The process that is continuous proximally with the anterodistal edge of the lateral part of the prementum and distally with the anterior glossal sclerite.	http://purl.obolibrary.org/obo/HAO_0000408	(Mikó et al., 2009; Snodgrass, 1956)
ligular plate	lp	The sclerite extending from the end of the prementum, and lateral to the ligular arms.		(Snodgrass, 1956)
lateral arm of the prementum	lapmt	The dorsal sclerotic projection of the prementum, bounded laterally by the hypopharynx, and being in connection with the epistipes.		(Seltmann, 2010)
longitudinal muscle of the sucking pump	lmsp	The epipharyngeal muscle that is unpaired and arises from the distal part of the pharyngeal wall and inserts along most of the length of the proximal epipharyngeal wall.	http://purl.obolibrary.org/obo/HAO_0000500	(Vilhelmsen & Mikó, 2010; Vilhelmsen, 1996)

mandible	md	The sclerite that is connected to the cranium along the anterior margin of the oral foramen via the anterior and posterior cranio-mandibular articulations.	http://purl.obolibrary.org/obo/HAO_0000506	(Karlsson & Ronquist, 2012; Mason & Huber, 1993; Mikó, 2009)
mandibular condyle	mdc	The condyle that is located ventrally (posteriorly) on the proximolateral edge of the mandible and inserts into the pleurostomal fossa.	http://purl.obolibrary.org/obo/HAO_0000508	(Karlsson & Ronquist, 2012; Michener, 2000; Mikó, 2009; Ronquist & Nordlander, 1989)
mandibular muscle	ms	The head muscle that inserts on the mandible.	http://purl.obolibrary.org/obo/HAO_0001779	(Vilhelmsen & Mikó, 2010)
mandibular tooth	md1-md4	The projection that is located distally on the mandible.	http://purl.obolibrary.org/obo/HAO_0001019	(Mikó, 2009; Sharkey & Wharton, 1997)
maxillo-labial complex		The anatomical cluster that is composed of the labium and maxillae.	http://purl.obolibrary.org/obo/HAO_0000452	(Deans, 2009)
maxilla		The anatomical cluster that consists of cardo, stipes, galea, lacinia and maxillary palps. The maxilla is situated lateral to the labium.	http://purl.obolibrary.org/obo/HAO_0000513	(Gibson et al., 1998; Mikó, 2009)

maxillary palp	mp	The palp that is located on the maxilla articulating laterally on the stipes.	http://purl.obolibrary.org/obo/HAO_0000515	(Deans, 2009; Karlsson & Ronquist, 2012; Sharkey & Wharton, 1997)
muscle of the maxillary palpus	spm	A large muscle arising in stipes this muscle as an extensor of the galea inserted on the base of the latter near the palpus. In Vespula, according to Duncan, there are two muscles with a common insertion on the palpus.		(Snodgrass, 1942)
mentum		The sclerite that articulates basally with the submentum and apically with the prementum.	http://purl.obolibrary.org/obo/HAO_0000534	(Mason & Huber, 1993; Mikó, 2009; Vilhelmsen, 1996)
meres		The sclerite that is ring-like and is not attached to muscles.	http://purl.obolibrary.org/obo/HAO_0000096	(Mikó et al., 2009)
mouthparts		The anatomical cluster that is composed of the labrum, epipharyngeal wall, hypopharyngeal wall (including the sitophore), mandibles, maxillae, labium and conjunctivae connecting them.	http://purl.obolibrary.org/obo/HAO_0000639	(Mason & Huber, 1993)
occipital foramen	of	The foramen that is delimited dorsally by the postocciput.	http://purl.obolibrary.org/obo/HAO_0000347	(Mikó, 2009; Ronquist & Nordlander, 1989)

ocelli	oc	The multi-tissue structure that is located on the top of the head, composed of the corneal lens, pigment cell, rhabdoms and synaptic plexus.	http://purl.obolibrary.org/obo/HAO_0000661	(Gibson et al., 1998; Karlsson & Ronquist, 2012; Mikó, 2009)
oesophagus	oes	The tube-like area with longitudinal and transverse folds which extends through the thorax and petiole.		
proboscis fossa	pf	The anatomical space that contains the functional mouth.	http://purl.obolibrary.org/obo/HAO_0000669	(Deans, 2009)
palp		The anatomical cluster that is composed of the palpal segments.	http://purl.obolibrary.org/obo/HAO_0000683	(Mikó et al., 2009)
palpal fossa of the prementum		The fossa that accommodates the base of the labial palp.	http://purl.obolibrary.org/obo/HAO_0001978	(Popovici, Seltmann, & Mikó, in prep)
palpus		The anatomical cluster that is composed of the palpal segments.	http://purl.obolibrary.org/obo/HAO_0000683	(Deans, 2009; Snodgrass, 1935)
paraglossa	pgl	The lobe that is submedian, paired, and is situated distally on the labium, laterally of the glossa.	http://purl.obolibrary.org/obo/HAO_0000686	(Deans, 2009; Karlsson & Ronquist, 2012)
paraglossal extension	pgex	A non-sclearotised portion of the paraglossa extending in a triangular shape, parallel to the glossa, and covering the paraglossa by a thin membrane.		(Seltmann, 2010)
posterior glossal plate	pgp	The sclerite that is located ventrally of the paraglossae and is composed of two fused posterior glossal sclerites.	http://purl.obolibrary.org/obo/HAO_0000747	(Deans, 2009; Vilhelmsen, 1996)
premento-glossal muscle	pgm	The labial muscle that arises on the ventral part of the prementum, laterally to the ventral premento-salivarial	http://purl.obolibrary.org/obo/HAO_0000377	(Vilhelmsen, 2010)

		muscle and inserts on the anterior glossal sclerite.		
premento-paraglossal muscle	ppm	The labial muscle that arises from the ventral part of the prementum, anterior to the origin of the premento-glossal muscle and ventral premento-salivarial muscle and inserts on the posterior glossal sclerite.	http://purl.obolibrary.org/obo/HAO_0000687	(Vilhelmsen, 2010)
prementum	prmt	The sclerite that is unpaired and is situated ventrally on the labium proximal to the endites and palps. The sclerite laterally extends towards the hypopharyngeal wall.	http://purl.obolibrary.org/obo/HAO_0000804	(Gibson et al., 1998; Karlsson & Ronquist, 2012; Mikó, 2009; Snodgrass, 1935)
salivary duct	sd	The gland canal that is part of the salivary gland.		(Snodgrass, 1935)
salivarial muscle		The head muscle that inserts on the dorsal wall of the salivarium or the salivarial sclerite.	http://purl.obolibrary.org/obo/HAO_0001110	(Vilhelmsen & Mikó, 2010)
salivarial orifice (=opening of the salivary duct)	so, osd	The anatomical space that is located on the boundary of the hypopharyngeal wall and the labium. The orifice is situated dorsally of the base of the glossa and corresponds to the opening for the salivarium.	http://purl.obolibrary.org/obo/HAO_0001683	(Vilhelmsen 2010; Seltmann 2010)
sitophore	sit	The sclerite that is located in the proximal part of the hypopharygeal wall delimited distally by the functional mouth and proximally by the proximal boundary of the cibarium. The tentorio-hypopharyngeal muscle inserts on the proximal	http://purl.obolibrary.org/obo/HAO_0000939	(Deans, 2009; Vilhelmsen, 1996)

		margin of the sitophore.		
stipes	st	The sclerite that is located in the maxilla and articulates proximally with the cardo, distally with the galea and lacinia, and laterally with the maxillary palp.	http://purl.obolibrary.org/obo/HAO_0000958	(Gibson et al., 1998; Karlsson & Ronquist, 2012; Mikó, 2009; Snodgrass, 1956)
submentum		The anatomical cluster that is composed of the sclerites that are located between the prementum and the hypostoma.	http://purl.obolibrary.org/obo/HAO_0000785	(Matsuda, 1957; Mikó et al., 2009; Vilhelmsen, 1996)
tentorial pits		The pit on the cranium that corresponds to the tentorium.	http://purl.obolibrary.org/obo/HAO_0000999	(Deans, 2009; Snodgrass, 1935)
tentorio-antennal muscles	tam	The muscle that arises from the dorsal surface of the anterior broadened part of the anterior tentorial arm and inserts on the scape.	http://purl.obolibrary.org/obo/HAO_0001000	(Mikó, 2009; Mikó, Vilhelmsen, Johnson, Masner, & Pénzes, 2007)
tentorio-stipital muscle	tsm	The maxillar muscle that arises on the posteroventral part of the anterior tentorial arm and inserts on the stipes.	http://purl.obolibrary.org/obo/HAO_0001002	(Mikó, 2009; Mikó et al., 2007)

tentorium	ten	The apodeme that has its sites of origins marked by the anterior and posterior tentorial pits and gular sulci.	http://purl.obolibrary.org/obo/HAO_0001003	(Gibson et al., 1998; Karlsson & Ronquist, 2012; Mikó et al., 2007)
tonofibrillae	ton	The bundles of fine threads that are points of attachment that connect muscles to apodemes or tendons.		(Snodgrass, 1956)
ventral cibarial dilatator	dvc	The hypopharyngeal muscle that is unpaired and arises from the median anterior part of the tentorial bridge and inserts on the proximal part of the ventral wall of the cibarium or on the proximal margin of the sitophore, immediately distally to the point where the alimentary canal passes between the lateral arms of the sitophore.	http://purl.obolibrary.org/obo/HAO_0001057	(Vilhelmsen, 1996, 2010)
ventral premental adductor	avpr	The tentorio-labial muscle that arises from the tentorial arm and inserts on the proximolateral corner of the ventral part of the prementum.	http://purl.obolibrary.org/obo/HAO_0001064	(Vilhelmsen, 1996, 2010)
ventral salivarial dilator	dvs	A flat muscle arising on anterior lateral margin of prementum, inserted medially on anterior wall of salivary syringe. (Protractor liguae Wolff; dilator ampullae superior Morison; dilator of the salivarium Duncan.)		(Snodgrass, 1942; Vilhelmsen, 1996)

Table 4: URI table for Mouthparts evaluation of *Acanthinevania sp*, as prepared by the author. Terms without URIs are still under review.

5. GENERAL DISCUSSION

This dissertation collected a range of research associated with the development of the Hymenoptera Anatomy Ontology and its use for the amelioration of descriptions in Hymenoptera research literature, to the computer-assisted analysis of Hymenoptera literature, and to the application of the HAO in the domain of Hymenoptera mouthpart morphology.

5.1. The Hymenoptera Anatomy Ontology (HAO)

The core of this dissertation was based on contributions to, the development of tools for, and applications of, the Hymenoptera Anatomy Ontology. The HAO is a resource based on a foundation of explicitly defined anatomical concepts and a straightforward mechanism for referencing these concepts (URIs). This resource is intended for any and all users who reference Hymenoptera anatomy, not only descriptive taxonomists. The HAO, like other biological ontology efforts, is rapidly evolving, both in its underlying data and its application. The adoption to the addition of anatomy ontology first and foremost depends on the support of the hymenopterist community, whose papers represent the bulk of the fundamental anatomical expertise. Referencing the HAO in publication has the potential to increase the repeatability of our research, and open up their interpretation and use to a much broader array of biologists than just highly specialized taxonomists. Fundamentally, all aspects of science are potentially impacted which rely on the correct interpretation of anatomical structures including: biodiversity, host-parasite biology, collections digitization, genomics, ecology, evolutionary developmental biology (evo-devo), invasive species evaluation, agro-ecosystem management, and biological control. An example of the importance of correct interpretation of morphological structures is highlighted in Mikó et al. (2012), which demonstrated that the misinterpretation of morphological features had resulted in a wing-appendage hypothesis for the membracid 'helmet' (Prud'homme et al., 2011), where in actuality the helmet is the first tergite (T1) in its entirety.

5.2. Utilizing the HAO for Literature Analysis

Cluster analysis lends evidence for the observation that the Hymenoptera community tends to use granular, domain-specific (i.e., taxon-specific) terminology. Datasets were analyzed extensively using different permutations of clustering methods and datasets delimited by term occurrence. As expected, a large amount of variation in the number of groups recovered was observed in the analysis results. Not all hymenopteran families were in analysis, because taxonomic descriptions of some groups were not published in *JHR* between 1993-2007. Also, there is a strong bias in the number of papers concerning Ichneumonoidea and Chalcidoidea represented. This is due, in part, to the large number of taxonomists interested in these diverse superfamilies. Despite these idiosyncrasies in the data, obvious groupings for Ichneumonoidea, Chalcidoidea, 'Symphyta' and Aculeata were retrieved. On a more detailed level, many family level groupings were recovered, demonstrating that we can group articles, and those taxa described in the articles, simply by the terms used to describe those organisms. In the comparison of cluster analysis, the number of clusters recovered decreased with an increase of characters (terms) used in the analysis. The terms found more commonly are generally used across Hymenoptera. Terms like head, wing, and carina are almost universally used, and thus provide very little signal to group articles. In order to capture the variation in the terminology of the authors, to manifest any observable signal in the analysis, much less frequently used terms needed to be included.

5.3. Utilizing the HAO for Literature Annotation

The Proofer software application facilitated a computer-assisted analysis of the Hymenoptera literature. The workflow provided by this tool holds a great deal of potential as a tool for ontology development, and, in particular, for the efficient analysis of descriptive text to extract new terms that can supplement the construction of anatomy ontologies. The workflow requires significant input of domain experts, and open access publications, resulting in the collection of 1189 new terms for the HAO. Although the Proofer tool accumulated numerous terms for inclusion in the HAO, mapping terms to existing classes or creating ontology compliant definitions for those concepts require further expertise and citation. At present (November, 2012) only 144

of the collected terms are tied to HAO concepts. To define concepts, HAO curators initially selected literature that was generally inclusive, taxonomically, of Hymenoptera (including glossary and online resources). This process was done, in part, prior to looking at the BHL JHR articles. Most of the very common terms were already included in the database prior to the term discovery exercise, leaving predominantly highly granular, and superfamily-specific terms used in taxonomic descriptions to be discovered using the Proofer, accounting for the low number of these terms presently fully incorporated in the HAO. These terms will be utilized as the HAO continues to grow, and curators focus on publications from domain experts working exclusively within superfamilies, as this is where the term granularity is demonstrated.

The importance of domain expertise in the process of incorporating these terms cannot be overstated. Jensen & Bork (2010) clearly regards input from biologists as necessary for success in biomedical, ontology based literature mining and Dahdul et al. (2010a) described the importance of taxon experts for phenotype annotation curation. Our evaluation concurs with their observation and extends the thought to conclude that granular terminology is necessary to capture morphological variation, but it requires domain expertise, and evaluation of their publications, in the process, to identify these terms and fully utilize them in the ontology.

The corpus of biological literature will continue to grow and with it the need for more automated methods to utilize and discover the information contained within the articles. Natural language processing methods for biological data discovery is only possible through open access publications, and efforts such as the Biodiversity Heritage Library to make legacy literature freely available. The next necessary ingredient is the ability to reference controlled vocabularies, or ontologies to clearly define the contents of the articles. The core mechanism by which one can reference the Hymenoptera Anatomy Ontology, URIs, is implementable now via simple tables, whose building is facilitated via the Analyzer tool. This ability is an important first step toward transparently and seamlessly integrating the HAO into the scientists' workflow. However, URI tables are only a first step towards adoption of the great potential utilizing the Web and online publications for scientific discovery may hold. Mullins et al. (2012) demonstrated the potential of further linking controlled vocabularies to descriptive taxonomy as discussed in Deans et al. (2012), where entire descriptive statements are written in Entity-Quality format (Balhoff et al., 2010), allowing for the entire corpus of these statements to be reasoned over.

5.4. Morphological Exploration for HAO Augmentation

The final component of this dissertation consisted of a case-study application demonstrating the utility of the HAO in the domain of morphology, and in the descriptive characterization of Hymenoptera mouthpart morphology. The utilization of a URI table to summarize the results makes it possible to neatly clarify concepts, and allows for the discussion to focus on hypotheses regarding morphological features. Detailed and highly annotated work on the labial-maxillary complex in Hymenoptera is now possible through novel laser confocal imaging and dissection techniques, elucidating the musculature and other internal components, as evidenced in the work presented here. Future directions in this exploration are anticipated to include the following: further exploration of the variation of Mouthparts in Hymenoptera, a more comprehensive study of the gustatory system of evaniids, likely involving observations of living organisms feeding, and the illustration of the complex glandular system of the labium.

ACKNOWLEDGEMENTS

I would like to graciously thank my advisors Dr. Zsolt Pénzes and Dr. Andrew R. Deans for generous their support. Dr. Matthew J. Yoder, Dr. István Mikó, Dr. Matthew Bertone, Jim Balholf, Andrew Ernst, and Patricia Mullins who taught me a great deal. Elizabeth MacLeod and Kelly Dew who worked in the lab for many of the years of the project. Mattias Forshage for his philosophical insights regarding structural equivalency. Dr. Eva Johannes for assistance with laser confocal imaging. The International Society of Hymenopterists for having the forethought and inclination to put their journal on the Biodiversity Heritage Library Website. Additional thanks are extended to my doctoral committee members: Dr. George Melika, Dr. János Gausz, Dr. László Gallé, and Dr. Gábor Rákhely for their continued support and advisement, as well as all my past advisors, who there were many. The National Science Foundation primarily funded research for this dissertation, under grant DBI 0850223, with additional support by TÁMOP-4.2/B-09/1/KONV-2010-0005.

REFERENCES

Agosti, D., & Egloff, W. (2009). Taxonomic information exchange and copyright: the Plazi approach. *BMC research notes*, *2*, 53. doi:10.1186/1756-0500-2-53

Akella, L. M., Norton, C. N., & Miller, H. (2012). NetiNeti: discovery of scientific names from text using machine-learning methods. *BMC bioinformatics*, *13*(1), 211. doi:10.1186/1471-2105-13-211

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*.

Bertone, M.A., Yoder, M., Seltmann, K. C., Mikó, I., & Deans, A. R. (2010). *Ontology Primer*. Retrieved from http://bit.ly/T45vuQ

Beutel, R. G., & Vilhelmsen, L. (2007). Head anatomy of Xyelidae (Hexapoda: Hymenoptera) and phylogenetic implications. *Organisms Diversity & Evolution*, *7*(3), 207–230. doi:10.1016/j.ode.2006.06.003

Biodiversity Heritage Library. (2011). Biodiversity Heritage Library. Retrieved February 1, 2011, from http://www.biodiversitylibrary.org

Bolton, B. (1994). *Identification Guide to the Ant Genera of the World*. Cambridge, Massachusetts: Harvard University Press.

Bradly, J. C. (1908). The Evaniidae, ensign flies, an archaic family of Hymenoptera. *Transactions of the American Entomological Society*, *34*, 101–194.

Brothers, D. J. (1975). Phylogeny and classification of the Aculeate Hymenoptera with special reference to Mutillidae. *Heritage*, *50*(11), 483–648.

Brothers, D. J., Carpenter, J. M. (1993). Phylogeny of Aculeata: Chrysidoidea and Vespoidea (Hymenoptera). *Journal of Hymenoptera Research*, *2*(2), 227–304.

Brower, A. V. Z., & Schawaroch, V. (1996). Three Steps of Homology Assesstment. *Cladistics*, *12*(3), 265–272. doi:10.1111/j.1096-0031.1996.tb00014.x

Buffington, M. L., & Van Noort, S. (2012). Revision of the afrotropical oberthuerellinae (cynipoidea, liopteridae). *ZooKeys*, (202), 1–154. doi:10.3897/zookeys.202.2136

Carpenter, J. M., Engel, M., Sharkey, M. J., & Heraty, J. M. (2003). Hymenoptera Assembling the Tree of Life Proposal. *National Science Foundation Proposal*. Retrieved from http://www.hymatol.org

Carpenter, J. M., & Wheeler, W. C. (1999). Towards simultaneous analysis of morphological and molecular data in Hymenoptera. *Zoologica Scripta*, *28*(1-2), 251–260. doi:10.1046/j.1463-6409.1999.00009.x

Castro, L. R., & Dowton, M. (2006). Molecular analyses of the Apocrita (Insecta: Hymenoptera) suggest that the Chalcidoidea are sister to the diaprioid complex. *Invertebrate Systematics*, *20*, 603–614. Retrieved from http://www.publish.csiro.au/?act=view_file&file_id=IS06002.pdf

Castro, L. R., & Dowton, M. (2007). Mitochondrial genomes in the Hymenoptera and their utility as phylgenetic markers. *Systematic Entomology*, *32*, 60–69.

Cui, H. (2012). CharaParser for fine-grained semantic annotation of organism morphological descriptions. *Journal of the American Society for Information Science and Technology*, *63*(4), 738–754. doi:10.1002/asi.22618

Dahdul, W. M., Balhoff, J. P., Engeman, J., Grande, T., Hilton, E. J., Kothari, C., Lapp, H., et al. (2010a). 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. (2010b). The teleost anatomy ontology: anatomical representation for the genomics age. *Systematic biology*, *59*(4), 369–83. doi:10.1093/sysbio/syq013

Davis, R. B., Baldauf, S. L., & Mayhew, P. J. (2010). The origins of species richness in the Hymenoptera: insights from a family-level supertree. *BMC evolutionary biology*, *10*, 109. doi:10.1186/1471-2148-10-109

Day, D., Aberdeen, J., Hirschman, L., Kozierok, R., Robinson, P., & Vilain, M. (1997). Mixed-initiative development of language processing systems. *Proceedings of the fifth conference on applied natural language processing* (pp. 348–355). Morristown, NJ, USA: Association for Computational Linguistics. doi:10.3115/974557.974608

Day-Richter, J. (2012). OBO Flat File Format. Retrieved from http://www.geneontology.org/GO.format.obo-1_2.shtml

De Pinna, M. G. G. (1991). Concepts and tests of homology in the cladistic paradigm. *Cladistics*, *7*, 367–394.

