Curriculum Vitae

KIMMEN SJÖLANDER, PHD
Amsterdam, Netherlands

Email: kimmen@berkeley.edu
URL: www.kimmen.amsterdam

PERSONAL PROFILE
- US Citizen
- Netherlands resident

PROFESSIONAL PROFILE
- Accomplished scientist in data science, genomics, computational biology and bioinformatics, with expertise spanning computer science, machine learning, statistics and biology.
- Extensive experience in the development of effective undergraduate and graduate training programs, courses and curricula in Data Science and bioinformatics.
- Proven ability in writing successful grant proposals to support university research and training objectives.

SCIENTIFIC EXPERTISE
- **Machine learning algorithms** including clustering, hidden Markov models and other statistical models, information theory
- **Bioinformatics and biological applications**: whole genome annotation, protein structure prediction, automatic protein function prediction, protein-protein interaction, evolutionary tree estimation, ortholog identification, multiple sequence alignment, homology clustering, plant disease resistance, operon evolution

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

POSITIONS AND EMPLOYMENT
- **Professor Emeritus,** Department of Bioengineering and Department of Plant and Microbial Biology, University of California, Berkeley, 2018-Present
- **Professor**, Department of Bioengineering and Department of Plant and Microbial Biology, University of California, Berkeley, 2012-2017
- **Associate Professor**, Department of Bioengineering and Department of Plant and Microbial Biology, University of California, Berkeley, 2006-2012
- **Assistant Professor**, Department of Bioengineering, University of California, Berkeley, 2001-2006
- **Principal Scientist**, Protein Informatics, Celera Genomics, California, 1999-2001
- **Chief Scientist**, Molecular Applications Group, Palo Alto, California, 1997-1999

SELECTED HONORS
- 1993   Elected to Phi Beta Kappa
- 1993   National Science Foundation Three-Year Graduate Research Fellowship
- 1996   Program in Mathematics and Molecular Biology Fellowship
- 2003   National Science Foundation CAREER Award
- 2004   National Science Foundation Presidential Early Career Award in Science and Engineering (PECASE)

PEER-REVIEWED PUBLICATIONS (IN CHRONOLOGICAL ORDER)
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.
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 (ISMB) 1993 1:47-55.
3. 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.
4. 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.
5. 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.
6. 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.
7. Sjölander, K , "Phylogenetic inference in protein superfamilies: Analysis of SH2 domains," Proceedings of the Conference Intelligent Systems for Molecular Biology 1998 6:165-74.
8. Venter, C. et al (including Sjölander, K), "The sequence of the human genome," Science, 2001 Feb 16;291(5507):1304-51.
9. 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.
10. Edgar, R., and Sjölander, K., "SATCHMO: Sequence Alignment and Tree Construction using Hidden Markov models," Bioinformatics. 2003 Jul 22;19(11):1404-11. Selected by the Faculty of 1000 as a "Must Read" for Technological Advance (rating 6.0).
11. 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. Selected by the Faculty of 1000 as a "Must Read".
12. 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.
13. Edgar, R. and Sjölander, K., "A comparison of scoring functions for protein sequence profile alignment," Bioinformatics 2004 May 22;20(8):1301-8.
14. Edgar, R., and Sjölander, K., "COACH: profile-profile alignment of protein families using hidden Markov models," Bioinformatics 2004 May 22;20(8):1309-18.
15. Sjölander, K., "Phylogenomic inference of protein molecular function: advances and challenges," Bioinformatics 2004 (20)2:170-179.
16. Nandini Krishnamurthy and Kimmen Sjölander, "Basic protein sequence analysis". Current Protocols in Protein Science, Unit 2.11, 2005.
17. Nandini Krishnamurthy and Kimmen Sjölander, "Phylogenomic inference of protein molecular function," Current Protocols in Bioinformatics, Unit 6.9, 2005.
18. Nandini Krishnamurthy and Kimmen Sjölander, "Basic protein sequence analysis". Current Protocols in Molecular Biology, Unit 19.5, 2005.
19. von Zychlinski A, Kleffmann T, Krishnamurthy N, Sjölander K, Baginsky S, Gruissem W., "Proteome analysis of the rice etioplast: metabolic and regulatory networks and novel protein functions", Mol Cell Proteomics. 2005 May 20.
20. Lillian Fritz-Laylin, Nandini Krishamurthy, Mahmut Tor, Kimmen Sjölander* and Jonathan Jones. "Phylogenomic analysis of the receptor-like proteins of rice and Arabidopsis", Plant Physiology, June 2005, Vol. 138, pp. 611-623   *Corresponding author (and joint last-author).

