

Adam Siepel

Assistant Professor
Department of Biological Statistics and Computational Biology
Cornell University, Ithaca, NY

CONTACT INFORMATION	101 Biotechnology Building, Cornell University, Ithaca, NY 14853 Phone: (607) 254-1157; Fax: (607) 255-4698 Email: acs4@cornell.edu; Web: http://compgen.bscb.cornell.edu/~acs
RESEARCH INTERESTS	Comparative genomics, molecular evolution, phylogeny reconstruction, population genetics, gene prediction, conserved noncoding sequences, Bayesian statistics, probabilistic graphical models
EDUCATION	University of California , Santa Cruz, CA USA Ph.D., Computer Science, 2005 Adviser: David Haussler Dissertation: "Comparative mammalian genomics: models of evolution and detection of functional elements" University of New Mexico , Albuquerque, NM USA M.S., Computer Science, December, 2001 Adviser: Bernard Moret Master's Thesis: "Exact algorithms for the reversal median problem" Cornell University , Ithaca, NY USA B.S., Agricultural and Biological Engineering, <i>cum laude</i> with distinction in research, 1994
PROFESSIONAL EXPERIENCE	Cornell University , Ithaca, NY <i>Assistant Professor of Biological Statistics and Computational Biology</i> 2006– University of California , Santa Cruz, CA <i>Graduate Student Researcher</i> 2002–2005 National Center for Genome Resources (NCGR) , Santa Fe, NM <i>Director of Information Technology (Acting)</i> 2000–2001 <i>Software Development Group Leader</i> 1997–2001 <i>Bioinformatics Programmer</i> 1996–1997 Los Alamos National Laboratory , Los Alamos, NM <i>Graduate Research Assistant, HIV Database</i> 1994–1996
HONORS AND AWARDS	Packard Fellowship, 2007 (1 of 20) Microsoft Research New Faculty Fellowship, 2007 (1 of 5) NSF CAREER Award, 2007 Named one of "Tomorrow's PIs," <i>Genome Technology</i> magazine, 2007 (1 of 30) Graduate Research and Education in Adaptive bio-Technology (GREAT) Fellowship, University of California Biotechnology Research and Education Program (UCBREP), 2004-2005 Achievement Rewards for College Scientists (ARCS) Scholarship, 2003–2004 Best Student Paper, 6th Int'l Conf. on Research in Computational Molecular Biology, 2002 UC Santa Cruz Chancellor's Fellowship, 2002 Sustained Performance Award, National Center for Genome Resources, 1999 Tau Beta Pi, National Engineering Honor Society, inducted 1993 Ho-Nun-De-Kah, Cornell University Agriculture and Life Sciences Honor Society, inducted 1993 Empire State Scholarship, 1990 Valedictorian, West Valley Central School (NY), 1990

EDITING AND
REVIEW

Editorial Board, *Genome Research* (2007–present)

Associate Editor, *PLoS Computational Biology* (2008–present)

Program Committees: Workshop on Algorithms in Bioinformatics (WABI), 2006–2008; Intelligent Systems for Molecular Biology (ISMB) / European Conference on Computational Biology (ECCB), 2004–2007; International Conference on Machine Learning (ICML), 2007; Research in Computational Molecular Biology (RECOMB), 2008–2009

NSF Panelist, Directorate for Biological Sciences (2007)

Reviewer for *Nature*, *Nature Genetics*, *PLoS Biology*, *Genome Research*, *PNAS*, *Molecular Biology and Evolution*, *PLoS Computational Biology*, *PLoS Genetics*, *Bioinformatics*, *The Plant Cell*, and other journals.

Advisory committee for sequencing of primate genomes, National Human Genome Research Institute, 2007–present

PUBLICATIONS

Journal Articles

- 1 Holloway A, Begun D, **Siepel A**, Pollard KS. Adaptive evolution drives accelerated sequence divergence of conserved genomic elements in *Drosophila melanogaster*. To appear, *Genome Res*.
- 2 Wang Y, Diehl A, Wu F, Vrebalov J, