

KIMMEN SJÖLANDER, PH.D

473 Evans Hall #1762
Department of Bioengineering
University of California
Berkeley, CA 94720-1762

(510) 642-9932 FAX (510) 642-5835
<http://phylogenomics.berkeley.edu>
kimmen@berkeley.edu

EDUCATION

University of California, Santa Cruz Computer Science	B.A., 1993
University of California, Santa Cruz Computer Science	Ph.D., 1997

APPOINTMENTS

11/01 – Present	Assistant Professor, Dept. of Bioengineering, UC Berkeley.
12/99-11/01	Principal Scientist, Protein Informatics, Celera Genomics.
9/97-12/99	Chief Scientist, Molecular Applications Group.

MAJOR AWARDS

- Elected to Phi Beta Kappa (1993);
- National Science Foundation Three-Year Graduate Research Fellowship (1993);
- Program in Mathematics and Molecular Biology Fellowship (1996);
- University of California Presidential Dissertation-Year Fellowship (1996);
- NSF Career Award (2003);
- Presidential Early Career Award in Science and Engineering (PECASE) Award (2004).

RESEARCH INTERESTS: Method/algorithm development in computational molecular biology; molecular evolution; phylogenetic tree reconstruction; remote homolog recognition; protein structure prediction and comparative modeling; subfamily classification; prediction of critical positions in molecules; multiple sequence alignment; protein-protein interaction; pathway inference; computational prediction of protein domain structure. Applications of these algorithms to analysis of disease resistance proteins in plants and corresponding effector molecules in viruses, bacteria, and fungi; analysis of G protein coupled receptors and ion channel proteins. Investigation into eukaryotic innate immunity.

SELECTED PUBLICATIONS

Refereed Archival Journals

1. Krogh, A., Brown, M., Mian, S., *Sjölander, K.* and Haussler, D., "Hidden Markov Models in Computational Biology: Applications to Protein Modeling," *Journal of Molecular Biology* 1994 Feb 4;235(5):1501-31
2. Sakakibara, Y, Brown, M., Hughey, R., Mian, S., *Sjölander, K.*, Underwood, R., Haussler, D. "Stochastic Context-Free Grammars for tRNA Modeling," *Nucleic Acids Research*. 1994 Nov 25;22(23):5112-20.
3. *Sjölander, K.*, Karplus, K., Brown, M.P., Hughey, R., Krogh, A., Mian, I.S., Haussler D., "Dirichlet Mixtures: A Method for Improved Detection of Weak but Significant Protein Sequence Homology," *Computing Applications in the Biosciences (CABIOS)* 1996 Aug;12(4):327-45.
4. Karplus, K, *Sjölander, K.*, Barrett C, Cline M, Haussler D, Hughey R, Holm L, Sander C., "Predicting protein structure using hidden Markov models," *Proteins: Structure, Function and Genetics*, Suppl 1:134-139. 1997. Invited paper for special issue covering the second Critical Assessment for Protein Structure Prediction (CASP) competition.
5. Venter, C. et al, "The sequence of the human genome," *Science*, 2001 Feb 16;291(5507):1304-51. (KS contributions: the algorithms used for the Panther HMM library construction and functional classification of the human genome. (1) FlowerPower clustering and alignment of homologs; (2) Bayesian Evolutionary Tree Estimation and subfamily identification; (3) Subfamily HMM construction.)
6. Edgar, R., and *Sjölander, K.*, "SATCHMO: Sequence Alignment and Tree Construction using Hidden Markov models," *Bioinformatics*. 2003 Jul 22;19(11):1404-11. (Listed as a "must-read" by the Faculty of 1000.)
7. *Sjölander, K.*, "Phylogenomic inference of protein molecular function: advances and challenges," *Bioinformatics* 2004 (20)2:170-179.
8. Edgar, R., and *Sjölander, K.*, "COACH: profile-profile alignment of protein families using hidden Markov models," *Bioinformatics* 2004 May 22;20(8):1309-18.
9. Edgar, R. and *Sjölander, K.*, "A comparison of scoring functions for protein sequence profile alignment," *Bioinformatics* 2004 May 22;20(8):1301-8.
10. Mahmut Tör, M., Brown D., Cooper, A., Woods-Tör, A., *Sjölander, K.*, Jones, J.D.G. and Holub, E. "Arabidopsis Downy Mildew Resistance Gene RPP27 Encodes a Receptor-Like Protein Similar to CLAVATA2 and Tomato Cf-9," *Plant Physiology*, 2004 Jun;135(2):1100-12.
11. Kleffmann, T., Russenberger, D., von Zychlinski, A., Christopher, W., *Sjölander, K.*, Gruissem, W., and Baginsky, S. "The Arabidopsis thaliana chloroplast proteome reveals pathway abundance and novel protein functions," *Current Biology* 2004 Mar 9;14(5):354-62.
12. Magnani, E., *Sjölander, K.*, and Hake S., "From endonucleases to transcription factors: evolution of the AP2 DNA-binding domain in plants", *Plant Cell*, 2004 Sep;16(9):2265-77.
13. Rowland O, Ludwig AA, Merrick CJ, Baillieux F, Tracy FE, Durrant WE, Fritz-Laylin L, Nekrasov V, *Sjölander, K.*, Yoshioka H, Jones JD. "Functional Analysis of Avr9/Cf-9 Rapidly Elicited Genes Identifies a Protein Kinase, ACIK1, That Is Essential for Full Cf-9-Dependent Disease Resistance in Tomato", *Plant Cell*. 2004
14. Rebecca Middleton, *Kimmen Sjölander*, Nandini Krishnamurthy, Jonathan Foley, and Patricia Zambryski, "Predicted hexameric structure of the Agrobacterium VirB4 C terminus suggests VirB4 acts as a docking site during type IV secretion", *Proceedings of the National Academy of Sciences* Feb 1, 2005;102(5):1685-90.
15. Stephen T. Chisholm, Douglas Dahlbeck, Nandini Krishnamurthy, Brad Day, *Kimmen Sjölander*, and Brian J. Staskawicz, "Molecular characterization of proteolytic cleavage sites of the Pseudomonas syringae effector AvrRpt2", *Proceedings of the National Academy of Sciences*, February 8, 2005, vol. 102, no. 6, 2087-2092.