21. 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.

22. 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 2005 Feb 1;102(5):1685-90.

23. Brown D, Krishnamurthy N, Dale J, Christopher W, and Sjölander K, "Subfamily HMMs in Functional Genomics", Proceedings of the Pacific Symposium on Biocomputing, 2005.

24. Rowland O, Ludwig AA, Merrick CJ, Baillieul 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. Jan;17(1):295-310. 2005.

25. Nandini Krishnamurthy, Duncan Brown, Dan Kirshner and Kimmen Sjölander, "PhyloFacts: An online structural phylogenomic encyclopedia for protein functional and structural classification," Genome Biology 2006, 7:R83. Noted as "Highly Accessed" by the journal.

26. Leebens-Mack J, Vision T, Brenner E, Bowers JE, Cannon S, Clement MJ, Cunningham CW, Depamphilis C, Desalle R, Doyle JJ, Eisen JA, Gu X, Harshman J, Jansen RK, Kellogg EA, Koonin EV, Mishler BD, Philippe H, Pires JC, Qiu YL, Rhee SY, Sjolander K, Soltis DE, Soltis PS, Stevenson DW, Wall K, Warnow T, Zmasek C. "Taking the First Steps towards a Standard for Reporting on Phylogenies: Minimum Information about a Phylogenetic Analysis (MIAPA)." OMICS 2006

27. Duncan Brown and Kimmen Sjölander, "Functional Classification using Phylogenomic Inference." PLoS Computational Biology, Vol 2, Issue 6, June 2006. Invited review article.

28. Samart Wanchana, Supat Thongjuea, Victor Jun Ulat, Mylah Anacleto, Ramil Mauleon, Matthieu Conte, Mathieu Rouard, Manuel Ruiz, Nandini Krishnamurthy, Kimmen Sjölander, Theo van Hintum and Richard M. Bruskiewich "The Generation Challenge Programme comparative plant stress-responsive gene catalogue," Nucleic Acids Research 2007; doi:10.1093/nar/gkm798.

29. Brown DP, Krishnamurthy N, Sjölander K, "Automated Protein Subfamily Identification and Classification," PLoS Computational Biology 2007, 3(8): e160 doi:10.1371/journal.pcbi.0030160. Selected by the Faculty of 1000 as a technological advance.

30. Jake Gunn Glanville, Dan Kirshner, Nandini Krishnamurthy and Kimmen Sjölander, "Berkeley Phylogenomics Group web servers: resources for structural phylogenomic analysis" Nucleic Acids Research Web Server Issue, 2007; doi:10.1093/nar/gkm325

31. Nandini Krishnamurthy, Duncan Brown and Kimmen Sjölander, "FlowerPower: clustering proteins into domain architecture classes for phylogenomic inference of protein function", BMC Evolutionary Biology 2007, 7 Suppl 1:S12 doi:10.1186/1471-2148-7-S1-S12.

32. Sriram Sankararaman and Kimmen Sjölander, "INTREPID - INformation-theoretic TREe traversal for Protein functional site IDentification," Bioinformatics 2008; doi: 10.1093/bioinformatics/btn474.

