

**Annual Report for Period:**08/2005 - 08/2006**Submitted on:** 06/27/2006**Principal Investigator:** Schatz, Bruce R.**Award ID:** 0425852**Organization:** U of Ill Urbana-Champaign**Title:**  
FIBR: BeeSpace - An Interactive Environment for Analyzing Nature and Nurture in Societal Roles**Project Participants****Senior Personnel****Name:** Schatz, Bruce**Worked for more than 160 Hours:** Yes**Contribution to Project:****Name:** Robinson, Gene**Worked for more than 160 Hours:** Yes**Contribution to Project:****Name:** Fahrbach, Susan**Worked for more than 160 Hours:** Yes**Contribution to Project:****Name:** Rodriguez-Zas, Sandra**Worked for more than 160 Hours:** Yes**Contribution to Project:****Name:** Zhai, ChengXiang**Worked for more than 160 Hours:** Yes**Contribution to Project:****Name:** Bruce, Bertram**Worked for more than 160 Hours:** Yes**Contribution to Project:**

Project Lead for Education and Outreach.

Senior graduate student in education supported on grant, plus summer salary for Biology Teacher at University Laboratory High School.

**Post-doc****Graduate Student****Undergraduate Student****Technician, Programmer****Name:** Littell, Todd**Worked for more than 160 Hours:** Yes**Contribution to Project:**

Chief Programmer developing the software system

**Other Participant**

**Research Experience for Undergraduates****Organizational Partners****Texas A&M University Main Campus**

computational annotation of the honeybee genome

**CORNELL UNIVERSITY**

fulltext digitalization of the beekeeping literature

**Other Collaborators or Contacts**

Chris Elsik, Department of Animal Sciences, Texas A&M University

Janet McCue, Mann Biology Library, Cornell University

**Activities and Findings****Research and Education Activities: (See PDF version submitted by PI at the end of the report)**

See attached file with details of our substantial efforts into  
Biology and Informatics and Education.

**Findings:**

The Biology experiments are still underway and not yet been analyzed.

The Informatics developments are still underway and not yet been validated, although there is some preliminary usage.

**Training and Development:**

We have been focusing on providing experiences with real science research as appropriate for students at levels ranging from postgraduate to middle school. A small group of 5-10 for each level.

Graduate Students. Work closely with postdocs and PhDs for judging results and helping their research.

Building early adopter community at external sites.

Undergraduate. Planning for freshman bioinformatics course in Fall at Wake Forest. Students will then train middle school students at Urbana during the summer as their service requirement.

High School. Paying biology teacher at University Laboratory High School to integrate materials into Field Biology course.

**Outreach Activities:**

We host an extensive website of slides and videos, including hosting the campuswide Biomedical Informatics seminar.

See <http://www.beespace.uiuc.edu>

We did extensive outreach to Campus Middle School for Girls, including investigator lectures on their research and field trips to bee research facility.

**Journal Publications**

Gene Robinson, "Beyond Nature and Nurture", Science, p. 397, vol. 304, (2004). Published

Honey Bee Genome Sequencing Consortium, "The genome sequence of the honey bee, *Apis mellifera*, a highly social animal", Nature, p. , vol. , ( ). Accepted

R. Velarde, G. Robinson, S. Fahrbach, "Nuclear receptors of the honey bee", Insect Molecular Biology. Honey bee genome special issue., p. , vol. , ( ). Accepted

A. Hummon, T. Richmond, P. Verleyen, G. Baggerman, J. Huybrechts, M. Ewing, E. Vierstraete, S. Rodriguez-Zas, L. Schoofs, G. Robinson, J. Sweedler, "From the Genome to the Proteome: Uncovering Peptides in the Apis Brain", Science, p. , vol. , ( ). Accepted

S. Rodriguez-Zas, B. Southey, C. Whitfield, G. Robinson, "Characterization of unique gene expression trajectories across behavioral maturation in honey bees using a semiparametric model", Genome Research, p. , vol. , ( ). Submitted

C. Whitfield, Y. Ben-Shahar, C. Brilllet, I. Leoncini, D. Crauser, Y. LeConte, S. Rodriguez-Zas, G. Robinson, "Genomic dissection of behavioral maturation in the honey bee", PNAS, p. , vol. , ( ). Submitted

X. Ling, J. Jiang, X. He, Q. Mei, C. Zhai, B. Schatz, "Automatically Generating Gene Summaries from Biomedical Literature", Proceedings of Pacific Symposium on Biocomputing 2006 (PSB'06), p. 40-51, vol. , (2006). Published

J. Jiang, C. Zhai

, "Exploiting Domain Structure for Named Entity Recognition", Proceedings of HLT/NAACL, p. , vol. , (2006). Accepted

Q. Mei, C. Zhai, "A Mixture Model for Contextual Text Mining", Proceedings 2006 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (KDD'06), p. , vol. , (2006). Accepted

