

# Eagu Kim

Department of Biostatistics and Medical Informatics  
University of Wisconsin-Madison  
Madison, Wisconsin 53706

`egkim@biostat.wisc.edu`  
`www.biostat.wisc.edu/~egkim`

voice 608.265.2388

fax 608.265.7916

## Research interests

Computational biology, design and analysis of algorithms, combinatorial optimization, coding theory. Specifically, biological sequence analysis, such as inverse parametric alignment and local similarity.

## Education

Ph.D. in Computer Science, University of Arizona, Tucson, Arizona, August 2008.

Dissertation: *Inverse Parametric Alignment for Accurate Biological Sequence Comparison*

Advisor: John Kececioglu, minor: Mathematics

M.S. in Computer Science, Yonsei University, Seoul, Korea, February 1998.

Thesis: *An Efficient Public-Key Cryptosystem Based on Rabin's Oneway Function*

Advisor: Jooseok Song

B.S. in Computer Science, Hanyang University, Ansan, Korea, February 1995.

## Professional experience

Research associate, Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, Wisconsin, September 2009 to present. Advisor: Colin Dewey.

Research Assistant, Department of Computer Science, University of Arizona, Tucson, Arizona, August 2001 to July 2008. Advisor: John Kececioglu.

Research Assistant, Department of Computer Science, University of Arizona, Tucson, Arizona, August 2000 to May 2001. Advisor: Stephen Pink.

Instructor, High-Technology Institute, Samsung Electronics, Kiheong, Korea, January 1998 to July 1998.

Teaching Assistant, Department of Computer Science, Yonsei University, Seoul, Korea, September 1996 to August 1997.

Research Engineer, Switching Division, LG Information and Communication (now LG Electronics), Anyang, Korea, January 1995 to July 1995.

## Research projects

*Large scale genome alignment*, September 2009 to present.

Design and implement time-efficient algorithm for large scale genome alignment: the problem of building homology forest of DNA sequences in sub-quadratic time in the length of input sequences. Currently develop algorithms for finding pairs of homologous sequences and building a forest consisting of phylogenetic trees from these pairs.

*Inverse parametric sequence alignment*, September 2005 to July 2008.

Designed and implemented efficient algorithms for inverse alignment: the problem of learning optimal parameter values for the sequence alignment objective function from examples of biological reference alignments. Developed general alignment models for protein sequences that incorporate predicted secondary structure in the examples to further improve alignment accuracy.

*Local alignment by maximum surprise*, March 2001 to August 2005.

Studied local alignment under a nonlinear similarity function, namely finding the most surprising local alignment between two strings that has the least probability of occurring by chance. Derived an accurate analytic approximation to this probability function that generalizes known special cases.

## Technical skills

C/C++ languages, Perl script, Maple, GLPK (GNU Linear Programming Kit), GMP (GNU Multi-Precision library), MFC (Microsoft Foundation Class library).

## Honors and awards

Graduate College Fellowship, College of Science, University of Arizona, Tucson, Arizona, August 1999 to July 2000.

Undergraduate Tuition Scholarship, Hanyang University, Ansan, Korea, March 1993 to February 1995.

## Teaching interests

Bioinformatics, design and analysis of algorithms, theory of computation, data structures, programming languages.

## Publications

Kim, E., T. Wheeler, and J. Kececioglu. "Learning models for aligning protein sequences with predicted secondary structure." Proceedings of the 13th *Conference on Research in Computational Molecular Biology* (RECOMB), Springer-Verlag Lecture Notes in Bioinformatics (LNBI) 5541, 512–532, 2009.

Kim, E. and J. Kececioglu. "Learning parameters for sequence alignment from examples with missing data." *IEEE/ACM Transactions on Computational Biology and Bioinformatics* (TCBB) 5:4, 546–556, 2008. (Invited paper for the TCBB special issue on WABI 2007.)