33. P. Gaudet, L. Lane, P. Fey, A. Bridge, S. Poux, A. Auchincloss, K. Axelsen, S. Braconi Quintaje, E. Boutet, P. Brown, E. Coudert, R.S. Datta, W.C.de Lima, T. de Oliveira Lima, S. Duvaud, N. Farriol-Mathis, S. Ferro Rojas, M. Feuermann, A. Gateau, U. Hinz, C. Hulo, J. James, S. Jimenez, F. Jungo, G. Keller, P. Lemercier, D. Lieberherr, M. Moinat, A. Nikolskaya, I. Pedruzzi, C. Rivoire, B. Roechert, M. Schneider, E. Stanley, M. Tognolli, K. Sjölander, L. Bougueleret, R.L. Chisholm, and A. Bairoch, "Collaborative annotation of genes and proteins between UniProtKB/Swiss-Prot and dictyBase," Database 2009:bap016, DOI 10.1093/database/bap016.

34. Sriram Sankararaman, Bryan Kolaczkowski and Kimmen Sjölander "INTREPID: a web server for prediction of functionally important residues by evolutionary analysis," Nucleic Acids Research 2009; doi: 10.1093/nar/gkp339

35. Ron Alterovitz, Aaron Arvey, Sriram Sankararaman, Carolina Dallett, Yoav Freund, and Kimmen Sjölander, "ResBoost: characterizing and predicting catalytic residues in enzymes," BMC

Bioinformatics 2009, 10:197doi:10.1186/1471-2105-10-197. Designated as "Highly accessed" by BMC Bioinformatics.

36. Ruchira S. Datta, Christopher Meacham, Bushra Samad, Christoph Neyer and Kimmen Sjölander "Berkeley PHOG: PhyloFacts Orthology Group Prediction Web Server," Nucleic Acids Research Web Server Issue 2009; doi: 10.1093/nar/gkp373.

37. Pieper, U., Webb, B., Barkan, D., Schneidman-Duhovny, D., Schlessinger, A., Braberg, H., Yang, Z., Meng, E., Pettersen, E., Huang, C., Datta, R., Sampathkumar, P., Madhusudhan, M.S., Sjölander, K., Ferrin, T., Burley, S., and Sali, A., "ModBase, a database of annotated comparative protein structure models, and associated resources," Nucleic Acids Research, 2010, 1–10 doi:10.1093/nar/gkq1091.

38. Jung H-S, Okegawa Y, Shih PM, Kellogg E, Abdel-Ghany SE, Pilon M, Sjölander K, Shikanai T, and Niyogi KK, "Arabidopsis thaliana PGR7 Encodes a Conserved Chloroplast Protein That Is Necessary for Efficient Photosynthetic Electron Transport," PLoS One 5(7): e11688, doi:10.1371/journal.pone.0011688.

39. Hagopian, R., Davidson, J., Datta, R., Samad, B., Jarvis, G., and Sjölander, K. "SATCHMO-JS: a webserver for simultaneous protein multiple sequence alignment and phylogenetic tree construction," Nucleic Acids Research 2010, doi:10.1093/nar/gkq298. Selected as a Featured Article by NAR. From the NAR website: "Featured Articles represent the top 5% of NAR papers in terms of originality, significance and scientific excellence."

40. Sjölander K (2010) Getting Started in Structural Phylogenomics". PLos Comput Biol 6(1): e1000621. doi:10.1371/journal.pcbi.1000621

41. Sriram Sankararaman, Fei Sha, Jack F. Kirsch, Michael I. Jordan, and Kimmen Sjölander, "Active Site Prediction using Evolutionary and Structural Information," Bioinformatics 2010; doi: 10.1093/bioinformatics/btq008.

41. Sjölander, K., Datta, R., Shen, Y., Shoffner, G., "Ortholog identification in the presence of domain architecture rearrangement," Briefings in Bioinformatics 2011; doi: 10.1093/bib/bbr036.

