

**Annual Report for Period:**09/2008 - 08/2009**Submitted on:** 09/09/2009**Principal Investigator:** Schatz, Bruce R.**Award ID:** 0425852**Organization:** U of Ill Urbana-Champaign**Submitted By:**

Schatz, Bruce - Principal Investigator

**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****Name:** Sen Sarma, Moushumi**Worked for more than 160 Hours:** Yes**Contribution to Project:**

Biology Advisor for software functionality and user testing

**Graduate Student****Undergraduate Student****Technician, Programmer**

**Name:** Buell, Jim

**Worked for more than 160 Hours:** Yes

**Contribution to Project:**

project coordinator for user community and education/outreach

**Name:** Arcoleo, David

**Worked for more than 160 Hours:** Yes

**Contribution to Project:**

Lead programmer for system infrastructure and user interface.

**Name:** Sanders, Barry

**Worked for more than 160 Hours:** Yes

**Contribution to Project:**

Informatics Programmer for database integration and interface design

**Other Participant**

**Research Experience for Undergraduates**

**Organizational Partners**

**Georgetown University**

Computational annotation of the honey bee genome, working with BeeBase researchers

**CORNELL UNIVERSITY**

fulltext digitalization of the beekeeping literature

**Indiana University**

FlyBase curator site, working to evaluate our analysis software

**Harvard University**

FlyBase curator site, working to evaluate analysis software

**Kansas State University**

Community organization of Arthropod Base Consortium, working with BeetleBase researchers

**Other Collaborators or Contacts**

Chris Elsik, Department of Biology, Georgetown University

Nan Hyland, Mann Biology Library, Cornell University

Kathy Mathews, Department of Biology, Indiana University

Bill Gelbart, Department of Biology, Harvard University

Susan Brown, Department of Genetics, Kansas State University

At University of Illinois at Urbana-Champaign:

May Berenbaum, Department of Entomology

Nathan Price, Department of Chemical & Biomolecular Engineering

Saurabh Sinha, Department of Computer Science

Sheng Zhong, Department of Bioengineering

Matthew Hudson, Department of Crop Sciences

John MacMullen, School of Library and Information Science

## 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 Research (Informatics and Biology) and Education (including Outreach).

### **Findings:**

For Biology Research, the nature-nurture dissection has been completed with microarray experiments and the analyses are proceeding. For Aggression (Defense), the molecular mechanisms appear to be the same for nature (evolutionary) and for nurture (societal), in terms of common gene clusters. These results have been published in PNAS.

For Informatics Research, the final software system for functional analysis has been released. This version 4 has a much streamlined user interface and fully supports creation of user spaces that can be used as background collections for genelist analysis from microarray experiments. These results have been released to the Arthropod Base Consortium for genomics researchers.

For Education Research, an electronic curriculum for modern genomics based on BeeSpace has been created and distributed via our website.

### **Training and Development:**

Graduate Students. Work closely with 20 postdocs and PhDs for judging results and helping their research, in departments ranging from neuroscience to computer science.

Building early adopter community at external sites, notably the Arthropod Base Consortium spread across the US and Europe.

Undergraduate Students. Developed freshman bioinformatics course based on BeeSpace at Wake Forest. Taught for third time this year, including lectures by project investigators.

High School Students. Paying biology teacher at University High School to plan and teach summer workshop for 15 Uni students, who will leverage the materials into national science competition projects.

### **Outreach Activities:**

We support an extensive website of slides and videos of our research. The educational materials are packaged for external distribution. See <http://www.beespace.uiuc.edu>

We set up a computer laboratory for outreach partners at University Laboratory High School and at Campus Middle School for Girls. Our project coordinator salvaged and refurbished all these machines.

## 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. 443, (2006). Published,

R. Velarde, G. Robinson, S. Fahrback, "Nuclear receptors of the honey bee: Annotation and Expression in the Adult Brain", *Insect Molecular Biology*. Honey bee genome special issue., p. , vol. 15, (2006). Published,

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. 314, (2006). Published,

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. Brillet, I. Leoncini, D. Crauser, Y. LeConte, S. Rodriguez-Zas, G. Robinson, "Genomic dissection of behavioral maturation in the honey bee", *PNAS*, p. , vol. 103, (2006). Published,