Q. Mei, C. Liu, H. Su, C. Zhai, "A Probabilistic Approach to Spatiotemporal Theme Pattern Mining on Weblogs", Proceedings of the World Wide Web Conference 2006 ( WWW'06), p. , vol. , (2006). Accepted

Q. Mei, D. Xin, H. Cheng, J. Han, C. Zhai, "Generating Semantic Annotations for Frequent Patterns with Context Analysis", Proceedings 2006 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '06), p. , vol. , (2006). Accepted

B. Chee, B. Schatz, "Document Clustering using Small Worlds Communities", Proceedings 2006 ACM Conference on Information and Knowledge Management (CIKM '06), p. , vol. 15, (2006). Submitted

### **Books or Other One-time Publications**

### **Web/Internet Site**

**URL(s):**

<http://www.beespace.uiuc.edu>

**Description:**

The project website. created this first year.  
 Contains detailed information on the background behind the project.  
 Contains slides and videos of the lectures given during the project.  
 Contains prototype software system for test users.

### **Other Specific Products**

**Product Type:**

**Software (or netware)****Product Description:**

User manual and sample sessions.

**Sharing Information:**

Freely available at project website  
<http://www.beespace.uiuc.edu>

**Product Type:****Audio or video products****Product Description:**

Slides and Videos of Project Lectures.  
 Overviews and Details.

**Sharing Information:**

Freely available on project website.

**Contributions****Contributions within Discipline:**

coPI Gene Robinson was elected to the  
 National Academy of Sciences  
 in the section on Evolutionary Biology

**Contributions to Other Disciplines:****Contributions to Human Resource Development:**

Established Bioinformatics Laboratory in  
 new Institute for Genomic Biology at the  
 University of Illinois at Urbana-Champaign

**Contributions to Resources for Research and Education:**

the various informatics investigators  
 (Schatz, Zhai, Rodriguez-Zas) have helped  
 establish a new Master's degree program in  
 Bioinformatics at the University of Illinois

Preliminary Work to establish PhD program  
 (campuswide interdisciplinary) in biomedical informatics.

**Contributions Beyond Science and Engineering:**

coPI Gene Robinson published an OpEd piece in the New York Times about Nature-Nurture, a popular version of the project topic.

Schatz and Robinson authored an essay on BeeSpace and its implications for understanding nature/nurture for a public magazine sponsored by the Institute for Genomic Biology.

**Special Requirements**

**Special reporting requirements:** None

**Change in Objectives or Scope:** None

**Unobligated funds:** \$ 0.00

**Animal, Human Subjects, Biohazards:** None

**Categories for which nothing is reported:**

Any Book

Contributions: To Any Other Disciplines

## Research and Education, BeeSpace project August 2005 to August 2006

This year was the second. We made progress on all aspects of Research, both Biology and Informatics, and of Education, both training and outreach. These are outlined below.

### Biology

The goal of the Biology Research is to carefully dissect the relative contributions of nature and nurture for social behavior in the honey bee. In particular, we are experimentally measuring brain gene expression for important societal roles during normal behavior, with different experiments varying heredity (nature) and environment (nurture) for different societal roles. Last year, we laid out our detailed master plan; this year we have begun to substantially execute it.

The honey bee genome is now complete; it was generated as part of the second wave of NIH sequencing at Baylor. We have completed a genome annotation, using a computational pipeline based on ortholog comparisons to assign Gene Ontology categories to each generated sequence. We have been running initial experiments with a previous partial microarray based on an EST library. The complete microarray based on the genome sequence is now fabricated, and we shall be using it in our full experiments being carried out in the laboratory of coPI Robinson.

It is expected that the genome sequence of the honey bee will be published in a special issue of *Nature* in late summer-early fall. The special issue will include cover art, news coverage, and specially commissioned News & Views pieces on this milestone. There also will be related articles at the same time in *Science*, *Genome Research* and *Insect Molecular Biology*. While this is preliminary work for this project, our articles will explicitly mention BeeSpace as **the** environment for post-genome functional analysis.

Our discussions between investigators Robinson and Schatz transformed the master list of expression experiments from a standard behavior list for social insects into a detailed list of nature/nurture dissections for social behavior. We are focusing on Foraging and Defense, as symbolic of animal behavior for food and warfare. Hopefully, this will enable comparisons to higher organisms. The honey bee is a model for natural behavior; we are collecting bees during their normal behavior in the field. Half the bees for the experiments were collected last summer and the rest are being collected this summer.