42. Shen, Y., Bonnot, F., Imsand, E., Rosefigure, J., Sjölander, K., Kilnman, J., "Distribution and Properties of the Genes Encoding the Biosynthesis of the Bacterial Cofactor, Pyrroloquinoline Quinone," Biochemistry 2012; doi: 10.1021/bi201763d

43. Dessimoz, C., Gabaldon, T., Roos, D., Sonnhammer, E., Herrero, J., and the Quest for Orthologs Consortium, "Toward community standards in the quest for orthologs," Bioinformatics 2012; doi: 10.1093/bioinformatics/bts050 (Members of the Quest for Orthologs Consortium: Adrian Altenhoff, Rolf Apweiler, Michael Ashburner, Judith Blake, Brigitte Boeckmann, Alan Bridge, Elspeth Bruford, Mike Cherry, Matthieu Conte, Durand Dannie, Ruchira Datta, Christophe Dessimmoz, Jean-Baka Domelevo Entfellner, Ingo Ebersberger, Toni Gabaldon, Michael Galperin, Javier Herrero, Jacob Joseph, Tina Koestler, Evgenia Kriventseva, Odile Lecompte, Jack Leunissen, Suzanna Lewis, Benjamin Linard, Michael S. Livstone, Hui-Chun Lu, Maria Martin, Raja Mazumder, David Messina, Vincent Miele, Matthieu Muffato, Guy Perriere, Marco Punta, David Roos, Mathieu Rouard, Thomas Schmitt, Fabian Schreiber, Alan Silva, Kimmen Sjölander, Nives Skunca, Erik Sonnhammer, Eleanor Stanley, Radek Szklarczyk, Paul Thomas, Ikuo Uchiyama, Michiel Van Bel, Klaas Vandepoele, Albert J. Vilella, Andrew Yates and Evgeny Zdobnov)

44. Liberles, D., Teichmann, S., Bahar, I., Bastolla, U., Bloom, J., Bornberg-Bauer, E., Colwell, L., de Koning, J., Dokholyan, N., Echave, J., Elofsson, A., Gerloff, D., Goldstein, R., Grahnen, J., Holder, M., Lakner, C., Lartillot, N., Lovell, S., Naylor, G., Perica, T., Pollock, D., Pupko, T., Regan, L., Roger, A., Rubinstein, N., Shakhnovich, E., Sjölander, K., Sunyaev, S., Teufel, A., Thorne, J., Thornton, J., Weinreich, D., Whelan, S., "The interface of protein structure, protein biophysics, and molecular evolution," Protein Science 2012; doi: 10.1002/pro.2071

45. Brian P. Anton, Yi-Chien Chang, Peter Brown, Han-Pil Choi, Lina L. Faller, Jyotsna Guleria, Zhenjun Hu, Niels Klitgord, Ami Levy-Moonshine, Almaz Maksad, Varun Mazumdar, Mark McGettrick, Lais Osmani, Revonda Pokrzywa, John Rachlin, Rajeswari Swaminathan, Benjamin Allen, Genevieve Housman, Caitlin Monahan, Krista Rochussen, Kevin Tao, Ashok S. Bhagwat, Steven E. Brenner, Linda Columbus, Valérie de Crécy-Lagard, Donald Ferguson, Alexey Fomenkov, Giovanni Gadda, Richard D. Morgan, Andrei L. Osterman, Dmitry A. Rodionov, Irina A. Rodionova, Kenneth E. Rudd, Dieter Söll, James Spain, Shuang-yong Xu, Alex Bateman, Robert M. Blumenthal, J. Martin Bollinger,

Woo-Suk Chang, Manuel Ferrer, Iddo Friedberg, Michael Y. Galperin, Julien Gobeill, Daniel Haft, John Hunt, Peter Karp, William Klimke, Carsten Krebs, Dana Macelis, Ramana Madupu, Maria J. Martin, Jeffrey H. Miller, Claire O'Donovan, Bernhard Palsson, Patrick Ruch, Aaron Setterdahl, Granger Sutton, John Tate, Alexander Yakunin, Dmitri Tchigvintsev, Germán Plata, Jie Hu, Russell Greiner, David Horn, Kimmen Sjölander, Steven L. Salzberg, Dennis Vitkup, Stanley Letovsky, Daniel Segrè, Charles DeLisi, Richard J. Roberts, Martin Steffen, Simon Kasif. "The COMBREX Project: Design, Methodology, and Initial Results," PLoS Biol 11(8): e1001638. doi:10.1371/journal.pbio.1001638
46. Afrasiabi, C., Samad, B., Dineen, D., Meacham, C. Sjölander, K., "The PhyloFacts FAT-CAT Webserver: Ortholog Identification and Function Prediction using Fast Approximate Tree Classification," Nucleic Acids Research 2013; doi: 10.1093/nar/gkt399
47. Erik Sonnhammer, Toni Gabaldón, Alan Wilter Sousa da Silva, Maria Martin,Marc Robinson-Rechavi, Brigitte Boeckmann, Paul Thomas, Christophe Dessimoz, and the Quest for Orthologs consortium. "Big Data and Other Challenges in the Quest for Orthologs," Bioinformatics (2014) doi: 10.1093/bioinformatics/btu492
48. Adrian M Altenhoff, Brigitte Boeckmann, Salvador Capella-Gutierrez, Daniel A Dalquen, Todd DeLuca, Kristoffer Forslund, Jaime Huerta-Cepas, Benjamin Linard,   Cécile Pereira, Leszek P Pryszcz, Fabian Schreiber, Alan Sousa da Silva, Damian Szklarczyk, Clément-Marie Train, Peer Bork, Odile Lecompte, Christian von Mering, Ioannis Xenarios, Kimmen Sjölander, Lars Juhl Jensen, Maria J Martin, Matthieu Muffato,, Quest for Orthologs consortium, Toni Gabaldón, Suzanna E Lewis et al. "Standardized benchmarking in the quest for orthologs" Nature Methods 13, 425–430 (2016) doi:10.1038/nmeth.3830

BOOK CHAPTERS:
1. Kimmen Sjölander and Chelsea Specht, "Functional prediction through phylogenetic inference and structural classification of proteins" Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics, John Wiley and Sons (Short Specialist Review) July 2006.
2. Nandini Krishnamurthy, Jim Leebens-Mack, Kimmen Sjölander, "Structural Phylogenomic Inference of Plant Gene Function" (Chapter 17) in The Handbook of Plant Functional Genomics: Concepts and Protocols. Edited by Dr. Günter Kahl and Dr. Khalid Meksem, Wiley Online. Published Online: 11 JUN 2008. DOI: 10.1002/9783527622542.ch17
3. Kimmen Sjölander "Phylogenomic databases and orthology prediction" (Chapter 5) in Protein Families: Relating Protein Sequence, Structure, and Function. Edited by Christine Orengo, & Alex Bateman, John Wiley & Sons (Hoboken, New Jersey) 2013.

PATENT HELD:
- Sjölander, K. (2000) U.S. Patent No. 6,128,587: Method and apparatus using Bayesian subfamily identification. (Licensed by Celera Genomics for the functional annotation of the human genome reported in Venter et al, Science 2001.)

CURRICULUM AND COURSE DEVELOPMENT
As a professor at UC Berkeley, I developed bioinformatics and data science curricula and courses for both undergraduate and graduate students in diverse programs and departments. Some of the courses were aimed at students in the life sciences while others were developed for students from engineering and maths backgrounds. Many classes combined students from both backgrounds; these were the most challenging, as the gaps in student training and understanding (as well as their interests and goals) seldom overlapped.

BIO ENG 144 Introduction to Protein Informatics (undergraduate)
BIO ENG 244 Introduction to Protein Informatics (graduate)
This course will introduce students to the bioinformatics algorithms used by biologists to identify homologs, construct multiple sequence alignments, predict protein structure, estimate phylogenetic

trees, identify orthologs, predict protein-protein interaction, and build hidden Markov models. The focus is on the algorithms used, and on the sources of various types of errors in these methods.

Course Objectives: This course is designed to provide a theoretical framework for protein sequence and structure analysis using bioinformatics software tools. Students completing this course will be prepared for subsequent in-depth studies in bioinformatics, for algorithm development, and for the use of bioinformatics methods for biological discovery. It is aimed at two populations: students in the life sciences who need to become expert users of bioinformatics tools, and students in engineering and mathematics/computer science who wish to become the developers of the next generation of bioinformatics methods. As virtually all the problems in this field are very complex, there are many opportunities for research and development of new methods.