- 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). Published,
- 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). Published,
- 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). Published,
- 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). Published,
- B. Chee, B. Schatz, "Document Clustering using Small Worlds Communities", Proceedings 2007 Joint (ACM/IEEE) Conference on Digital Libraries, p. , vol. , (2007). Published,
- Xu Ling, Jing Jiang, Xin He, Qiaozhu Mei, Chengxiang Zhai, Bruce Schatz, "Generating Semi-Structured Gene Summaries from Biomedical Literature", Information Processing & Management, p. 1777, vol. 43, (2007). Published, 10.1016/j.ipm.2007.01.018
- Yue Lu, Xin He, Sheng Zhong, "Cross-species microarray analysis with the OSCAR system suggests an INSR->Pax6->NQO1 neuro-protective pathway in ageing and Alzheimer's disease", Nucleic Acids Research, p. , vol. , ( ). Accepted,
- S. Sinha, X. Ling, C. Whitfield, C. Zhai, G. Robinson, "Genome scan for cis-regulatory DNA motifs associated with social behavior in honey bees", PNAS, p. , vol. 103, (2006). Published,
- C. Whitfield, Y. Ben-Shahar, C. Brillet, I. Leoncini, D. Crauser, Y. LeConte, S. Rodriguez-Zas, G. Robinson, "Thrice Out of Africa: ancient and recent expansions of the honey bee", Science, p. , vol. 314, (2006). Published,
- T. Tao, X. Wang, Q. Mei, C. Zhai, "Language Model Information Retrieval with Document Expansion", Proceedings of HLT/NAACL, p. , vol. , (2006). Published,
- G. Robinson, J. Evans, R. Maleszka, H. Robertson, D. Weaver, K. Worley, R. Gibbs, G. Weinstock, "Sweetness and Light: Illuminating the Honey Bee Genome", Insect Molecular Biology, p. , vol. 15, (2006). Published,
- Y. Lu, X. Ling, X. He, C. Zhai, "Language Models for Genomics Information Retrieval", TREC Genomics Track, p. , vol. Jun, (2007). Published,
- J. Jiang, C. Zhai, "Instance Weighting for domain adaptation in NLP", Proc 45th Annual Meeting of Association for Computational Linguistics (ACL), p. 264, vol. Jun, (2007). Published,
- X. Ling, Q. Mei, B. Schatz, C. Zhai, "Mining multifaceted overviews of arbitrary topics in a text collection", Proc 15th ACM KDD Knowledge Discovery and Data Mining, p. 497, vol. Aug, (2008). Published,
- B. Schatz, "Interactive Functional Analysis of Arthropod Genome Databases", Arthropod Genomics Symposium annual meeting, p. , vol. April, (2008). Published,
- X. Ling, J. Jiang, Q. Mei, M. Sen Sarma, C. Zhai, B. Schatz, "Gene Summarizer: Software for Automatically Generating Structured Summaries from Biomedical Literature", 2nd International Biocuration Meeting, p. , vol. Oct, (2007). Published,
- B. Sanders, D. Arcoleo, B. Schatz, "BeeSpace: Interactive Analysis of Genome Databases", BOSC Bioinformatics Open Source Conference, p. , vol. Jul, (2008). Published,