Kim, E. and J. Kececioglu. “Inverse sequence alignment from partial examples.” Proceedings of the 7th ISCB/EATCS *Workshop on Algorithms in Bioinformatics* (WABI), Springer-Verlag Lecture Notes in Bioinformatics (LNBI) 4645, 350–370, 2007.

Kececioglu, J. and E. Kim. “Simple and fast inverse alignment.” Proceedings of the 10th ACM *Conference on Research in Computational Molecular Biology* (RECOMB), Springer-Verlag Lecture Notes in Bioinformatics (LNBI) 3909, 441–455, 2006.

Nyang, D., E. Kim, and J. Song. “An interactive identification scheme based on quadratic residue problem.” *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E80-A(7), 1330–1335, 1997.

### **In preparation**

Kim, E. and J. Kececioglu. “Gila: Inverse parametric alignment via linear programming.” 2009.

### **Thesis**

Kim, E. *Inverse Parametric Alignment for Accurate Biological Sequence Comparison*. Ph.D. dissertation, Department of Computer Science, University of Arizona, Tucson, Arizona, 2008.

Kim, E. *An Efficient Public-key Cryptosystem Based on Rabin’s Oneway Function*. M.S. Thesis, Department of Computer Science, Yonsei University, Seoul, Korea, 1998.

### **Software systems**

Kim, E. and J. Kececioglu. IPA: Software for inverse parametric sequence alignment. Version 0.9, February 2008. <http://inversealign.cs.arizona.edu>

### **Oral presentations**

“Learning models for aligning protein sequences with predicted secondary structure.” Paper presentation at the 13th *Conference on Research in Computational Molecular Biology* (RECOMB), Tucson, AZ, 20 May 2009.

“Inverse parametric alignment for accurate biological sequence comparison.” Invited talk at the Computation and Informatics in Biology and Medicine (CIBM) seminar, University of Wisconsin-Madison, Madison, WI, 10 March 2009

“Learning parameters for sequence alignment from examples with missing data.” Invited talk at the Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, 12 May 2008

“Inverse sequence alignment from partial examples.” Paper presentation at the 7th ISCB/EATCS *Workshop on Algorithms in Bioinformatics* (WABI), University of Pennsylvania, Philadelphia, PA, 9 September 2007.

“Simple and fast inverse alignment.” Paper presentation at the 10th ACM *Conference on Research in Computational Molecular Biology* (RECOMB), Venice, Italy, 4 April 2006.

Computational Biology Seminar (14 lectures), Department of Computer Science, University of Arizona, Tucson, AZ, March 2001 to July 2008. <http://www.cs.arizona.edu/news/seminars.html>

## References

John Kececioglu, Associate Professor, Department of Computer Science, University of Arizona, Tucson, AZ 85721. Phone: 520.621.4526 Email: [kece@cs.arizona.edu](mailto:kece@cs.arizona.edu)

Colin Dewey, Assistant Professor, Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI 53706. Phone: 608.809.1589 Email: [cdewey@biostat.wisc.edu](mailto:cdewey@biostat.wisc.edu)

Peter Downey, Professor, Department of Computer Science, University of Arizona, Tucson, AZ 85721. Phone: 520.621.2207 Email: [pete@cs.arizona.edu](mailto:pete@cs.arizona.edu)

John Leonard, Emeritus faculty, Department of Mathematics, University of Arizona, Tucson, AZ 85721. Phone: 520.621.6874 Email: [jdb@math.arizona.edu](mailto:jdb@math.arizona.edu)

Gregory Andrews, Professor and Head, Department of Computer Science, University of Arizona, Tucson, AZ 85721. Phone: 520.621.4239 Email: [greg@cs.arizona.edu](mailto:greg@cs.arizona.edu)

## Hobbies

Modern Korean poetry (on nature and the human condition), classical music (high Baroque and late Romantic Russian composers), hiking (off the beaten path), cinema (with unexpected endings).