16. Nandini Krishnamurthy and *Kimmen Sjölander*, "Sequence analysis" Current Protocols in Molecular Biology (Invited unit; *in press*).
17. Nandini Krishnamurthy and *Kimmen Sjölander*, "Phylogenomic inference of protein function," Current Protocols in Bioinformatics (Invited unit; *to appear*).

Refereed Conference and Symposium Proceedings

1. Haussler, D., Krogh, A., Mian, I.S., *Sjölander, K.*, "Protein Modeling using Hidden Markov Models: Analysis of Globins", Proceedings of the Hawaii International Conference on System Sciences, 1993. Voted best in the category AI Technologies for Molecular Biology Analysis. (The first publication from UCSC on HMMs for proteins.)
2. Brown, M.P., Hughey, R., Krogh, A., Mian, I.S., *Sjölander, K.*, Haussler, D., "Using Dirichlet mixture priors to derive hidden Markov models for protein families," Proceedings of the First International Conference on Intelligent Systems for Molecular Biology 1993 1:47-55.
3. *Sjölander, K.*, "Bayesian Evolutionary Tree Estimation", Proceedings of the Eleventh International Conference on Mathematical and Computer Modelling and Scientific Computing , Computational Biology Session: "Conference Computing in the Genome Era" 1997
4. *Sjölander, K.*, "Phylogenetic inference in protein superfamilies: Analysis of SH2 domains," Proceedings of the Conference Intelligent Systems for Molecular Biology 1998 6:165-74.
5. Edgar, R., and *Sjölander, K.*, "Simultaneous sequence alignment and tree construction using hidden Markov models." Proceedings of the Pacific Symposium on Biocomputing , HI. 2003; 180-91.
6. Brown D, Krishnamurthy N, Dale J, Christopher W, and *Sjölander, K* "Subfamily HMMs in Functional Genomics", Proceedings of the Pacific Symposium on Biocomputing, 2005.