- M. Sen Sarma, C. Whitfield, G. Robinson, "Species differences in brain gene expression profiles in honey bees", *BMC Genomics*, p. 202, vol. 8, (2007). Published,
- C. Alaux, G. Robinson, "Releaser pheromone provokes immediate-early gene and slow behavioral response", *J. Chemical Ecology*, p. 1346, vol. 33, (2007). Published,
- R. Kantorovitz, G. Robinson, S. Sinha, "A statistical method for alignment-free comparison of regulatory sequences", *Proc 15th Intl Conf Intelligent Systems for Molecular Biology (ISMB)*, p. , vol. Jul, (2008). Published,
- A. Toth, K. Varala, T. Newman, J. Simons, M. Egholm, J. Hunt, M. Hudson, G. Robinson, "Wasp brain gene expression supports an evolutionary link between maternal behavior and eusociality", *Science*, p. 441, vol. 318, (2007). Published,
- A. Toth, G. Robinson, "Evo-Devo and the evolution of social behavior", *Trends in Genetics*, p. 334, vol. 23, (2007). Published,
- N. Ismail, S. Christine, G. Robinson, S. Fahrbach, "Pilocarpine improves recognition of nestmates in young honey bees", *Neuroscience Letters*, p. , vol. , (2008). Accepted,
- C. Smith, A. Toth, A. Suarez, G. Robinson, "Genetic and genomic analyses of division of labour in the insect societies", *Nature Reviews Genetics*, p. , vol. , (2008). Accepted,
- M. Sen Sarma, S. Rodriguez-Zas, F. Hong, S. Zhong, G. Robinson, "Transcriptomic Profiling of Central Nervous System Regions in Three Species of Honey Bee During Dance Communication Behavior", *PLoS One*, p. , vol. , (2009). Published,
- C. Alaux, N. Duong, S. Schneider, B. Southey, S. Rodriguez-Zas, G. Robinson, "Modulatory communication signal performance is associated with a distinct neurogenomic state in honey bees", *PLoS One*, p. , vol. , (2009). Published,
- C. Alaux, S. Sinha, L. Hasadsri, G. Hunt, E. Guzman-Novoa, G. Degrandi-Hoffman, J. Uribe-Rubio, S. Rodriguez-Zas, G. Robinson, "Honeybee Aggression Supports a Link Between Gene Regulation and Behavioral Evolution", *Proceedings National Academy Sciences PNAS*, p. , vol. , (2009). Published,
- R. Velarde, S. Fahrbach, "Neuropeptide precursor genes in the honey bee: Expression profiles and mRNA localization in the adult worker honey bee", *Annual Meeting of Society for Neuroscience, Chicago, October*, p. , vol. , (2009). Published,
- J. Buell, D. Stone, N. Naeger, S. Fahrbach, C. Bruce, B. Schatz, "Experiencing BeeSpace: Educational Explorations in Behavioral Genomics for High School and Beyond", *Annual Meeting of American Association for Advancement of Science, Chicago, February*, p. , vol. , (2009). Published,
- D. Stone, J. Buell, N. Naeger, "Electronic BeeSpace: A Video Curriculum for Behavioral Genomics", *Science Teacher, National Science Teachers of America journal*, p. , vol. , (2009). Submitted,
- X. He, M. Sen Sarma, X. Ling, B. Chee, C. Zhai, B. Schatz, "Analyzing gene lists from literature: a statistical approach based on Poisson mixture model", *Bioinformatics*, p. , vol. , (2010). Submitted,
- B. Chee, R. Berlin, B. Schatz, "Measuring Population Health using Personal Health Messages", *Annual Meeting of American Medical Informatics Association (full paper)*, p. , vol. , (2009). Accepted,
- Y. Lv, C. Zhai, "Adaptive Relevance Feedback in Information Retrieval", *Proceedings 18th ACM International Conference Information and Knowledge Management (full paper)*, p. , vol. , (2009). Accepted,
- Y. Ko, C. Zhai, S. Rodriguez-Zas, "Inference of Gene Pathways using Mixture Bayesian Networks", *BMC Systems Biology*, p. 54, vol. 3, (2009). Published,
- Y. Lu, C. Zhai, N. Sundaresan, "Rated Aspect Summarization of Short Comments", *Proceedings of the World Wide Conference*, p. 131, vol. , (2009). Published,

Y. Lu, H. Fang, C. Zhai, "An Empirical Study of Gene Synonym Query Expansion in Biomedical Information Retrieval", Information Retrieval, p. 51, vol. 12, (2009). Published,

Q. Mei, D. Zhang, C. Zhai, "Smoothing Language Models with Document and Word Graphs", Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, p. 611, vol. , (2008). Published,

A. Shakery, C. Zhai, "Smoothing Document Language Models with Probabilistic Term Count Propagation", Information Retrieval, p. 139, vol. 11, (2008). Published,

### Books or Other One-time Publications

#### Web/Internet Site

**URL(s):**

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

**Description:**

The project website.

Contains detailed information on the background behind the project.

Contains slides and videos of the lectures given during the project.

Contains released informatics software for biology users.

Contains released education curriculum for outreach students.

Contains scientific publications and educational materials.

#### Other Specific Products

**Product Type:**

**Audio or video products**

**Product Description:**

Slides and Videos of Project. Overviews and Training.

**Sharing Information:**

Freely available on project website.

**Product Type:**

**Software (or netware)**

**Product Description:**

BeeSpace analysis environment supporting concept navigation.

**Sharing Information:**

available from website <http://www.beespace.uiuc.edu>

#### Contributions

**Contributions within Discipline:**

Informatics Research is pioneering conceptual navigation of biological literature and distributing related software to the Arthropod Base Consortium of NSF-funded biologists. We gave big demonstration and plenary talk at annual Arthropod Genomics Symposium for second year in row.