Student Learning Outcomes: Students completing this course are likely to find several potential areas of research of interest, which they may want to work on as independent study projects during undergraduate work, or take on as Master's or Ph.D. thesis topics for advanced work.

Prerequisites: Prior coursework in algorithms. No prior coursework in biology is required. This course includes no programming projects and prior experience in programming is not required

BIO ENG 144L Protein Informatics Laboratory (undergraduate)
BIO ENG 244L Protein Informatics Laboratory (graduate)
This course is intended to provide hands-on experience with a variety of bioinformatics tools, web servers, and databases that are used to predict protein function and structure. This course will cover numerous bioinformatics tasks including: homolog detection using BLAST and PSI-BLAST, hidden Markov model construction and use, multiple sequence alignment, phylogenetic tree construction, ortholog identification, protein structure prediction, active site prediction, cellular localization, protein-protein interaction and phylogenomic analysis. Some minimal programming/scripting skills (e.g., Perl or Python) are required to complete some of the labs.

Prerequisites: One upper-division course in molecular biology or biochemistry and Python programming and experience using command-line tools in a Unix environment

BIO ENG 145 Intro to Machine Learning in Computational Biology (undergraduate)
BIO ENG 245 Intro to Machine Learning in Computational Biology (graduate)
This course will review the fundamentals of Data Science and data mining techniques. We will begin by reviewing Data Science across the disciplines, including guest lectures from data scientists on campus. As the semester progresses, we will focus increasingly on data science techniques in computational biology and bioinformatics, illustrating major methods and issues from these fields. Finally, we will discuss ethical issues related to data from biomedical research and genomics.

Course Objectives: This course aims to equip students with a foundational understanding of machine learning techniques used in genomics and computational biology. Desired Course Outcomes: Students completing this course should have stronger programming skills, the ability to apply simple machine learning techniques to complex biosequence and genomics data, and an understanding of some of the challenges in genomics and bioinformatics.

Student Learning Outcomes: Students completing this course should have stronger programming skills, the ability to apply simple machine learning techniques to complex biosequence and genomics data, and an understanding of some of the challenges in genomics and bioinformatics.

Prerequisites: Strong programming skills, and one or more courses in algorithms, data structures, statistics and/or discrete math.

GRANT SUPPORT (US FEDERAL FUNDING AGENCIES):

National Science Foundation, Presidential Early Career Award in Science and Engineering (PECASE) Program, DBI-0238311,
Title: "CAREER: Investigation of Disease-Resistance Proteins in Flowering Plants."
PI: Kimmen Sjölander.

Project Period: 2/1/2003 – 1/31/2008; Total award: $698,619

NIH, National Human Genome Research Institute R01HG02769
Title: "High throughput phylogenomic analysis of animal proteins."
PI: Kimmen Sjölander
Project Period: 09/30/2003 - 06/30/2008; Total award: $1,802,976

National Science Foundation/USDA CSREES, Microbial Genome Sequencing Program
Title: "EF-0626651 The Berkeley-TIGR Phylogenomic Encyclopedia of Microbial Protein Families."
PI: Kimmen Sjölander.
Project Period: 11/1/2006 – 10/31/2007; Total award: $239,326

National Science Foundation: Division of Molecular and Cellular Biosciences
Title "MCB-0732065 The PhyloFacts Phylogenomic Encyclopedia of Microbial Protein Families."
PI: Kimmen Sjölander
Project Period: 11/1/2008-10/31/2010; Total award: $1,899,499

Department of Energy, Division: Biological and Environmental Research
Grant number: DE-SC0004916
Title: "Phylogenomic tools and web resources for the Systems Biology Knowlegebase."
PI: Kimmen Sjölander
Project duration: 2010-2013; Total award: $1,412,313

National Science Foundation, Division: Information & Intelligent Systems, Directorate for Computer & Information Science & Engineering
Title: "EAGER: Towards a self-organizing map and hyper-dimensional information network for the human genome"
PI: Kimmen Sjölander
Grant number: #1355632
Project duration: 2013-2015; Total award: $250,000