Deans, A. R. (2009). Curator. *Hymenoptera Anatomy Ontology*.

Deans, A. R., Mikó, I., Wipfler, B., & Friedrich, F. (2012). Evolutionary phenomics and the emerging enlightenment of arthropod systematics. *Invertebrate Systematics*, *26*(3), 323. doi:10.1071/IS12063

Deans, A. R., & Ronquist, F. (2006). Ontologizing morphological terms for Hymenoptera - implementing and benefiting from a controlled vocabulary. *6th International Conference of Hymenopterists*. Sun City, South Africa.

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

Deans, A.R., Huben, M. (2003). World, Annotated Key To The Ensign Wasp (Hymenoptera: Evaniidae) Genera of the World, with Descriptions of Three New Genera. *Proceedings of the Entomological Society of Washington*, *105*(4), 859–875.

Dowton, M., & Austin, A. (2001). Simultaneous analysis of 16S, 28S, CO1 and morhology in teh Hymenoptera: Apocrita - evolutionary transistions among parasitic wasps. *Biological Journal of the Linnean Society*, *74*, 87–111.

Dowton, M., & Austin, A. D. (1994). Molecular phylogeny of the insect order Hymenoptera: apocritan relationships. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(21), 9911–5. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=44927&tool=pmcentrez&rendertype=abstract

Dowton, M., & Austin, A. D. (1999). Models of analysis for molecular datasets for the reconstruction of basal hymenopteran relationships. *Zoologica Scripta*, *28*(1-2), 69–74. doi:10.1046/j.1463-6409.1999.00012.x

Drysdale, R. (2001). Phenotypic data in FlyBase. *Briefings in Bioinformatics*, *2*(1), 68–80. doi:10.1093/bib/2.1.68

Endnote. (2010). Endnote. Retrieved March 18, 2010, from http://endnote.com/

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

Gibson, G. A. P. (1985). Some pro-and mesothoracic structures important for phylogenetic analysis of Hymenoptera, with a review of terms used for the structures. *The Canadian Entomologist*, *118*, 205–240.

Gibson, G. A. P. (1997). Morphology and Terminology. In G. Gibson, J. T. Huber, & J. B. Woolley (Eds.), *Annotated Keys to the Genera of Nearctic Chalcidoidea (Hymenoptera)* (pp. 16–44). National Research Council Canada , NRC Research Press.

Gibson, G. A. P., Read, J. D., & Fairchild, R. (1998). Chalcid wasps (Chalcidoidea): illustrated glossary of positional and morphological terms. Retrieved November 1, 2011, from http://www.canacoll.org/Hym/Staff/Gibson/apss/chglintr.htm

Harris, N., Day-Richter, J., Mungall, C., Abdulla, A., & Deegan, J. (n.d.). OBO-EDIT. Retrieved November 1, 2012, from http://oboedit.org/

Harris, R. (1979). A Glossary of Surface Sculpture. *Occasional Papers in Entomology*, *28*, 31.

Haszprunar, G. (1992). The types of homology and their significance for evolutionary biology and phylogenetics. *Journal of Evolutionary Biology*, *5*(1), 13–24. doi:10.1046/j.1420-9101.1992.5010013.x

International Society of Hymenopterists (2012). International Society of Hymenopterists. Retrieved October 1, 2012, from http://www.hymenopterists.org

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

Jardine, N. (1969). The observational and theoretical components of homology: a study based on the morphology of the dermal skull-roofs of rhipidistian fishes. *Biological Journal of the Linnean Society*, *1*(4), 327–361. doi:10.1111/j.1095-8312.1969.tb00125.x

Jensen, L. J., & Bork, P. (2010). Ontologies in quantitative biology: a basis for comparison, integration, and discovery. *PLoS biology*, *8*(5), e1000374. doi:10.1371/journal.pbio.1000374

Johnson, N. F., & Musetti, L. (2011). Redescription and revision of the Neotropical genus Pseudoheptascelio Szabó (Hymenoptera, Platygastridae, Scelioninae), parasitoids of eggs of short-horned grasshoppers (Orthoptera, Acrididae). *ZooKeys*, (136), 93–112. doi:10.3897/zookeys.136.1580

Jurasinski, G., & Retzer, V. (2012). A Collection of functions for similarity analysis of vegetation data. Retrieved from http://cran.r-project.org/package=simba

Karlsson, D., & Ronquist, F. (2012). Skeletal morphology of Opius dissitus and Biosteres carbonarius (Hymenoptera: Braconidae), with a discussion of terminology. *PLoS ONE*, *7*(4), e32573. doi:10.1371/journal.pone.0032573

Kaufman, L., & J., R. P. (1990). *Finding Groups in Data*. (L. Kaufman & P. J. Rousseeuw, Eds.). Hoboken, NJ, USA: John Wiley & Sons, Inc. doi:10.1002/9780470316801

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

Königsmann, E. (1977a). Das phylogenetische System der Hymenoptera Teil 1: Einführung, Grundplanmerkmale, Schwestergruppe und Fossilfunde. *Deutsche Entomologische Zeitschrift*, (23), 253–279.

Königsmann, E. (1977b). Das phylogenetische System der Hymenoptera Teil 2: Symphyta. *Deutsche Entomologische Zeitschrift*, (24), 1–40.

Königsmann, E. (1978a). Das Phylogenetische System der Hymenoptera Teil 3: Terebrantes (Unterordnung Apocrita). *Deutsche Entomologische Zeitschrift*, *a*(25), 1–55.

Königsmann, E. (1978b). Das Phylogenetische System der Hymenoptera Teil 4: Aculeata (Unterordnung Apocrita). *Deutsche Entomologische Zeitschrift*, *a*(25), 365–435.

Lankester, R. E. (1870). On the use of the term homology in modern zoology, and the distinction between homogenetic and homoplastic agreements. *Annals and magazine of natural history*, *6*(4).

Leary, P. R., Remsen, D. P., Norton, C. N., Patterson, D. J., & Sarkar, I. N. (2007). uBioRSS: tracking taxonomic literature using RSS. *Bioinformatics (Oxford, England)*, *23*(11), 1434–6. doi:10.1093/bioinformatics/btm109

Legendre, P., & Legendre, L. (1998). *Numerical Ecology* (2nd ed., p. 839). Amsterdam: Elsevier.

Mabee, P. M., Ashburner, M., Cronk, Q., Gkoutos, G. V, Haendel, M., Segerdell, E., Mungall, C., et al. (2007). Phenotype ontologies: the bridge between genomics and evolution. *Trends in ecology & evolution*, *22*(7), 345–50. doi:10.1016/j.tree.2007.03.013

Malpas, J., & Zalta, E. N. (2012). Mereology. *The Stanford Encyclopedia of Philosophy*. Retrieved November 23, 2012, from http://plato.stanford.edu/entries/mereology/

Mason, W., & Huber. (1993). Order Hymenoptera. *Hymenoptera of the world: An identification guide to families*. (pp. 4–12).

