

1. Title. Transforming the Universal Chalcidoidea Database into a Community-Supported Resource for Evolutionary Biology.

2. Short Title. “Transforming the UCD”.

3. Project Leader:

James B. Woolley
Department of Entomology
Texas A&M University
jimwoolley@tamu.edu
979-845-9349

Project Co-Leader:

Matt Yoder
Department of Entomology
North Carolina State University
diapriid@gmail.com

4. Project Summary

Evolutionary and systematics research on species in the superfamily Chalcidoidea (19 families, currently about 31,000 species) has become critically dependent on the Universal Chalcidoidea Database. The UCD is a catalog of all scientific names and taxonomic actions that have occurred in the family and a comprehensive database of biological information and host records from the scientific literature. Every piece of information in UCD is directly linked to a scientific publication. The UCD was the sole achievement of John Noyes, Natural History Museum, London. John’s retirement has created an urgent need to move UCD to a new platform that will provide a stable foundation for future development. We propose to form a Working Group composed of chalcidoid taxonomists, including taxonomists in training, and database and IT experts, to move UCD web-based platform. One person will coordinate the activities of a group of researchers, each of which will take responsibility for updating information for particular families of wasps. UCD will continue to be served over the Internet and it will become fully incorporated into the semantic net. The project will facilitate and enable new research in the evolutionary biology of parasitoids and it will have profound broader impacts, given the widespread use of chalcidoid species as natural enemies of pest arthropods. The Working Group will address an important question in the evolution of science itself: how to convert an essential resource compiled by a single person into a distributed resource, supported by an entire research community.

5. Introduction and Goals

The Chalcidoidea are a mega-diverse group of insect parasitoids, with as many as 500,000 species estimated to exist. They are numerous and ubiquitous components of all terrestrial and even aquatic ecosystems. Hosts are known in virtually all orders of insects, as well as arachnids and other arthropods, and many chalcidoids are phytophagous (phytophagous chalcidoids are almost invariably gall-forming or at least, endophytic phytophages). As one of the largest and most diverse radiations of insect parasitoids, chalcidoids provide ideal systems for studying evolutionary processes such as the evolution of host relationships in parasitoids, transitions from entomophagic to plant-parasitic biologies, transitions to and from endoparasitic and ectoparasitic development, and the evolution of hyper-parasitoids (parasitoids of parasitoids). All of these phenomena are common and well represented in Chalcidoidea, in many different lineages resulting from many different evolutionary histories. Many evolutionary questions are currently of interest, of which the following are only a sample. What is the role of host-switching as a component in or even as a driver of speciation in Chalcidoids? Do phylogenetically related parasitoids attack phylogenetically related hosts? What is the role of host specialization in the evolution of chalcidoids? Why are some lineages obligate parasitoids of certain host groups (e.g. all known Eucharitidae are parasitoids of ants), while other lineages have hosts in multiple orders of insects and other arthropods? Why do some lineages consist entirely of egg parasitoids (Trichogrammatidae), plant parasitoids (Tanaostigmatidae), hyperparasitoids (many genera in many different families), whereas other lineages of comparable size have members with all of these biologies and more? Although chalcidoids have been widely and successfully used as biological control agents, the potential impact of natural enemies on non-target species is now a central concern. Therefore, understanding the evolution of host-specificity in chalcidoids has very direct broader impacts to agriculture and forestry, as well as human and veterinary medicine. Many evolutionary questions not associated with host relationships are also of interest, for example, why are some lineages characterized by exuberant and extravagant displays of morphological diversity (Encyrtidae), whereas others seem to be characterized by morphological uniformity (Pteromalinae)? What is the role of bacterial symbionts such as *Wolbachia* and *Cardinium*, which appear to be ubiquitous in some lineages and less common in others?

As phylogenies of parasitoid lineages become available, they provide a powerful context for comparative studies of these evolutionary questions, and more. Phylogenies provide the essential framework for designing and interpreting experimental studies, such as the choice of species for experiments and controls. However, design of contemporary research programs to address these questions, whether experimental or comparative, are also dependent on the uniform application of taxonomic names, the ability to track particularly taxonomic names through nomenclatural history, and the widespread availability of accurate host records. Fortunately, we now have a comprehensive database that provides this information for chalcidoid species, but in order for it to continue to serve in this role and to continue to evolve as new information develops, it must itself undergo a metamorphosis into a new form.

The Universal Chalcidoidea Database (UCD) is a resource compiled during the career of a single individual, John Noyes (Natural History Museum, London), and hosted by the NHM

(<http://www.nhm.ac.uk/research-curation/research/projects/chalcidoids/>). The UCD currently contains original citations for all 31,000 taxonomic names in Chalcidoidea, and nearly 50,000 different combinations and spellings are recorded. Also included are over 120,000 records of insect and plant hosts, and more than 140,000 distribution records. The database lists over 40,000 literature references relevant to Chalcidoidea and this can be searched using 120 predefined keywords. The taxonomic literature referenced in UCD is essentially complete for the entire superfamily back to Jan. 1, 1758.