SELECTED TALKS AND TUTORIALS

1. "High-throughput phylogenomic inference of protein molecular function" Tutorial. ABRF 2005 Biomolecular Technologies: Discovery to Hypothesis, Savannah GA. Feb 7, 2005. (Invited tutorial)
2. "Methods for high-throughput phylogenomic inference of protein molecular function and structure", NCBI Aug 19, 2004. (Invited talk)
3. "Protein function prediction: getting it right." Gordon Research Conference, Bioinformatics: From Predictive Models to Inference. Oxford, UK. Aug. 2003. (Invited talk)
4. "Phylogenomic analysis of disease resistance proteins in plants", Gordon Research Conference, Plant Molecular Biology, Holderness School, Plymouth, New Hampshire, July 7-12, 2002. (Invited talk)
5. "The role of phylogenetic inference in proteomics", Plant Proteomes: Structure, Changes, Interactions and Function (Symposium), Iowa State University, June 20-23, 2002 (Invited plenary session talk)
6. "From structure to function: how bioinformatics can assist structural genomics", Special workshop on structural genomics sponsored by the NIGMS, December 7, 2001. (Invited talk)
7. "High-throughput Functional Classification of Proteins", Gordon Research Conference: Bioinformatics, From Inference to Predictive Models, Tilton, NH, August 2001 (Invited talk)
8. "Issues in Target Identification and Prioritization," Computational Genomics Conference, Baltimore, MD, November 1999 (Refereed talk)
9. "Automated domain identification in proteins using HMMs," TIGR Genome Sequencing and Analysis Conference, Miami, FL, September 1999 (Invited plenary session talk)
10. "Tutorial: Remote homolog prediction using HMMs and Threading," Mathematical Sciences Research Institute Conference, Understanding the Genome: Technological and Mathematical Challenges, Berkeley, CA, May 1998 (Refereed tutorial)
11. "Recognizing protein folds using Hidden Markov Models: A Behind-the-Scenes look at the UCSC method," Center for Molecular Biotechnology, Seattle, WA September 1997 (Invited seminar)
12. "Bayesian Evolutionary Tree Estimation," Center for Molecular Biotechnology, Seattle, WA September 1997 (Invited seminar)

13. "Inferring a protein fold using Hidden Markov models," The Second Asilomar Meeting for the Critical Assessment of Structure Prediction (CASP2), Asilomar, CA. December, 1996 (Invited talk)
14. "Tutorial: Stochastic Modeling and Hidden Markov Models," Conference: Intelligent Systems in Molecular Biology, St. Louis, MO. June, 1996. (Refereed tutorial)
15. "Dirichlet Mixture Priors for Improved Remote Homolog Recognition," European Molecular Biology Laboratory, Heidelberg, Germany, August, 1995 (Invited talk)
16. "Hidden Markov Models for Protein Sequence Analysis," LAFORIA, Université de Pierre et Marie Curie, Paris, France, July, 1995 (Invited talk)
17. "Dirichlet Mixture Priors for Hidden Markov Models and Stochastic Context-Free Grammars," Workshop on Machine Learning and Computational Biology, Conference on Learning Theory, Rutgers, New Jersey, July, 1994 (Refereed talk)

PATENTS

U.S. Patent No. 6,128,587: Method and apparatus using Bayesian subfamily identification for sequence analysis. Issued Oct. 3, 2000. Patent for algorithm for protein superfamily analysis and modeling: constructing a phylogenetic tree, identifying subfamilies, predicting critical positions, and constructing subfamily HMMs. This algorithm is being licensed (non-exclusively) to Celera Genomics, as the key technology employed in the Celera Discovery System for classification of novel genes produced during sequencing.

GRANTS

National Science Foundation Presidential Early Career Award in Science and Engineering (PECASE)
 Title: *PECASE: Investigation of Disease-Resistance Proteins in Flowering Plants.*

PI: Kimmen Sjölander (no Co-PIs)

February 1, 2003- January 31, 2008. Expected Total Amount \$698,619

National Human Genome Research Institute. RO1 Award.

Title: *High-throughput phylogenomic analysis of animal proteins.*

PI: Kimmen Sjölander (no Co-PIs)

Sept 30, 2003 – June 30, 2008. Expected Total Amount: \$1,820,265.

Grant proposals submitted (NIH RO1 and NIH P20 Center Grant, Improving the Accuracy of Comparative Modeling).

MISCELLANEOUS

1. **Curriculum development:** Developed a new course for the UC Berkeley Bioengineering Department, "Introduction to Protein Informatics." Attended by students across campus from many departments (both life sciences and engineering).
2. **Service to the scientific community outside of immediate organization:**
 - a. Chair, Gordon Research Conference "Bioinformatics: From inference to predictive models", Oxford University, August, 2003.
 - b. Conference Program Committees: Genome Informatics 1999; Intelligent Systems in Molecular Biology, 2002
 - c. External Advisory Board member: University of Texas, Austin, Bioengineering Department; University of California Santa Cruz Extension, Bioinformatics Certificate Program.
 - d. Referee: Journal of Molecular Biology, Bioinformatics, Nucleic Acids Research, PNAS, Plant Physiology, BMC Bioinformatics, Protein Science, Jnl of Theoretical Biology, FEBS Letters.
 - e. Associate Editor, IEEE Transactions in Computational Biology and Bioinformatics.