Matsuda, R. Y. U. I. C. H. I. (1957). Morphology of the Head of a sawfly, Macrophya pluricincta. *Society*, *30*(3).

Melo, G. A. R. (1997). Silk glands in adult sphecid wasps (Hymenoptera, Sphecidae, Pemphredoninae). *Journal of Hymenoptera Research*, *6*, 1–9.

Merrill, G. H. (2006). Ontological realism : methodology or misdirection ? *Applied Ontology*, *3*, 1–3. doi:10.3233/AO-2010-0076

Michels, J., & Gorb, S. N. (2012). Detailed three-dimensional visualization of resilin in the exoskeleton of arthropods using confocal laser scanning microscopy. *Journal of Microscopy*, *245*(1), 1–16. Retrieved from http://doi.wiley.com/10.1111/j.1365-2818.2011.03523.x

Michener, C. D. (2000). *The bees of the world* (p. 913). Baltimore, Maryland: Johns Hopkins University Press.

Mikó, I. (2009). Curator. *Hymenoptera Anatomy Ontology*.

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

Mikó, I., Masner, L. L., & Deans, A. R. (2010). World revision of Xenomerus Walker (Hymenoptera: Platygastroidea, Platygastridae). *ZooTaxa*, *2708*, 1–73.

Mikó, I., Vilhelmsen, L., Johnson, N., Masner, L., & Pénzes, Z. (2007). Skeletomusculature of Scelionidae (Hymenoptera: Platygastroidea): head and mesosoma. *Zookeys*, *1571*, 1–78.

Mikó, I., Yoder, M. J., Bertone, M. A., & Deans, A. R. (2009). The Hymenoptera Anatomy Ontology Curation Team.

Mikó, I., Yoder, M. J., Seltmann, K. C., Bertone, M. A., & Deans, A. R. (2011). Functional morphology in descriptive taxonomy: leave the muscle on! *Entomological Society of America SysEB Section Symposium: Illuminating the Phenome: the Future of Morphology in Entomology*. Reno, NV.

Mullins, P., Kawada, R., Balhoff, J., & Deans, A. (2012). A revision of Evaniscus (Hymenoptera, Evaniidae) using ontology-based semantic phenotype annotation. *ZooKeys*, *223*, 1–38. doi:10.3897/zookeys.223.3572

Mungall, C. J., Torniai, C., Gkoutos, G. V, Lewis, S. E., & Haendel, M. A. (2012). Uberon, an integrative multi-species anatomy ontology. *Genome biology*, *13*(1), R5. doi:10.1186/gb-2012-13-1-r5

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

Nichols, S. W. (Ed.). (1989). *Torre-Bueno Glossary of Entomology*. New York: New York Entomological Society and the American Museum of Natural History.

Nixon, K. C., & Carpenter, J. M. (2012). On homology. *Cladistics*, *28*(2), 160–169. doi:10.1111/j.1096-0031.2011.00371.x

Oeser, R. (1961). Vergleichend-morphologische Untersuchungen über den Ovipositor der Hymenopteren. *Mitteilungen aus dem Zoologischen Museum in Berlin*, *37*, 3–119.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., O'Hara, R. B., Simpson, G. L., Solymos, P., et al. (2010). vegan: Community Ecology Package. Retrieved from http://cran.r-project.org/package=vegan

Ontology, R. (2012). Relationship Ontology. Retrieved from http://code.google.com/p/obo-relations

Owens, R. (1843). *Lectures on the Comparative Anatomy and Physiology of the Invertebrate Animals*. London: Longman, Brown, Green and Longmans. Retrieved from http://archive.org/details/lecturesoncompar02owen

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, *20*(2), 289–290. doi:10.1093/bioinformatics/btg412

Patterson, C. (1982). Morphological characters and homology. In K. A. Joysey & A. E. Friday (Eds.), *Problems of Phylogenetic Reconstruction* (pp. 21–74). London & New York: Academic Press.

Plazi. (2012). Plazi: Search and Retrieval Server. Retrieved December 10, 2012, from http://plazi.org:8080/dspace/community-list

Popovici, O., Seltmann, K. C., & Mikó, I. (in prep). The maxillo-labial complex in Sparasion (Platygastroidea: Platygastridae).

Prud'homme, B., Minervino, C., Hocine, M., Cande, J. D., Aouane, A., Dufour, H. D., Kassner, V. A., et al. (2011). Body plan innovation in treehoppers through the evolution of an extra wing-like appendage. *Nature*, *473*(7345), 83–6. doi:10.1038/nature09977

R Development Core Team. (2010). *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org

Ramírez, M. J., Coddington, J. a, Maddison, W. P., Midford, P. E., Prendini, L., Miller, J., Griswold, C. E., et al. (2007). Linking of digital images to phylogenetic data matrices using a morphological ontology. *Systematic biology*, *56*(2), 283–94. doi:10.1080/10635150701313848

Rasnitsyn, A. P. (1969). Origin and evolution fo the lower Hymenoptera. *Trudy Paleontologicheskii Instituta*, (123), 1–196.

Rasnitsyn, A. P. (1980). Origin and evolution of Hymenoptera. *Transactions of the Paleontological Institute of the Academy of Sciences of the USSR*, (174), 1–192.

Rasnitsyn, A. P. (1988). An outline of the evolution of hymenopterous insects (order Vespida). *Oriental Insects*, *22*(22), 115–145.

Remane, A. (1952). *Die Grundlagen des natürlichen Systems, der vergleichenden Anatomie und der Phylogenetik*. Leipzig: Geest und Portig.

Richards, O. W. (1997). *Hymenoptera. Introduction and key to families. Handbooks for the Identification of British Insects* (2nd ed., pp. 1–100).

Rieppel, O. (1980). Homology, a deductive concept? *Zeitschrift für Zoologische Systematik und Evolutionsforschung*, *18*, 315–319.

Rieppel, O. (1988). *Fundamentals of Comparative Biology*. Basel: Birkhäuser.

Rokas, A., Williams, B. L., King, N., & Carroll, S. B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *October*, *425*(October).

Ronquist, F. (1999). Phylogeny of the Hymenoptera (Insecta): The state of the art. *Zoologica Scripta*, *28*(1-2), 3–11. doi:10.1046/j.1463-6409.1999.00019.x

Ronquist, F., & Nordlander, G. (1989). Skeletal morphology of an archaic cynipoid, Ibalia rufipes (Hymenoptera: Ibaliidae). *Entomologica Scandinavica*, *33*, 1–60.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, *4*(4), 406–25.

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

Savard, J., Tautz, D., Richards, S., Weinstock, G. M., Gibbs, R. A., Werren, J. H., Tettelin, H., et al. (2006). Phylogenomic analysis reveals bees and wasps (Hymenoptera) at the base of the radiation of Holometabolous insects. *Genome research*, *16*(11), 1334–8. doi:10.1101/gr.5204306

Scholar, G. (2010). Google Scholar. Retrieved March 18, 2010, from http://scholar.google.com/

Schulmeister, S. (2001). Functional morphology of the male genitalia and copulation in lower Hymenoptera, with special emphasis on the Tenthredinoidea s. str. (Insecta, Hymenoptera, "Symphyta"). *Acta Zoologica*, *82*, 331–349.