The database is a monumental achievement by John Noyes, and it has become an absolutely essential resource and an indispensable tool for research and teaching in systematics, phylogenetics, parasitoid evolution and biology, biological diversity and faunistics, and integrated pest management and biological control. The UCD is at the heart of new emerging bioinformatics initiatives in Chalcidoidea associated with numerous research activities worldwide. However, due to the retirement of John Noyes on April 2, 2009, the UCD is now at a critical transition point. Although Andy Polaszek, NHM, has taken over the management of UCD, it is clear that the time has come to substantially rethink both the platform and structure for the database, and the manner in which information will be updated and maintained.

The UCD was first implemented in 1991 by keypunching the information that had been painstakingly accumulated over the years in a card catalog at NHM into a PARADOX database on an MS/DOS platform. The database became commercially available on CD in August 1998 in collaboration with ETI Amsterdam and Dicky Yu. A second, improved edition was published in 2002 by Dicky Yu using the Taxapad© data management system. The database is still maintained in PARADOX, and updated to the on-line version on an intermittent basis by NHM computing staff.

While UCD is severely non-normalized at present, it is meticulously curated and well positioned to parsing via script-based translation to an RDMS. Presently at the NHM, the data are periodically parsed and re-indexed for display on their web site. However, it is not accessible via web services or other automated APIs.

Our Goals for UCD:

1) The UCD should be moved to a contemporary platform. We propose to incorporate UCD into MX (<http://sourceforge.net/projects/mx-database/>) an open source collaborative content-management framework developed originally as part of NSF PEET awards to Bob Wharton and Jim Woolley and now under the Hymenoptera Anatomy Ontology (<http://hymao.org>). MX uses MySQL and Ruby on Rails to provide the user interface. Mapping of the data model of the UCD to MX is already underway and the two are largely compatible.

MX is currently serving as a central hub for data storage for several large-scale collaborative research projects. Of particular relevance to this proposal are two taxonomic catalogs, *Evanioidea Online* (<http://evanioidea.info/>) and HISL Xyleborini (<http://xyleborini.tamu.edu/>) which have been developed in MX and encapsulate the basic capabilities needed by the UCD. Also relevant to this proposal has been an initiative organized by John Heraty, University of

California at Riverside, as part of the AToL project in Hymenoptera, to bring together over thirty experts on chalcidoid morphology and systematics to design and populate a large data matrix consisting of over 400 morphological characters to be scored eventually for hundreds of taxa of chalcidoids, representing thorough taxon sampling across the superfamily. The scale of this ambitious project has required building new functionality into MX, in order to provide tools for discussion of coding, recoding and reinterpretation of characters as the project has evolved. The project is an example of the continued ability of this chalcid research community to identify common goals and work cooperatively to achieve them.

2) Andy Polaszek will continue as Project Manager of UCD, but responsibility for update and management of content will be distributed among specialists for particular families of Chalcidoidea. This research community has a long history of productive and collegial collaborations (for example, see G.A.P. Gibson, J.T. Huber and J.B. Woolley, 1997, *Annotated Keys to the Genera of Nearctic Chalcidoidea (Hymenoptera)*, NRC Research Press, co-authored by 18 specialists). All data in UCD will be available on publicly accessible web-sites, however, to recognize the contributions of the Natural History Museum in developing and maintaining UCD we will include appropriate NHM branding on web interfaces and recognize the primary contributions of John Noyes. Source code for MX is currently available on SourceForge (see url above, Goal 1) and we will continue this policy.

3) The UCD should be broadly integrated with the semantic web. It is critical that the UCD data are exposed not only to human-accessible interfaces but also to algorithmic queries. This will facilitate the use of UCD data for a broad range of biological and evolutionary questions and further enables unanticipated usages. Potential technologies include LSIDs (or equivalent GUIDs), CDAO, DarwinCore, Nexcel, LinkedData, and RESTful APIs.

4) The new UCD will have an array of tools required to allow a community of specialists to curate the data in an efficient manner. This will include tools to alert other curators of changes or needed input (e.g. by triggered Twitter or Email events), interfaces to query the broader web (e.g. BHL <http://www.biodiversitylibrary.org/>) for additional data, and markup technologies (Tagging) to allow for rich data annotation. User roles, authentication (OpenID), and mechanisms for non-authenticated user input (e.g. authors alerting the UCD of new data to be curated) are potential technologies

Goals of the Working Group:

1) During two meetings at NESCENT, we will bring together the new Project Manager of UCD, Andy Polaszek, the original developer, John Noyes, a group of four prospective curators (domain experts in various groups of Chalcidoidea, including two graduate students), five experts in database structure and the semantic web, and at least one programmer dedicated to the project, in order to move UCD to a contemporary platform.