Biology Research is pioneering large-scale gene expression analysis of insect social behavior. We published big study in PNAS about aggression gene clusters in nature and in nurture, comparing evolutionary to societal changes.

**Contributions to Other Disciplines:**

Collaboration with FlyBase is leading towards automatic support for biology curation. Important for future of biology as genome sequencing becomes routine and each community must provide their own annotations. We gave plenary talk at International Biocuration Meeting this year.

**Contributions to Human Resource Development:**

Established research facility in new Institute for Genomic Biology at University of Illinois, training new generation of integrative biologists across wet lab biology and dry lab informatics.

Trained high school students in fundamental of modern genomics generating electronic curriculum distributed on website.

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

PI Schatz is the campus leader to develop new interdisciplinary PhD program in informatics at University of Illinois, with special focus on biomedical informatics. He chaired the campus committee this year to establish the new degree. The new degree has now been approved by the Graduate College.

coPI Rodriguez-Zas has become campus chair of the Bioinformatics MS degree, succeeding coPI Zhai.

**Contributions Beyond Science and Engineering:**

BeeSpace-catalyzed genomics research being used to study agricultural crisis of honeybee disappearance (CCD Colony Collapse Disorder). We have been awarded a USDA grant to extend BeeSpace technology to predictive epidemiology of honeybee disease for both Biology and Informatics.

PI Schatz doing industrial outreach, e.g. invited big Tech Talk at Google on BeeSpace-catalyzed semantic searching in July 2007.

**Conference Proceedings****Special Requirements**

**Special reporting requirements:** None

**Change in Objectives or Scope:** None

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

**Categories for which nothing is reported:**

Any Book

Any Conference

## **BeeSpace Research and Education, 9/1/2008 – 8/31/2009**

This year was the fifth of five. This is the last year of grant funding, but we have requested an additional year of no-cost extension to finish some tasks. We made progress on all aspects of Research, both Biology and Informatics, and of Education, both training and outreach. These are outlined below. See the BeeSpace website <http://www.beespace.uiuc.edu> for more information, including software and slides, curriculum materials and workshop videos, among other digital materials.

In brief, the Informatics demonstrated the final major version of the analysis software with streamlined interface intended for widespread distribution, the Biology published the first major expression analysis from nature-nurture dissection experiments, the Education developed a new major curriculum for teaching modern genomic biology to high school students.

During the extension year, the Informatics will evolve the system for several thousand users within the biology community of arthropod researchers, in addition to demonstrating the next generation system for entity-relation extraction. During this year, the Biology will construct a master database for all the completed expression experiments and perform a meta-analysis for common patterns of social behavior across the database.

We hosted a big project workshop for the last time in May 2009, as documented by the slides and videos available at [http://www.beespace.uiuc.edu/news\\_workshop.php](http://www.beespace.uiuc.edu/news_workshop.php). The agenda for the Fifth Annual BeeSpace Project Workshop included summary talks on each part of the project, including Biology, Informatics, Education. As before, there was then an afternoon session with demonstrations of the software, this time both the production v4 and the research v5. Since this was the last workshop, we also hosted a prior afternoon of student talks, to highlight the graduating professionals in each of the parts. More than 60 scientists attended our workshop, including the project members, the remote collaborators, and a variety of invited participants interested in using the results for their own research.

### **Informatics Research**

The goal of the Informatics Research is to build BeeSpace, an interactive environment for functional analysis. More specifically, to interactively annotate gene functions using concept-based navigation of biological literature. This year, we completed our version 4, which has a greatly streamlined user interface. This redesign is based upon experience of many sessions with local biologists, making easy the most frequent interactions.

The new interface is a single screen with multiple panes so that long interactive sessions are possible, without losing the context of where you are. You can easily make your own custom space for your particular problem, by leveraging from previous spaces either standard or custom. Subspaces can be created from existing spaces by filtering via text

search or clustering via our special partitioning. The steerable cluster is a computer science research result fully integrated into our space system for concept navigation.

Different spaces can be selected, then different categorizations can be made of selected documents using the dynamic indexing. That is, the particular space within the biological literature that corresponds to your exact problem can be easily and quickly located within an interactive session. Then functional analysis can be performed against this space, this analysis is new bioinformatics research that statistically analyzes the gene functions mentioned within the literature space, unconstrained by pre-existing categories. The current functional analysis is keyed off the services Summarize and Analyze.