Schulmeister, S. (2003a). Simultaneous analysis of basal Hymenoptera (Insecta): introducing robust-choice sensitivity analysis. *Biological Journal of the Linnean Society*, *79*(2), 245–275. doi:10.1046/j.1095-8312.2003.00233.x

Schulmeister, S. (2003b). Review of morphological evidence on the phylogeny of basal Hymenoptera (Insecta), with a discussion of the ordering of characters. *Biological Journal of the Linnean Society*, *79*(2), 209–243. doi:10.1046/j.1095-8312.2003.00232.x

Scotland, R. W. (2000). Taxic Homology and Three-Taxon Statement Analysis. *Systematic Biology*, *49*(3), 480–500. doi:10.1080/10635159950127358

Seltmann, K.C. (2010). Curator, Hymenoptera Anatomy Ontology. *Hymenoptera Anatomy Ontology*.

Seltmann, K.C., Pénzes, Z., Yoder, M. J., Bertone, M. A., & Deans, A. R. (in press). Utilizing Descriptive Statements from the Biodiversity Heritage Library to Expand the Hymenoptera Anatomy Ontology. *PLoS ONE*.

Seltmann, K.C., Yoder, M., Mikó, 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), 101–12. doi:10.1016/j.ympev.2010.07.006

Sharkey, M. J., Yu, D. S., Noort, S., Seltmann, K.C., 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.

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., (2007). Phylogeny and Classification of Hymenoptera. *Zootaxa*, *548*, 521 – 548.

Sharkey, M. J., Carpenter, J. M., Vilhelmsen, L., Heraty, J., Liljeblad, 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., & Roy, A. (2002). Phylogeny of the Hymenoptera: a reanalysis of the Ronquist et al. (1999) reanalysis, emphasizing wing venation and apocritan relationships. *Zoologica Scripta*, *31*(1), 57–66. doi:10.1046/j.0300-3256.2001.00081.x

Sharkey, M. J., & Wharton, R. (1997). *Manual of the New World genera of the family Braconidae (Hymenoptera)*. (R. Wharton, M. J. Sharkey, & P. Marsh, Eds.). International Society of Hymenoptera Special Publication 1.

Smith, B. (2005). The logic of biological classification and the foundations of biomedical ontology. *Invited Papers from the 10th International Conference in Logic Methodology and Philosophy of Science (Oviedo, Spain, 2003)* (pp. 505–520). London: King's College Publications.

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

Snodgrass, R. E. (1910). *Anatomy of a Honey Bee*. *Library* (p. 236). Cornell University Library.

Snodgrass, R. E. (1935). *Principles of Insect Morhology* (p. 667). New York & London: McGraw-Hill Book Co., Inc.

Snodgrass, R. E. (1942). Skeleto-muscular mechanisms of the honey bee. *Smithsonian Miscellaneous Collections*, *103*(2), 1–120.

Snodgrass, R. E. (1956). *Anatomy of the Honey Bee*. *Smithsonian* (1st ed., Vol. 103, pp. 1–334). Ithaca, New York: Comstock Publishing Associates.

Sokal, R. R., & C. D. Michener. (1958). A statistical method for evaluating systematic relationships. *Univ. Kans. Sci. Bull.*, *38*, 1409–1438.

SourceForge. (*2012*). SourceForge. Retrieved from sourceforge.net

Talamas, E. J., Masner, L., & Johnson, N. F. (2011). Revision of the Paridris nephtaspecies group (Hymenoptera, Platygastroidea, Platygastridae). *ZooKeys*, (133), 49–94. doi:10.3897/zookeys.133.1613

Topalis, P., Tzavlaki, C., Vestaki, K., Dialynas, E., Sonenshine, D. E., Butler, R., Bruggner, R. V, et al. (2008). Anatomical ontologies of mosquitoes and ticks, and their web browsers in VectorBase. *Insect molecular biology*, *17*(1), 87–9. doi:10.1111/j.1365-2583.2008.00781.x

Vilhelmsen, L. (2007). The phylogeny of lower Hymenoptera (Insecta), with a summary of the early evolutionary history of the order. *Journal of Zoological Systematics and Evolutionary Research*, *35*(2), 49–70. doi:10.1111/j.1439-0469.1997.tb00404.x

Vilhelmsen, L. (1996). The preoral cavity of lower Hymenoptera (Insecta): comparative morphology and phylogenetic significance. *Zoologica Scripta*, *321*(2), 129–170. doi:10.1111/j.1463-6409.1996.tb00156.x

Vilhelmsen, L. (1999). The occipital region in the basal Hymenoptera (Insecta): a reappraisal. *Zoologica Scripta*, *28*(1-2), 75–85. doi:10.1046/j.1463-6409.1999.00008.x

Vilhelmsen, L. (2001). Phylogeny and classification of the extant basal lineages of the Hymenoptera (Insecta). *Zoological Journal of the Linnean Society*, *131*(4), 393–442. doi:10.1006/zjls.2000.0255

Vilhelmsen, L. (2010). Curator. *Hymenoptera Anatomy Ontology*.

Vilhelmsen, L., & Mikó, I. (2010). Curator. *Hymenoptera Anatomy Ontology*.

Vilhelmsen, L., Mikó, I., & Krogmann, L. (2010). Beyond the wasp-waist: structural diversity and phylogenetic significance of the mesosoma in apocritan wasps (Insecta: Hymenoptera). *Zoological Journal of the Linnean Society*, *159*(1), 22–194. doi:10.1111/j.1096-3642.2009.00576.x

Von Ahn, L., Maurer, B., McMillen, C., Abraham, D., & Blum, M. (2008). reCAPTCHA: human-based character recognition via Web security measures. *Science (New York, N.Y.)*, *321*(5895), 1465–8. doi:10.1126/science.1160379

Wagner, G. P. (1989). The Biological Homology Concept. *Annual Review of Ecology and Systematics*, *20*(1), 51–69. doi:10.1146/annurev.es.20.110189.000411

Wharton, R., Ward, L., & Mikó, 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

Wiegmann, B. M., Trautwein, M. D., Kim, J. W., Cassel, B. K., Bertone, M. A., Winterton, S. L., & Yeates, D. K. (2009). Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. *BMC biology*, *7*, 34. doi:10.1186/1741-7007-7-34

Wiley, E. O., & Lieberman, B. S. (2011). *Phylogenetics: Theory and Practice of Phylogenetic Systematics*. Hoboken: Wiley-Blackwell.

Yoder, M. J. (2009). *Curator, Hymenoptera Anatomy Ontology*.
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
Youssef, N. N. (1971). Topography of the Cephalic Musculature and Nervous System of the Honey Bee Apis mellifera Linaeus. *Smithsonian Contributions to Zoology*, (99).
Zotero. (2010). Zotero. Retrieved March 18, 2010, from http://www.zotero.org

Summary

The work in this dissertation was focused on the further development and application of the Hymenoptera Anatomy Ontology (HAO), a resource based on a foundation of explicitly defined anatomical concepts and a straightforward mechanism for referencing these concepts (URIs).