2) Ensuring that the data in the new system are semantically sound, meet the latest taxonomic standards, and are maximally useful to the systematics and evolutionary biology communities, requires a number of more practical goals:

- a) UCD should incorporate LSID's (or equivalent GUIDs) to become part of the semantic web, and to promote maximal integration with other biological databases including as MorphBank, BHL, Hymenoptera Online, Encyclopedia of Life, Catalogue of Life, Species 2000, ITIS etc.
 - b) Discussions between domain experts will resolve issues associated with the structure of fields and tables in the database, and provide an opportunity for improvements.
 - c) At the first meeting of the Working Group we will outline the goals for the programmers who will there begin to hack on the necessary framework. This work will include mapping all existing data into the new structure, extending the data model where necessary, outlining protected access via OPENid, and outlining the API which will expose the UCD to the web.
- 3) We wish to develop, with the assistance of NESCENT participants, novel approaches to extracting data from the UCD such that they can be quantitatively analyzed to address questions in evolutionary theory and processes. We will have a broad range of available data (UCD, HAO, Chalcid supematrix) with which to experiment on and integrate across.
- 4) In the second meeting of the Working Group, we will include representatives from Hymenoptera Online, EoL, Biodiversity Heritage Library and other large-scale bioinformatics activities to be sure that linkages between UCD and these activities are in place, and functional.

6. Proposed Activities. Two meetings of the Working Group at NESCENT are proposed.

Meeting One.

Day One

Morning sessions (all participants):

Graduate students will act as scribes, recording minutes to the UCD wiki.

- a. Demonstration of current implementation of UCD so that everyone is familiar with the current structure and current capabilities of the system (Noyes and Polaszek).
- b. Demonstration of MX content management system, focusing on aspects of the system relevant to potential porting of UCD to MX (Yoder).
- c. Overview of plan to port UCD to MX (Woolley, Yoder, Polaszek).
- d. Discussion of pitfalls we may face during the workshop and solutions, as well as any suggestions for changes to workshop organization (roundtable discussion).

Afternoon sessions (all participants)

- e. All participants including domain experts and database experts will exchange ideas for improving the structure and function of UCD in its new implementation, and discuss required functionality, including modification to existing data structures in UCD, new data structures, and functions that will be required to provide a platform for distributed curation of data.
- f. Set and modify goals for days 2-4.

Day Two

Morning sessions: (concurrent)

- a. Chalcid experts will resolve issues associated with actual data capture and outline the *preferred* workflow- independent of the application
- b. Programmers and semantic experts will further resolve the data model for the new UCD and the method of import of UCD to MX.

Afternoon sessions (all participants):

- c. Domain experts will report to programmers on morning activities outlining their preferences, and programmers will demonstrate the prototype data-model and any existing functionality that may meet requirements. Discussion between the groups will provide an opportunity for synthesis of ideas and practical realities.
- d. Final afternoon session will set goals for Day Three.

Day Three

Morning sessions (concurrent):

- a. Domain experts will discuss and record use cases that represent particularly problematic challenges to the structure and function of UCD.
- b. Programmers will begin to hack on the prototype and set up required repositories, stubbing libraries, methods and tests, focusing on particular issues that were identified in Day 2.

Afternoon Sessions (all participants):

- c. Each group will report to the other on morning activities and recommendations. Challenging use cases and other challenges to implementation of UCD will be discussed and solutions proposed.
- d. Final afternoon session will set goals for Day Four.

Day 4.

The goal during day four will be to provide a roadmap and clear guidelines to the programmers that will continue development of UCD in the period between working group meetings.

Morning sessions (concurrent):

- a. Domain experts will discuss the features necessary to incorporate into protected-access portals for data updating and curation.
- b. Programmers will outline and stub the authentication and roles for the web portal(s) for protected-access and public access to UCD.

Afternoon sessions (all participants):

- c. Each group will report to the other, and the combined group will synthesize ideas and propose solutions to issues that have arisen.
- d. A final afternoon session will provide an opportunity for final comments and resolution of any questions raised by the programmers relative to practical implementation of goals or features for UCD.

Day 5 (Project Leaders and graduate students Only).

- a. The Project Leader, Project Co-Leader and two participating graduate students will summarize their notes and the notes of any scribes during the working group meetings to produce a written report of the meeting, containing a set of detailed recommendations for the programmers.

Interim Period (between meetings at NESCENT, 3-4 months)

- a. Programmers will produce working code for the new implementation of UCD, and for both protected access and public portals via the Internet.
- b. As prototype pieces of the system are ready for testing, domain experts will road-test them and identify bugs, problems, and opportunities for improvement. At this time we will involve other specialists in groups of Chalcidoidea not involved in the Working Group meetings in beta-testing of data entry and data curation functions and obtain their feedback.
- c. A Wiki site and password protected user accounts on MX or other suitable content management system will provide appropriate access for participants, and a forum for electronic communication.