For Defense, we will dissect nature via different races of bees (European, African) and nurture via different levels of threat (manipulations of alarm pheromones). For Foraging, our primary focus will be on when a bee transitions her societal role from nurse to forager (age of onset of foraging). We will again dissect nature via different races of bees and dissect nurture via social manipulations (e.g. manipulating food supply in the hive to create precocious foragers or overage nurses) and physiological manipulations (e.g. manipulating NPF with vitellogenin or JH Juvenile Hormone with octopamine). The dissection of the brains from collected bees and the statistical analysis of the microarray expressions by coPI Rodriguez-Zas will be carried out during the next year.

Our plan for localization of gene expression using in situ hybridization focuses on analysis of lists of candidate genes based on whole genome microarray studies of the brains of bees collected in different behavioral states. This plan is being executed at Wake Forest University in the laboratory of coPI Fahrbach. Since the expression experiments have been delayed due to the delay of the genome sequence, we are instead using a list of genes encoding neuropeptides with key roles in regulation of behavior. We have separately identified 36 neuropeptide-encoding genes as the basis for our initial BeeSpace in situ hybridization studies. We will develop a standardized method for probe design and a “bank” of information on construction of probes for bee tissue, using vitellogenin as a universal “positive control” for hybridization of honey bee tissues.

### **Informatics**

The goal of the Informatics Research is to build BeeSpace, an interactive environment for functional analysis. The technologies are general for semantic text analysis, but the collections are special for insect social behavior. In particular, our software environment will interactively annotate gene functions for important gene clusters using concept navigation across biological literature representing community knowledge.

Last year, we developed an initial prototype for concept navigation, which has had some initial usage from local biologists and at our annual workshop. This year, we are developing the first fully fledged system for our early adopter users. This is being tested by our local users and should be available by the end of this project year (August). A preliminary version is already available (as is the first prototype) on our website <http://www.beespace.uiuc.edu>.

The system components are being developed by graduate students in computer science and bioinformatics, who are publishing papers on the research aspects of each component. These students are supervised by coPI Zhai and PI Schatz. The system integration and architectural foundations are being developed by our senior research programmer Todd Littell. Version 2 will have a full complement of C3 Concepts and Categories and Collaborations.

For community collections, we continue to use the 10-15 carefully generated from Biological Abstracts relevant to social behavior (averaging 5-10K abstracts in size). These partition the scientific literature in different ways across organism (bee, fly), method (chemistry, physiology), behavior (foraging, agonism), and so on. We have obtained all of Medline from the National Library of Medicine and are in negotiations for all of Biosis from Thomson Scientific. These collections have been semantically indexed, using concept graphs of mutual information to support conceptual navigation.

Users can navigate search text, using system support to build up themes of related concepts. They can also navigate from concept to concept, then view the corresponding documents. Each collection is automatically grouped into related clusters, which can be

interactively viewed for category navigation. There are initial links into the genome databases through gene names, to provide relevant sentences for an automatically generated structured gene summary. The goal is easy navigation from high-level concepts from one viewpoint to low-level concepts in another, and thence into genomic data.

The research components include: a noun phrase parser that automatically recognizes biological entities, a theme extractor that automatically discovers related terms in a collection, a concept navigator with visualization across multiple levels of topic resolution, a gene summarizer that automatically extracts relevant sentences from a collection, and an annotation pipeline that automatically transforms gene expressions into functional sentences from a community collection.

## **Education**

We are training students at many levels, concentrating on giving research experiences to small groups of focused persons. Our investigators and researchers are directly involved in education and outreach. Extensive photos and videos of these activities are available.

At the Graduate level, we support 10 students in Biology and in Informatics. Their department affiliations include entomology and neuroscience, computer science and statistics. The interdisciplinary interactions are facilitated by our new Institute for Genomic Biology. We also work closely with our early adopter community of biologists at other universities and research institutions.

At the Undergraduate Level, we have developed a new bioinformatics course based on BeeSpace. This will be taught over the next several years for freshman at Wake Forest University by coPI Fahrback. The service component of this course will be to have the students plan and execute our summer workshop for disadvantaged students at the middle school level in Champaign County of Illinois.

For K12, each summer we are paying the biology teacher at the University High School to integrate social insects and BeeSpace into his Field Biology course. We are also relying on the extensive contacts of coPI Chip Bruce for science education.

For outreach, we use our arrangement with the Campus Middle School for Girls. One of our female biology students (actually the same one who was judging the informatics results) described her thesis research to middle school girls in a hands-on interaction. Previously, our project coordinator had lectured on bee social behavior with slides and videos, with a field trip to the Bee Research Facility with lecture by coPI Robinson.

## **Special Requirements**

Our funds from NSF were frontloaded, in that more money actually arrived in the first two years than we requested. Because of this, we have underspent the funds and wish to push them forward into later years. The PI (Schatz) has significant experience in running large NSF systems projects and this pattern is quite typical. The planning years underspend, but the development years overspend, as the project ramps up with real systems and real users. So we appreciate the flexibility of NSF in permitting us to responsibly spend the funds when the needs of the project dictate.