We previously published the Gene Summarizer, which inputs a gene name and outputs the function sentences about this gene from the space collection, classified into functional categories. This classification is based upon machine learning, with training provided by the bee lab at Illinois and the beetle lab at Kansas State that we collaborate with. This year we submitted a paper on the Genelist Analyzer, which inputs a list of genes and outputs the list of concepts from the space collection, which are most discriminating for those genes. The Analyzer is meant to interactively categorize the expression results from microarray experiments, which represent the main biology of the BeeSpace project.

The Genelist Analyzer software was revised significantly this year, to improve the precision of the resulting concepts that are most discriminating for genelist mentions against the background collection. Thus different spaces produce different results, with an example such as behavior maturation space showing the power of the system against manual curation categorization such as via the Gene Ontology. We have placed a concept clusterer into the Analyzer, which appears effective in some cases. With the integration available in version 4, the steerable clusterer can be used for more power.

The new BeeSpace Navigator v4 was demonstrated at the project workshop in May, then again at the Arthropod Genomic Symposium in June. This latter meeting in Kansas City is our major scientific community with 200 scientists attending as representatives of their laboratories. PI Schatz was again granted a plenary lecture to publicize the analysis software, this time just before the demonstrations during the poster session. We gathered many comments and much interest, for early adopters when the software is released.

The updating process for new literature is now complete and the pipeline for adding new materials in final debugging. We are releasing the software to the community in September as this report is being submitted, the final version at the end of the grant period. We expect some evolution of software over the extension year, as substantial usage begins within the community, including users of BeeBase, BeetleBase, ButterflyBase, wFleaBase, and many others.

To reach the extended scientific community, we will be distributing BeeSpace Navigator as part of the GMOD archive of open source software (Generic Model Organism Database). We have developed a smooth interface to call the Gene Summarizer from Gene Browser (GBrowse), the primary GMOD service for displayed genome sequences.

We gave a talk about this interface and our software again at the GMOD meeting in San Diego in January, and followed up with further conversations with coordinator Scott Cain at the Arthropod Genomics Symposium.

We supported 5 bioinformatics graduate students, who are co-supervised by coPI Zhai and PI Schatz. Two are graduating with PhDs this year who were leads on Genelist Analyzer (Xin He) and Gene Summarizer (Xu Ling), moving onto a postdoc at UCSF and a faculty position at Norte Dame. All together, they developed the prototype v5, which extracts entities and relations automatically from space collections, leveraging off the manually curated knowledge in FlyBase for *Drosophila* literature.

## **Biology Research**

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, using the whole genome microarray. The focus of our genomic studies into societal roles is on aggression (defense) and foraging.

This year, our experimental pipeline was in full production and the biology experiments emerging rapidly to be analyzed by the informatics. The analysis is using our newly developed Genelist Analyzer as a supplement to the standard analysis of Gene Ontology categories, found by looking up fly genes in FlyBase after finding the fly orthologs to bee genes differentially expressed. We are focusing on Foraging and Defense, as symbolic of animal behavior for food and warfare. This may enable comparisons to higher organisms, as honey bee is a model for natural behavior including field samples.

Our expression microarrays comprise nearly 20 separate experiments in nature-nurture dissection, encompassing nearly 2000 microarrays on nearly 1300 honeybees. Most of these experiments are different genetic or environmental variations on aggression or foraging. But a variety of controls are also being measured, including scout behavior (looking for food but not foraging), molecular clocks (the timing of foraging activity), drone behavior (male flying but not foraging), and so on. All of these experiments have now been run, including brain dissections and microarray analyses.

This year we published the first major analysis of our nature-nurture dissection, which appeared in Proceedings National Academy of Science. These studies were performed by postdoctoral fellow Cedric Alaux supported full-time on the BeeSpace grant and concerned social aggression. He had previously determined that alarm pheromone increased the aggression level. Both nature and nurture were independently varied. Nature was dissected by comparing normal European honey bees versus Africanized European honey bees (so called killer bees). The variation on this genetic difference in brain gene expression was measured. Nurture was dissected by comparing normal honey

bees to those artificially injected with alarm pheromone. The variation on this environmental difference in brain gene expression was measured.

Using functional analysis of gene clustering using gene ontology and regulatory region similarity using cis-regulatory computations, common clusters of genes were located. The Africanized bees were similar in gene expression to the alarmed bees. That is, the killer bee molecular profile is like they are on alarm pheromone, even when they are not. This was a striking demonstration of a commonality of mechanism across nature and across nurture, across the timeline of social evolution and the timeline of social behavior. In particular, down regulation of brain metabolism was associated with behavioral arousal. Thus, nature and nurture at the molecular level are parts of a whole rather than completely separate mechanisms.