Meeting Two, at NESCENT

This meeting will include the domain experts and programmers from meeting one, and new participants selected to identify and create new linkages of UCD to related bioinformatics activities, and identify new opportunities for synthesis and synergy.

Day 1

Morning sessions (all participants):

- a. Demonstration by programmers of data entry portal to the new UCD, followed by demonstrations of search capabilities and formats for data visualization.
- b. Domain experts will practice entering data in their respective families, including problematical use cases. Programmers will assist domain experts with mastering the new UCD interface.

Afternoon Sessions (all participants)

- d. Domain experts will report to programmers on problems or issues encountered in data entry.
- e. The entire group will identify any remaining needs for the UCD structure or user interfaces, and discuss issues that either domain experts or programmers bring to the table.

Day 2

Morning sessions (all participants)

- a. Demonstrations by programmers of search and data visualization functions in new UCD interface.
- b. Practice of search and data visualization functions by domain experts, assisted by the programmers.

Afternoon Session (all participants)

- c. Identify opportunities for linkages and cross-talk with related initiatives in bioinformatics, including Hymeoptera-On-Line and Encyclopedia of Life.

Day 3

Morning sessions (concurrent)

- a. Domain experts will identify any remaining needs for search and data visualization functions.
- b. Programmers and IT experts will identify remaining issues with structure and implementation of the new platform, and opportunities for improving functionality.

Afternoon Session (all participants)

- c. Identify any remaining needs for UCD and assemble a list of priorities for final fine-tuning of the system.
- d. Identify and describe opportunities for Google Summer of Code student programmers to extend UCD and integrate it with other data portals, for example TreeBaseII. These students could be sponsored by NESCENT in summer 2011, and mentored by one or two participants from academic institutions.

Day 4 (Project Leaders and two graduate students only):

- a. The organizers and graduate students will summarize their notes and the notes of any scribes during the working group meetings to produce a final, written report of the meeting for NESCENT and all participants.
- b. Any remaining issues or suggestions for UCD will be prioritized and summarized.

7. Names of proposed participants, both meetings of Working Group:

John Noyes, Natural History Museum, London. Already committed. John is the original author and ultimate domain expert for the UCD system. His participation is absolutely essential, because the project cannot move forward without him.

Andrew Polaszek, Natural History Museum, London. Already committed. Andy has taken over responsibility for maintaining UCD, and he will be the Project Leader for the new UCD, responsible for quality control and coordination and management of all participant curators. His participation is absolutely essential, because the project cannot move forward without him.

John Heraty, University of California, Riverside. Already committed. John has been involved in the production and use of several different databases in Chalcidoidea, and he is a primary user of the MX system in several different taxonomic and systematics projects.

Gary Gibson, Canadian National Collection of Insects, Arachnids and Nematodes. Already committed. Gary has developed the BASIS system, a very similar system to UCD, with the particular capability to produce hard copy taxonomic catalogs. He has a lot of detailed, practical experience with the issues and challenges to be faced in porting UCD to a new platform, and we regard his participation as essential to the project's success.

Katja Seltsmann, North Carolina State University. Already committed. Katja has quite a wide experience base and expertise in many aspects of contemporary biodiversity bioinformatics, particularly with web-based information delivery and the handling of images, through her experience on the MorphBank project.

AnaMaria Dal Molin, Graduate Student Participant, Texas A&M University. Already committed. Ana is a graduate student in Chalcidoidea systematics, originally from Brazil, with very strong interests and expertise in bioinformatics. She represents the next generation that will use and develop UCD.

Jason Mottern, Graduate Student Participant, University of California, Riverside. Already committed. Jason is a graduate student in Chalcidoidea systematics that has done extensive database development work. He also represents the next generation that will use and develop UCD.

Andy Deans, North Carolina State University. Already committed. Andy represents organismal expertise in Hymenoptera outside of Chalcidoidea, thus he brings a different perspective. He has been active in the development of informatics tools and their application in contemporary systematics research, and he has successfully ported a similar database (Evanioidea Online) to the MX platform.

Todd Vision, Associate Director for Informatics, NESCENT. Todd is one of the PIs on the Phenoscope/Phenex projects which makes extensive use of semantic markup. We anticipate the UCD will have similar requirements.

Hilmar Lapp, Assistant Director for Informatics, NESCENT. Hilmar has broad experience with biodiversity data, including work on BioSQL.

Arlin Stoltzfus, National Institute for Standards. Interested, but not yet committed. Comparative Data Analysis Ontology (CDAO). Arlin will help integrate the UCD with the HAO and CDAO, ensuring that data can be integrated into large-scale evolutionary questions.