In Hymenoptera taxonomy we are in an interesting position, with an estimated million more species to describe. We have the potential to modify our approach, by incorporating semantic, repeatable, and machine understandable techniques into our descriptions, thus making them highly available for biological research in general. The first step toward this goal is the creation of a structured, controlled vocabulary of Hymenoptera terminology, the HAO. Simplified, an ontology is a set of concepts (definitions of morphological structures that follow specific rules) used to model a formalized domain (Hymenoptera anatomy), and the logical relationships between concepts. The goal of ontology creation is to enable computer-based reasoning on morphological concepts (linked to terms, or words in publications) that are defined based on structural similarity.

The primary ontology development software for the HAO is mx, a Ruby on Rails, MySQL-based open source content management system for descriptive taxonomy. The software for HAO development is outlined in Yoder et al. (2010), Seltmann et al. (2012), and Seltmann et al. (in press).

Anatomy ontology creation and implementation requires significant expertise in the domain it is describing. Terms and concepts were illustrated as part of this dissertation, 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 Bertone et al. (2012), Seltmann et al. (2012) and Miko et al. (2012).

In order to facilitate obscure term discovery an active learning, dictionary-based, natural language recognition tool, known as the 'Proofer' was implemented for examining text. 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. The sampled articles were then analyzed in Part 2 for occurrence of terms using a variety of clustering methods and 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).

I developed an easily comprehensible methodology for linking terminology in descriptive texts through Uniform Resource Identifier (URI) tables, which could be included in a manuscript. The tables are created using the 'Analyzer' mx-based tool, linking the specific word in the manuscript with a single, defined concept in the HAO. Once published, the links resolve to the HAO ontology through the Hymenoptera Portal Webpages. The general discussion of URI for the Hymenoptera community is derived from Seltmann et al. (2012).

I have 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. Additionally, I have demonstrated novel functionality for constructing anatomy ontologies, regardless of domain.

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. The main results of my work are summarized in the following points:

1. 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.

2. I have performed 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. 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. 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.

3. I have achieved a significant augmentation of the development of the Hymenoptera Anatomy Ontology based on the 1189 new terms that were collected through the computer assisted analysis and that were subsequently added to the mx database.

4. I examined the applicability of PATO (Phenotypic Anatomy Ontology) 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.

5. I promoted 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 systematist community. Since publication of the URI table concept was initiated, seven morphology publications have adopted HAO terminology and the URI/Analyzer methodology.

The work in this dissertation 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 Seltmann 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.

Összefoglalás

A jelen disszertáció alapjául szolgáló kutatás a Hymenoptera Anatómia Ontológia (HAO) fejlesztésével valamint lehetséges alkalmazásával kapcsolatos. A HAO explicit módon meghatározott anatómiai fogalmak kapcsolatrendszere, amely egyben egyszerű mechanizmust kínál a fogalmak hivatkozására (URI).

Szemantikus, és a kutatások megismételhetőségét elősegítő morfológiai leírások készítése különösen időszerű a hártyásszárnyúak taxonómiai kutatásában, hiszen még legalább egymillió darázsfaj vár leírásra. Az új fajok leírásának olyan metodikáját teremti meg, ami számítógéppel is értelmezhető. Ez kétség kívül javítaná a fajleírásban szereplő morfológiai adatok hozzáférhetőségét más tudományterületek számára is. Az első lépés ennek a célnak az eléréséhez a HAO létrehozása lehet. Leegyszerűsítve, a HAO egy olyan fogalomtár, amelyben a domén specifikus fogalmak (jelen esetben a hártyaszárnyúak anatómiájával kapcsolatosak) meghatározott szabályok, és az egymással kialakított kapcsolataik alapján vannak definiálva. Az anatómia ontológiákat általában abból a célból hozzák létre, hogy számítógépekkel is elérhetővé tegyék a morfológiai leírásokat úgy, hogy a leírások terminológiáját összekötik az ontológiákban szereplő, strukturális hasonlóságok alapján meghatározott morfológiai koncepciókkal.

A HAO fejlesztése elsődlegesen a módosított "mx" szoftverrel történik, amely Ruby on Rails és MySQL szoftvereken alapuló nyílt forráskódú rendszer és a leíró taxonómia számára készült. A szoftver leírását Yoder és mtsai. 2012, illetve Seltmann és mtsai. 2012 valamint közlésre elfogadott közlemények tartalmazzák.

Anatómiai ontológiák elkészítéséhez alapvető követelmény az adott domén (jelen esetben a hártyaszárnyú anatómia) alapos ismerete. Disszertációm részeként áttekintettem a hártyásszárnyúak szájszerv anatómiáját és definiáltam, valamint illusztráltam (konfokális lézer szkenning mikroszkópos, valamint fénymikroszkópos technikák alkalmazásával) az anatómiai régióval kapcsolatos morfológiai fogalmakat. Morfológusok, domén specialisták részvétele a fejlesztésben kulcsfontosságú, lásd például Bertone és mtsi. 2012, Seltmann és mtsi. 2012, valamint Mikó és mtsi. 2012 közleményeket.

Új, az ontológiában még nem szereplő, „rejtett" terminusok feltárása céljából készült az aktív tanuláson alapuló „Proofer", amely egy szótáralapú szövegfelismerő eszköz. Első lépésben, a HAO szókészletének növelése céljából, analizáltuk a Journal of Hymenoptera Research folyóiratban megjelent közleményekben szereplő leírásokat. A második lépésben többváltozós módszerek segítségével elemeztük a közleményeket a

bennük szereplő fogalmak, mint változók alapján. Az eredményeinket összehasonlítottuk a hártyásszárnyúak jelenleg elfogadott rendszerével. A „Proofer" fejlesztésével és alkalmazásával kapcsolatban Seltmann és mtsi. közlésre elfogadott kézirat foglalkozik részletesen.

Munkám során kidolgoztam egy metodológiát, aminek a segítségével morfológiai leírások összekapcsolhatóak az ontológiában található morfológiai koncepciókkal, egy adott cikk ún. „URI" (Unique Resource Identifier) táblázatán keresztül. A HAO koncepcióinak URI-jait tartalmazó táblázatot az mx-ben található „Analyzer" eszköz generálja. A publikációkban szereplő URI-k a „Hymenoptera Portal" internetes oldalon keresztül kapcsolódnak a HAO-hoz. URI táblázatok készítésével valamint azok használatával kapcsolatba Seltmann és mtsi. 2012 szolgáltatnak bővebb információt.

