Xu Ling

Department of Computer Science
University of Illinois at Urbana-Champaign
201 N. Goodwin Ave
Urbana, IL 61801
U.S.A.

Office Phone: (217) 244-4026
Office Fax: (217) 244-6500
Home Phone: (217) 721-7762
Email: `xuling@uiuc.edu`
`http://ews.uiuc.edu/~xuling/`

Research Interests

My general research interests are in the area of bioinformatics, targeting at developing principled methodology for solving computational problems related to molecular biology. I am especially interested in approaches that exploit probabilistic models, machine learning, data mining, and integrative analysis. My current focus is on comparative genomics studies, developing and applying both statistical models and data mining algorithms to problems of regulatory genomics and genome evolution. I also conducted research in mining and summarizing facts and knowledge from text collections including biomedical literature. I am becoming interested in the development and application of approaches that integrate different sources of information at different levels into a comprehensive framework to make new biological discoveries.

EDUCATION

July. 2009 (expected)	**University of Illinois at Urbana-Champaign**, Urbana, Illinois Ph.D. in Computer Science Advisor: Chengxiang Zhai, Saurabh Sinha
Aug. 2003 – Dec. 2004	**University of Illinois at Urbana-Champaign**, Urbana, Illinois Ph.D. student in Department of Electrical and Computer Engineering
July 2003	**Zhejiang University**, Hangzhou, China M.S. in Biomedical Engineering Advisor: Xiaoxiang Zheng
July 2001	**Zhejiang University**, Hangzhou, China B.E. in Biomedical Engineering

DISSERTATION RESEARCH

Title: Stochastic Evolutionary Modeling of cis-Regulatory Modules
Advisor: Saurabh Sinha, Chengxiang Zhai

In a cell, when and where genes are expressed is regulated by other genes, called *transcription factors*, and the complex networks of such regulatory interactions direct cells to perform their proper functions. The recognition of transcription factors by a target gene is through its regulatory DNA sequences, whose basic units are *transcription factor binding sites*, pieces of DNA that serve as molecular switches to turn on or off genes. Unlike coding sequences, regulatory elements do not have simple structural rules, thus it is a great challenge to identify these sequences in genomes. My dissertation research is to address this challenge through rigorous computational models. The central component of my work is a computational framework, STEMMA, that models the tendency of binding sites to form clusters using a Hidden Markov Model, and the evolution of binding sites, including both conservation and turnover among multiple species, using a stochastic evolutionary model. This framework serves as a general tool for deciphering the regulatory mechanism using cross-species comparisons. Using this framework, we are analyzing a large fruit fly gene expression dataset for constructing a global map of gene regulation related to fruit fly development. Besides more accurate decoding of the regulatory network, this work also provides a powerful manner of revealing why certain binding sites are created or destroyed by mutations, yielding new insights into the evolutionary forces that differentiate one species from the others.

OTHER RESEARCH PROJECTS

Title:	**Identification of gene clusters under evolutionary constraint**
Jan. 2006	University of Illinois at Urbana-Champaign, Urbana, Illinois
– Present	Graduate Research Assistant, Advisor: Chengxiang Zhai, Jiawei Han
	I developed both efficient data mining algorithms and a rigorous statistical evaluation method to study the gene arrangement patterns across multiple species. Using this framework, I performed extensive studies on more than one hundred bacterial genomes, made interesting findings about genome evolution, and predicted functions of many poorly characterized genes.
Title:	**Text mining in biomedical literature**
Jan. 2005	University of Illinois at Urbana-Champaign, Urbana, Illinois
– Present	Graduate Research Assistant, Advisor: Chengxiang Zhai, Bruce Schatz
	I developed software *Gene Summarizer*, that automatically generates structured summary of any gene from biomedical literature.

Title: **Identification of cis-regulatory DNA motifs associated with social behavior in honey bees**
Jun. 2004 – Oct. 2006 University of Illinois at Urbana-Champaign, Urbana, Illinois
Graduate Research Assistant, Advisor: Saurabh Sinha, Gene Robinson, Chengxiang Zhai

I conducted research on gene regulations associated with social behavior in honey bees. Interestingly, we found that transcription factors that perform nervous system-related functions in fruitfly are also likely regulating behavior-related genes. This study provides important clues to the social behavior of honey bees, and evidence to a general principle that the same set of genes may be reused to direct different developmental patterns as well as behavior in different species.

Title: **Multi-faceted text mining and summarization**
Jan. 2006 – Oct. 2008 University of Illinois at Urbana-Champaign, Urbana, Illinois
Graduate Research Assistant, Advisor: Chengxiang Zhai

I developed and applied topic modeling methods to specific text mining and retrieval tasks, such as extracting topics from blogs, and summarize/organize arbitrary topics in a text collection.

Title: **Mining search logs for query alteration**
May. 2008 – Aug. 2008 Microsoft Research, Redmond
Research Intern

I conducted original research on mining search logs for query alteration, with ultimate goal of improving Live search.

TEACHING AND PRESENTATION EXPERIENCE

Spring 2008 Guest lecture, at "*Probabilistic Methods for Biological Sequence Analysis*" (Graduate course CS598SS)
University Illinois at Urbana Champaign, Urbana, Illinois

Spring 2005 – Present Invited talks, at Bioinformatics seminar, Computer Science Department
University of Illinois at Urbana Champaign, Urbana, Illinois

Fall 2004 – Present Presentations and demos, at Beespace weekly seminar
University Illinois at Urbana Champaign, Urbana, Illinois

HONORS AND AWARDS

2006	**News report in Science Magazine (10/27/06), LA Times, Seattle Times, and other popular media** on work done on honeybee genome analysis *Genome scan for cis-regulatory DNA motifs associated with social behavior in honey bees* (published in PNAS).
2006	**Verizon Fellowship**, University of Illinois at Urbana-Champaign
2002	**First-Class Graduate Student Academic Fellowship**, Zhejiang University, China
2001	**Outstanding Graduate Student Award**, Zhejiang University, China
2001	**Mixed Class Honored Student Award**, Zhejiang University, China
2001	**Meritorious Prize in Interdisciplinary Contest in Modeling**, SIAM/NSA/INFORMS, U.S.A.
2000	**KuaiWei Academic Fellowship**, Zhejiang University, China
1999–2000	**First-Class Undergraduate Academic Fellowship**, Zhejiang University, China
1997	**First-Class Outstanding Freshman Fellowship**, Zhejiang University, China

PUBLICATIONS

1. Xu Ling, Xin He, and Saurabh Sinha, *Cis-regulatory module analysis using stochastic evolutionary model.* In preparation.

2. Xin He, Xu Ling, and Saurabh Sinha, *Alignment and prediction of regulatory sequences based on a probabilistic model of evolution.* To appear in **PLoS Computational Biology**.

3. Xu Ling, and Xin He, *Detecting gene clusters under evolutionary constraint in a large number of genomes.* To appear in **Bioinformatics**.

4. Xu Ling, Qiaozhu Mei, ChengXiang Zhai, and Bruce Schatz, *Mining multi-faceted overviews of arbitrary topics in a text collection.* In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD 2008**), Las Vegas, Aug. 2008, 497-505.

5. Xu Ling, Xin He, Dong Xin, and Jiawei Han, *Efficiently identify max-gap clusters in pairwise genome comparison.* In **Journal of Computational Biology**, July 2008, 15(6): 593-609.

6. Xu Ling, Jing Jiang, Xin He, Qiaozhu Mei, ChengXiang Zhai, and Bruce Schatz *Generating gene summaries from biomedical literature: A study of semi-structured summarization* In Information Processing and Management Pages (**IPM**), November 2007, 43(6): 1777-1791.

7. Xu Ling, Jing Jiang, Xin He, Qiaozhu Mei, Moushumi Sarma, ChengXiang Zhai, Gene Robinson, and Bruce Schatz, *Gene Summarizer: Software for automatically generating structured summaries from biomedical literature.* 2nd International Biocuration Meeting, San Jose, California, Oct. 2007.
8. Qiaozhu Mei, Xu Ling, Matthew Wondra, Hang Su, and ChengXiang Zhai, *Topic sentiment mixture: modeling facets and opinions in weblogs.* In Proceedings of the 16th International World Wide Web Conference (**WWW 2007**), Banff, Canada, May. 2007, 171-180.
9. Ying Wang, Mireia Jorda, Peter L. Jones, Ryszard Maleszka, Xu Ling, Hugh M. Robertson, Craig A. Mizzen, Miguel A. Peinado, and Gene E. Robinson, *Functional CpG methylation system in a social insect.* In **Science**, Oct. 2006, 314(5799): 578-579.
10. Saurabh Sinha, Xu Ling, Charles W. Whitfield, ChengXiang Zhai, and Gene E. Robinson, *Genome scan for cis-regulatory DNA motifs associated with social behavior in honey bees.* In Proceedings of National Academy of Sciences of the United States of America (**PNAS**), Oct. 2006, 103(44): 16352-16357.
11. The Honeybee Genome Sequencing Consortium, *Insights into social insects from the genome of the honeybee Apis mellifera.* In **Nature**, Oct. 2006, 443: 931-949.
12. Xu Ling, Jing Jiang, Xin He, Qiaozhu Mei, ChengXiang Zhai, and Bruce Schatz *Automatically generating gene summaries from biomedical literature* In Proceedings of the Pacific Symposium on Biocomputing (**PSB 2006**), Maui, Hawaii, Jan. 2006, 41-50.
13. ChengXiang Zhai, Xinghua Lu, Xu Ling, Xin He, Atulya Velivelli, Xuanhui Wang, Hui Fang and Azadeh Shakery, *UIUC/MUSC at TREC 2005 Genomics Track.* In Proceedings of the Text REtrieval Conference (**TREC 2005**), 2006.
14. Xu Ling, Jianfeng Ye, and Xiaoxiang Zheng, *Dynamic investigation of leukocyte-endothelial cell adhesion interaction under fluid shear stress in vitro.* In **Acta Biochimica et Biophysica Sinica**, Jun. 2003, 35(6): 567-572.

Professional service

External Referee: Bioinformatics, KDD, ICML, BIBE, WABI, AMIA Summit on Translational Bioinformatics

Personal Information

Gender:	Female
Citizenship:	P. R. China
Immigration status:	Permanent resident of United States

REFERENCES

Chengxiang Zhai
Associate Professor, Department of Computer Science
University of Illinois at Urbana-Champaign
Urbana, IL, 61801
(217) 244-4943
czhai@cs.uiuc.edu
`http://www.cs.uiuc.edu/homes/czhai/`

Saurabh Sinha
Assistant Professor, Department of Computer Science
University of Illinois at Urbana-Champaign
Urbana, IL, 61801
(217) 333-3233
sinhas@uiuc.edu
`http://www.cs.uiuc.edu/homes/sinhas/`

Bruce R. Schatz
Head, Department of Medical Information Science
Director, Bioinformatics Laboratory, Institute for Genomic Biology
Professor, Department of Computer Science, School of Information Science
Senior Research Scientist, National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign
Urbana, IL, 61801
(217) 244-0651
schatz@uiuc.edu
`http://www.canis.uiuc.edu/schatz.html`

Gene E. Robinson
Swanlund Chair Professor, Department of Entomology
Director, Neuroscience Program
University of Illinois at Urbana-Champaign
Urbana, IL, 61801
(217) 265-0309
generobi@uiuc.edu
`http://www.life.uiuc.edu/robinson/`

Jiawei Han
Professor, Department of Computer Science
University of Illinois at Urbana-Champaign
Urbana, IL, 61801
(217) 333-6903
hanj@uiuc.edu
`http://www.cs.uiuc.edu/homes/hanj/`

December, 2008