Proposed participants, second meeting only:

Norm Johnson, Ohio State University. Norm is the author and developer of Hymenoptera Online (<http://hol.osu.edu>), an extensive database with many unique capabilities.

Jim Balhoff, NESCENT. Jim has extensive experience with Phenoscope, and as such could provide input on capturing phenotypical annotation in the UCD.

Representative from EoL. We will request assistance from NESCENT to help identify the best person.

Chris Freeland, Biodiversity Heritage Library. Chris is one of the directors of the BHL and is responsible for IT matters.

Programmers:

Programmer (to be funded by Hymenoptera Anatomy Ontology project, not yet hired).

The HAO has two years of secure funding to hire a full time programmer, and it will commit 5% of this person's time for one year. The HAO will benefit from more accessible data in UCD (for example, pdf images of literature that have been scanned with OCR) and close integration of UCD into MX. This person will have an intimate knowledge of MX and will therefore be highly efficient.

Two student programmers, each to be supported by an award of \$1500 from NESCENT, will develop critical, additional modules and code for enhanced functionality of MX during the interim period between workshops. Their projects will be designed and conducted under the supervision of Yoder and the HAO programmer. We will welcome suggestions from NESCENT as we select these two individuals. Students will benefit by gaining a new toolset (software programming for biodiversity informatics).

Given the opportunity we would also pursue the following individuals:

An EOL fellow (<http://www.eol.org/content/page/fellows>). In conjunction with the EOL we would identify a fellow that would join the programming team. The EOL's goals for fellows are very broadly in sync with the given proposal.

Google Summer of Code Participants. By summer of 2011 we will be in position to identify additional opportunities to engage student programmers. One possibility would provide access to the Application Programming Interface (API) of UCD/MX and integrate it with other portals (e.g. TreebaseII) such that evolutionary-pertinent datasets can be extracted for analysis. These students will be sponsored by NESCENT and mentored by project participants.

8. Rationale for NESCENT support.

NESCENT provides an ideal location, a collegial and efficient setting, and unique resources for a highly integrative project of this sort. Most importantly, NESCENT staff and their associates will provide a broader view of contemporary research activities, emerging trends, and new opportunities in evolutionary biology that will be essential to us as we transform UCD from a resource used largely by taxonomists to a resource supporting diverse research and bioinformatics activities. In a practical sense, NESCENT staff are experienced with working groups, and we will rely on them to provide guidance and expertise in bringing discussions to closure, achieving synthesis of disparate ideas and points of view, and keeping meetings focused and on-task. We additionally recognize the broad expertise of NESCENT staff with respect to data standards and semantic markup, areas that we anticipate needing the most feedback and support on. Finally, three of the project principals are at NSCU (Yoder, Seltsmann and Deans), thus minimizing travel expenses and maximizing potential for value-added, spin-off collaborations from the Working Group meetings.

9. Anticipated IT needs.

During the Working Group meetings, wireless access to the Internet for all participants and standard computer projection facilities will be essential. Beyond that, we do not foresee any particular hardware needs, as all participants will bring their own laptops. A dedicated server computer for MX is already in place at Texas A&M University with more than adequate resources to support the incorporation of UCD (Mac Pro purchased summer 2009: two quad-core Xenon processors, 4 hard drives in two mirrored RAID arrays, 16 GB core memory). We do not foresee that NESCENT would be asked to maintain the server, as it is already maintained by grant funds from other sources and generous donations of time by several people. **We do request \$3000 from NESCENT** to provide two awards to student programmers to help write the code, in collaboration with the programmer to be supported on HAO funding.

10. Proposed time-table

Pre-meeting, Yoder and Dalmolin will work on cross-mapping the data models from UCD-MX, communicating with Noyes when clarifications are needed. Use-cases outlining development requirements, evolutionary questions, and semantic tie-ins will continue to be refined and added to the UCD wiki. Priorities for the first Working Group meeting will continue to be defined, and a more complete agenda will be developed.

First meeting of Working Group, 4 working days plus travel days (5 working days for Project Leaders and graduate students). Given existing commitments by participants, the earliest we can envision this would be late summer of 2010.

Interim Period, HAO programmer and two student programmers will implement recommendations of first meeting during three to four month interim period.

Second meeting of Working Group, 3 working days plus travel days (4 working days for Project Leaders and graduate students). Late winter or early spring, 2011.

Google Summer of Code participants, summer of 2011.

11. Anticipated Results.

1- The UCD database will be transferred to a modern, flexible, extensible programming platform and served up on the Internet from a server hosting MX at Texas A&M University. An initial prototype or individual modules should become available for beta-testing during the period between the two Working Group meetings, or fall of 2010. The full system consisting of records for over 30,000 valid species, over 120,000 host records and over 140,000 distribution records, should be in place by spring or summer of 2011. By this time, the network of specialists to be responsible for data entry and data curation in their respective groups of chalcidoids will have been trained in the system, through participation in the beta-testing period.