The functional analysis of our experiments is still on-going throughout the extension year. We are taking all of the experiments as a uniform whole and using systems biology to seek common patterns. We have generated a unique dataset of 2000 arrays on the same platform and the same organism. This meta-analysis is going back to the raw spot level from the microarray to avoid losing valuable signal if the cooked gene level was analyzed. The statistics are looking for groupings across different features in the dissection experiments, with common phenotypes for gene networks. We are supporting a graduate student here for Nathan Price, just arrived as a faculty in Biomolecular Engineering after doing a postdoc under Lee Hood at the Institute for Systems Biology.

We continue anatomical localization of gene expression, using in situ hybridization of whole bee brains. This is being executed at Wake Forest University by postdoc Rodrigo Velarde under coPI Fahrbach who earned a PhD at Illinois under coPI Robinson. The genes being localized were previously identified as important neuropeptides, e.g. pigment-dispersing factor (PDF). Our former student Amy Toth was awarded a USDA fellowship as a result of her BeeSpace work and is doing a postdoc at Penn State. She published a paper in Science on her work using bee gene expression techniques to study wasp maternal behavior; our best example for the Genelist Analyzer is her use with mouse orthologs of wasp genes against a mouse collection about maternal effects.

## **Education and Outreach**

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 our research activities are available on our website. Our overall strategy is to create a unique and useful collection of educational materials on biology and informatics, which are targeted at different audiences and which are freely available on our website.

At the Graduate level, throughout the project, we have supported more than 20 students in Biology and in Informatics, from entomology and neuroscience, computer science and

animal science. 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, most particularly the Arthropod Base Consortium organized by Susan Brown at Kansas State University. Our project coordinator Jim Buell is supporting the user community and educational activities; he is also finishing his dissertation in Educational Psychology under coPI Chip Bruce, using BeeSpace as example of inquiry learning.

At the Undergraduate Level, we have developed a new Bioinformatics for Beginners course based on BeeSpace. This is a freshman honors class at Wake Forest University taught by coPI Fahrbach. She has now taught this course three times to good effect. The materials and examples during the course are derived directly from BeeSpace research, and PI Schatz and coPI Robinson give lectures during the course via videoconference.

At the High School Level, each summer we are paying the biology teacher at University Laboratory High School to develop Beespace-related materials for use in his Field Biology course. The materials developed span a wide range of biological topics, and can be found on our website. The teacher (David Stone) is distributing these to other teachers in the regional school system throughout Champaign County Illinois.

This year, we conducted a weeklong education workshop in July 2008 for fourteen University High School students. This is an expanded version of a workshop that BeeSpace piloted in summer 2007 with seven middle school-age learners. The 30-hour workshop included in-depth learning about the genomic behavioral science behind BeeSpace, presentations by BeeSpace research scientists (Robinson, Fahrbach, Schatz), visits to the Bee Research Lab, evaluation of the Bioinformatics for Beginners educational materials, and use of the BeeSpace software to investigate the timely topic of Colony Collapse Disorder. We employed a graduate student working for BeeSpace Biology (Nick Naeger) to introduce the learners to how project researchers are raising and answering questions about genetic bases for bee social behaviors, using tools like the honey bee genome microarray and RT-PCR. Morning workshop sessions will present scientific content relevant to BeeSpace biology and informatics.

The workshop represented a rare instance of high school students exposed to frontier integrative biology research. The students heard lectures on the actual beespace research efforts and performed actual miniature versions of the beespace experiments, including brain dissection and microarray analysis. The sessions were carefully recorded and then extensively edited by Stone and Buell into a fully fledged electronic curriculum. This online education is called ebeespace and available at [www.beespace.uiuc.edu/ebeespace](http://www.beespace.uiuc.edu/ebeespace) . It is a comprehensive video curriculum in bee biology and behavioral genomics. It was highlighted as the Buzz of the Week in August by the Entomological Society of America, the major professional society in this subject. During the extension year, we will continue to publish articles and present at conferences to publicize this curriculum.

## **Special Requirements**

Our funds from NSF were frontloaded, in that more money arrived in the first half of the grant 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 as the needs of the project dictate, including the approved request for a no-cost extension after the formal end of the project in August 2009. This additional year will enable us to gain a user population from the developed software in Informatics and also to complete the meta-analysis of all the microarray experiments in Biology.