Munkám során szemléltettem, hogy a Hymenoptera rend taxonjainak morfológiai leírásaiban használatos terminusokat, és ezek által hivatkozott fogalmakat magában foglaló szótár, szinte nélkülözhetetlenen a hártyásszárnyúak rendszertanával kapcsolatos kutatásokban. Ennek legfőbb oka a különböző darázs családsorozatok valamint családok leírásában használatos morfológiai terminológia specifikussága.
Új ízeltlábú anatómiai ontológiák létrehozatalának elősegítése nagy jelentőséggel bír az ezzel foglalkozó közösség számára, valamint az egyre inkább elterjedtebb szemantikus publikációkon keresztül befolyásolhatja a tudományterületek fejlődését is. Munkám főbb eredményeit a következőkben foglalom össze:

1. Létrehoztam egy „szövegbányászati" eszközt („Proofer"), morfológiai terminusok faj leírásokból történő kinyerésének elősegítése céljából, amely jelentősen csökkenti a felhasználó által áttekintendő fogalmak számát. A "Proofer" hatékonyságának kvantitatív jellemzését 25 véletlenszerűen kiválasztott tudományos cikk elemzésével végeztem, az eszköz által kivont, és az eredeti leírásokban levő fogalmak számának összehasonlításával. A "Proofer" használatával 1/3-ra csökkent a felhasználó által ellenőrizendő szavak száma, az összetett morfológiai terminusok számát pedig 80%-kal csökkentette.
2. A "Proofer" felhasználásával elvégeztem 353 tudományos közlemény elemzését, ami alapján megállapítottam, hogy a taxonómusok taxon-doménekre specifikus terminológiát használnak, ami a családsorozatok illetve a családok tekintetében a legszembetűnőbb. A tanulmányozott közlemények közül 180 tartalmazott tudományra új taxon leírásokat. A legrövidebb fát az olyan családsorozatokon alapuló hierarchikus klaszter analízisekkel (Sorensen távolsággal és csoportátlag módszerrel) kaptuk,

amelyekbe csak kettő, vagy több terminálisban szerepelő karaktereket vontuk be. A legrövidebb fán 63 csoport különíthető el a fa ágainak taxonómiai besoroláson alapuló összevonása (monofiletikus csoportok „nyesése") után, legszembetűnőbbek az Ichneumonoidea, Chalcidoidea, Symphyta és Aculeate fajok leírásaival kapcsolatos közlemények csoportosulásai.

3. A "Prooferrel" végrehajtott szövegbányászat során a tudományos közleményekből kinyert 1189 terminus bekerült a HAO rendszerébe, jelentősen növelve annak fogalom és szókészletet.

4. Munkám során vizsgáltam a Fenotípus Anatómiai Ontológia, hártyásszárnyú taxonok leírásában való lehetséges használhatóságát (HAO és PATO egymással való átfedése). Ennek eredményeképpen kezdeményeztem több, a PATO-ból hiányzó, de a darázsfajok leírásában gyakran használt, terminus PATO-ba való felvételét.

5. Részt vettem egy, a taxonómiai leírásokat a HAO-ban tárolt fogalmakkal, azok URI-jain keresztül, összekötő metodológia kidolgozásában, valamint az új ontológiai filozófián alapuló koncepcióknak, a hártyásszárnyúakkal foglalkozó kutatók közötti népszerűsítésében. Megalkotása óta hét tudományos értekezés alkalmazta az URI-táblázatokon alapuló koncepciót és ezen keresztül a HAO nevezéktanát.

A jelen disszertáció fő témája egy a hártyásszárnyúak morfológiai leírásaiban található terminusokat tartalmazó rendszer, a HAO fejlesztése és alkalmazása a benne tárolt ontológiai koncepciók URI-jainak felhasználásával. Mint ahogy azt a megjelent publikációkban demonstráltuk, egy egységes terminológián alapuló eszköz rendkívül fontos a hártyásszárnyúak morfológiai leírásainak megértéséhez, hiszen ezeknek a terminológiája taxon (család és családsorozat) specifikus. A HAO alkalmazása sokrétű. Mindamellett hogy növeli a hártyásszárnyúak taxonómiájával kapcsolatos kutatások ismételhetőségét, és segíti a leíró taxonómusok munkáját, explicite definiált morfológiai koncepciói segítségével lehetővé teszi a leíró taxonómia eredményeinek szélesebb körű felhasználását. Hasonlóan más biológiával kapcsolatos ontológiákhoz, a HAO gyorsan fejlődik, bővül a benne található fogalmak száma, így használhatósága is.

A szemantikus publikációk számának növekedésével egyre nagyobb a jelentősége az ízeltlábú anatómiai ontológiáknak. Ezek nem csak az ontológiákkal foglalkozó szakemberek, hanem más biológiai diszciplínák képviselői számára is fontosak. Morfológiai struktúrák pontos és korrekt interpretációja pozitív hatással lehet minden olyan biológiai tudományterületre, ami valamilyen szinten kapcsolatban van a hártyásszárnyúak anatómiájával. A felsorolás teljessége nélkül megemlítenek néhányat

ezek közül: biodiverzitás, gazda-parazitoid biológia, rovargyűjtemények digitalizálása, genomika, ökológia, evolúciós fejlődésbiológia, invazív fajokkal és mezőgazdasági ökoszisztémákkal kapcsolatos vizsgálatok valamint a biológiai védekezés.

GLOSSARY

Word	Definition
Analyzer Tool	The mx software based tool for associating blocks of text with concepts in the HAO. The Analyzer facilitates URI table development.
Class	A synonym of *concept*.
Concept	The idea, represented by a definition, of our understanding of an anatomical feature. For example, “The tagma that is located anterior to the thorax” defines the concept commonly referenced by the *term* “head”.
Concept Drift	The application of a term to a modified set of structures over time and by different authors (ex. Paramere).
Figures	Figures are typically images, and are used to help illustrate concepts.
Genus-differentia	A type of definition structured so as to first describe a more inclusive class of *concepts* (genus) and then the characteristics differentiating (differentia) it from other children of that concept. Definitions in this format typically follow the pattern 'The *x* that is *y*.'
Homonym	A term that is used for two or more *concepts*.
Instance	A real-life exemplar of the *concept*. For example, the actual physical head of a specific specimen is an instance of the HAO *class* identified by the *URI* http://purl.obolibrary.org/obo/HAO_0000397.
Label	Similar to *term,* as it is a word representation of a concept. However, label is specifically the word associated with an ontology concept without further meaning.
mx	A Ruby on Rails, MySQL-based open source content management system for descriptive taxonomy.

Ontology	A set of *concepts* and *relationships* (properties) pertaining to a particular domain of knowledge (e.g. hymenopteran anatomy).
Obsolete Class	An ontology class that has been merged with another class or modified from its original form. Obsolete Classes are maintained and remain resolvable.
Proofer Tool	The mx software based tool for extracting new terms from the literature for incorporation in the HAO.
Tag	Tags are semi-informal ways to add short notes to the things in the ontology. They are stored in the mx database.
Relationship	A property linking two *concepts*, however when the ontology is applied to real-life examples, the relationships apply to instances of those concepts. For example, "the eye is not part of the head" (as recorded in the ontology), but rather "my eye is part of my head".
Semantic Phenotype	Structured annotations that represent observations of the phenome (see Deans et al. 2012; Mullins et al, 2012) and follow well-defined rules for annotation.
Sensu	The combination of a *concept, term,* and an associated reference. For example, Snodgrass (1941) used the *term* 'phallobase' in combination with the *concept* http://purl.obolibrary.org/obo/HAO_0000713 ("The anatomical cluster that is composed of the cupulae, gonostipites and volsellae."). For discussion of the term see Yoder et al. (2010).
Structural Equivalence	Topographical sameness of structures as implemented in anatomy ontology construction.
Synonyms	Two or more *terms* used to reference the same *concept*. For example, both *terms* 'phallobase' and 'paramere' have been used to reference the *concept* http://purl.obolibrary.org/obo/HAO_0000713.
Term	A name representing a *concept* to an author or person. Terms contain highly involved meaning outside of concepts defined in the HAO. Often

	terms contain evolutionary homology hypothesis, or more specific definitions than what is captured in the HAO. (compare with *label*)
URI	Short for Uniform Resource Identifier, a URI is a unique identifier for a *concept,* which is resolvable in a Web browser.