- 2- Incorporating UCD into the MX system will combine an authoritative and virtually complete taxonomic catalog with native capabilities for handling specimen data, DNA sequences and morphological character data, and it will provide direct linkages to Morphbank for storage and retrieval of digital images. This greatly expands the potential scope of the UCD.
- 3- Because the transformed UCD will incorporate standards-based markup and other globally accessible data elements (e.g. LSIDs), taxonomic names and associated information for over 30,000 species will be fully exposed to the broader semantic web. These updates in functionality will enhance not only the UCD but also all projects employing MX.
- 4- UCD will be made available for direct linkages to other international-scale bioinformatics resources, such as the EoL, Hymenoptera Online, Catalog of Life, ITIS, GBIF, and the Biodiversity Heritage Library. In particular, tens of thousands of data points (chalcid species pages) will be computable for the EoL.
- 5- UCD will be a model system applicable to other community supported efforts. Enabling continuing support of UCD by specialists in the chalcidoid research community assures its future development, and creates the possibility for incorporation of new data from a much broader spectrum of the scientific literature.
- 6- The new platform for UCD will provide opportunities for broad-scale integration of the data in UCD with other research efforts. For example, thousands of pdf images that have been OCR-scanned will become available for the Hymenoptera Anatomy Ontology project. This information, as well as an authoritative system of taxonomic nomenclature and access to digital images, cladograms and phylogenetic hypotheses, host associations, and history of use in managing pest species will all become immediately available to the large-scale collaborative project on chalcidoid phylogenetics managed by John Heraty, U.C. Riverside, as part of the AToL project in Hymenoptera. But these obvious applications are only the beginning. The possibilities for data-mining and the design of entirely new research strategies to address questions in the evolution of parasitoids and associated evolutionary processes are virtually without limits.

12. CV's of Project Leaders.

JAMES B. WOOLLEY

a. Professional Preparation: Ph.D. (Entomology), University of California, Riverside, 1983
B.S. (Entomology), Oregon State University, 1977.

b. Positions:

1995-present	Professor of Entomology, Texas A&M University
2003-2004	Program Director, Biodiversity Survey and Inventory and Systematic Biology Programs, Division of Environmental Biology, National Science Foundation
1997-2003	Assistant Dept. Head for Graduate Studies, Dept. of Entomology, Texas A&M University
1989-1995	Associate Professor of Entomology, Texas A&M University
1983-1989	Assistant Professor of Entomology, Texas A&M University

c. Five Related Publications

Gibson, G.A.P., J.T. Huber and J.B. Woolley. 1997. *Annotated keys to the genera of Nearctic Chalcidoidea (Hymenoptera)*. NRC Research Press, Ottawa, 794+xi pp.

Woolley, J.B. and P.E. Hanson. 2006. The Family Signiphoridae. Chapter 11.15 in Gauld, I and P.E. Hanson eds, "Neotropical Hymenoptera."

Gibson, G.A.P., J. Heraty and J.B. Woolley. 1999. Phylogenetics and Classification of Chalcidoidea and Mymarommatoidea - a review of current concepts (Hymenoptera, Apocrita). Zoological Scripta, In Press. Zoologica Scripta 28: 87-124.

Noyes, J.S. and J.B. Woolley. 1994. North American encyrtid fauna (Hymenoptera: Encyrtidae): taxonomic changes and new taxa. Journal of Natural History 28: 1327-1401.

Hunter, M.S.H. and J.B. Woolley. 2001. Evolutionary and behavioral ecology of heteronomous aphelinid parasitoids. Annual Review of Entomology 46:251-290.

Five Other Publications:

Unruh, T.R. and J.B. Woolley. 1999. Molecular Methods in Classical Biological Control. Chapter 4, pp. 57-85. in T. Bellows et al. (eds.) *Handbook of Biological Control*, Academic Press, San Diego.

Sharkov, A. and J.B. Woolley. 1997. A revision of the genus *Hambletonia* Compere (Hymenoptera: Encyrtidae). Journal of Hymenoptera Research 62(2): 191-218.

Woolley, J.B., M. Rose and P. Krauter. 1994. Morphometric comparisons of *Aphytis species* in the *lingnanensis* group (Hymenoptera: Aphelinidae). Chapter 12, pages 223-244, in D. Rosen, ed., *Advances in Aphytis Research*, Intecpt Ltd., Andover, UK, 362 pages.

Heraty, J.M., J.B. Woolley and D. C. Darling. 1994. Phylogenetic implications of the mesofurca and mesopostnotum in Hymenoptera. Journal Hymenoptera Research 3:241-277.

Heraty, J.M., J.B. Woolley and D.C. Darling. 1997. Phylogenetic implications of the mesofurca in Chalcidoidea (Hymenoptera). with emphasis on structures in Aphelinidae. *Systematic Entomology* 22:45-65.

d. Synergistic Activities:

Service at Texas A&M University: a) currently Chair of Interdisciplinary Research Program in Ecology and Evolutionary Biology, representing 68 faculty in 6 colleges and 16 departments.

Service to Systematics Community: a) Member, Steering Committee, NSF-funded Research Coordination Network to develop a community of natural history collections in U.S., 2007 to present, b) Co-chair of Steering Committee for LINNE, Legacy Infrastructure Network for Natural Environments, 2005 to present.

Service at NSF: a) Program Director in Biodiversity Survey and Inventory and Systematic Biology Programs, Chair of Working Group for Assembling the Tree of Life Program, served on Biology Working Group in Cyberinfrastructure, b) NSF Review Panels, Systematic Biology Program, April, 1990, April, 1991, April 1992, October 1992, April 1993, February 1998, April 2000, October 2007

Development of Professional Societies: a) currently President of International Society of Hymenopterists, participated in founding the society, b) active in Entomological Collections Network, maintained database of membership until 2003.

Editorial Service: a) Editor-In-Chief, Thomas Say Series in Entomology, b) North American Editor, *Cladistics*, 2000 to 2004, c) Associate Editor, *Systematic Biology*, 1994-1995.

e. Recent Grant Support:

NSF, REU Supplement to DEB 0730616, Consolidation of research and training activities in Chalcidoidea. Co-PI with John Heraty, UCR, \$16,000, 2009-2010.

NSF, REU Supplement to DEB 0730616, Consolidation of research and training activities in Chalcidoidea. Co-PI with John Heraty, UCR, \$14,000, 2008-2009.

NSF, DEB 0730616, Consolidation of research and training activities in Chalcidoidea. Co-PI with John Heraty, UCR, \$750,000, Sept 2007- Sept. 2112.

CONACYT, Encyrtid parasitoids of mealybugs in Mexico (Hymenoptera: Encyrtidae), Co-PI with Alejandro Gonzalez Hernandez, Univ. Auto. de Nuevo Leon, \$24,995, 2003-2004.

USDA/CSRS/NRI-CRG, CRIS 1926-22000-009-08S, Evolution of host range, courtship and morphology in sibling species, Co-PI with K. Hopper and J.M. Heraty, \$54,625 subcontract, February 1, 2003 – January 31, 2005.

Thesis Advisor and Post-Doctoral Sponsor: **Catanach**, Therese (M.S.), now in Ph.D. program at University of Illinois; **DalMolin, Ana**, (current, Ph.D.); **Gillooly**, Alan (Ph.D.), retired; **Goolsby**, John (Ph.D.), Research Scientist, USDA/ARS, Weslaco, Texas; **Heraty**, John (Ph.D.), Professor, University of California, Riverside; **Hunter**, Molly (Post-Doc), Professor, University of Arizona; **Johnson**, Roishene (M.S.), Centenary College, Shreveport, Louisiana; **Judd**, Darlene (Ph.D.), current status unknown; **Kim**, Jung Wook (Post-Doc), Postdoc, North Carolina State University; **Menard**, Katrina (current, Ph.D.); **Moomaw**, Charles (M.S.), current status unknown; **Rodriguez Velez**, Beatriz (Ph.D.), Postdoc at Universidad Nacional Autónoma de México; **Stauffer**, Steven (M.S.), deceased; **Yoder**, Matt (M.S.), Postdoc, North Carolina State University; **Zolnerowich**, Greg (Ph.D.), Associate Professor, Kansas State University.

[15 total; 7 women including 3 Latinas and 1 African-American]

MATTHEW YODER

EDUCATION

- **Ph.D.**, Entomology, Texas A&M University, College Station, Texas, 2007
Dissertation: *Advances in Diapriid (Hymenoptera: Diapriidae) systematics, with contributions to cybertaxonomy and the analysis of rRNA sequence data*
- **M.Sc.**, Entomology, Texas A&M University, College Station, Texas, 2002
Thesis: *A revision of the North American species of Entomacis Foerster (Hymenoptera: Diapriidae) with a preliminary phylogenetic analysis of the worldwide species*
- **Honors B.Sc.** (minor in biology), University of Waterloo, Ont. Canada, 1995

PROFESSIONAL EXPERIENCE

- *Researcher, NSF funded Advanced in Biological Informatics - “The Hymenoptera Anatomy Ontology”* (with Andy Deans), North Carolina State University. 2009 - present
- *Postdoc Research Assistant, NSF funded Platygastroidea Planetary Biodiversity Inventory* (under Dr. Norman Johnson), Ohio State University. 2007- 2009
- *Research Assistantship, NSF PEET Program* (under Dr. Jim Woolley and/or Dr. Bob Wharton). July 2000 - May 2007
- *Teaching Assistant – Insect Systematics* (under Dr. John Oswald). Aug. 1999 - Dec. 1999
- *Research Assistant – Insect Collection Database Developer* (under Dr. John Oswald). Aug. 1998 - July 2000

MOST RECENT PUBLICATIONS (13 TOTAL)

- **Yoder, M.J.**, Miko, I., Seltmann, K., Bertone, M. and A.R. Deans. The Hymenoptera Anatomy Ontology. (*Final stages of prep to BMC Bioinformatics*)
- Mallat, J., Waggoner Craig, C. and **M.J. Yoder**. (*In press*) Nearly complete rRNA genes assembled from across the metazoan animals: Effects of more taxa, improved alignment, and paired-sites evolutionary models on phylogeny reconstruction? *Molecular Phylogeny and Evolution*
- **Yoder, M.J.**, Valerio, A., Polaszek, A., and N.F. Johnson. 2009. A revision of the world species of the charismatic *Scelio pulchripennis* species group (Hymenoptera: Platygastriidae). *ZooKeys* 20: 53-118
- **Yoder, M.J.**, A.A. Valerio, L. Masner and N.F. Johnson. 2009. Identity and synonymy of *Dicroscelio* Kieffer and description of *Axea*, a new genus from tropical Africa and Asia (Hymenoptera: Platygastroidea, Platygastriidae). *Zootaxa* 2003: 1-45.
- Wharton, R. A., Roeder, K., and **M.J. Yoder**. 2008. A Monograph of the genus *Westwoodia* (Hymenoptera: Ichneumonidae). *Zootaxa* 1855: 1-40
- **Yoder, M.J.** 2007. *Mannomicrus* (Hymenoptera: Diapriidae), a new genus of myrmecophilic diapriid, with a digital version of Masner and García's (2002) key to New World Diapriinae and an illustration of digital description and key markup using an ontology. *Zootaxa* 1439: 47-55

INVITED PRESENTATIONS

- **Yoder, M.J.** (2009) The practical cybertaxonomist. Michigan State University, Department of Entomology
- **Yoder, M.J.** and Johnson, N. (2008) Large scale taxonomic efforts: The trials and tribulations of describing *lots* of species. Museo Argentino de Ciencias Naturales "Bernardino Rivadavia", Buenos Aires, Argentina
- **Yoder, M.J.** (2008) Managing systematics data in mx. NSF PEET Workshop, Riverside, California
- **Yoder, M.J.** (2007) Collaboration in systematic biology. Texas A&M Department of Entomology Seminar Series

BIODIVERSITY INFORMATICS

- Lead developer of 'mx', a web-based open-source collaborative on-line database
 - used by over 40 people from 6 labs in 3 countries
- Primary developer of the TAMU Insect Collection Curatorial and Specimen-Level Database
 - houses well over 120k specimen records (barcoded), 30k taxonomic names
- Strong knowledge of relational schemas pertinent to collections and systematics research
- Extensive experience with Perl, BioPerl, Ruby, and the Rails web application framework
- Web page design and implementation including experience in php, CSS, XML, XSLT
- Extensive experience using a wide range of phylogenetic inference software including TNT, PAUP*, MrBayes, and PHASE

SELECTED ELECTRONIC PUBLICATIONS and PROJECTS *(personal contributions)*

- Hymenoptera Anatomy Ontology (application, data) – <http://hymao.org/>
- mx (project leader, developer) – <http://hymenoptera.tamu.edu/wiki/>
- rubyBHL Ruby Gem (owner, developer)
- GENERator (owner, developer) – <http://github.com/mjy/GENERator>
- The Diapriidae – <http://www.diapriid.org/>
- jRNA/Psy (web site/Perl code/data archive) – <http://hymenoptera.tamu.edu/rna/>
- Catalogous Evaniidorum (database backend) – <http://evaniid.tamu.edu/>

GRANTS

- \$1,411,508 – NSF, ADVANCES IN BIOLOGICAL INFORMATICS, DBI 0850223: “THE HYMENOPTERA ONTOLOGY: PART OF A TRANSFORMATION IN SYSTEMATIC AND GENOME SCIENCE.” 4/2009-4/2012 (Co-written with A.R.Deans, P.I.)
- Travel funds - Museo Argentino de Ciencias Naturales, Buenos Aires (2008) ~ \$2000
- TAMU Department of Entomology Graduate Student Forum (2006) 1st Place – \$250
- TAMU Ecology and Evolutionary Biology Travel Grant to (2006) – \$800
- Texas A&M Entomology Graduate Student Grants (1999, 2001, and 2002) ~ \$1200 total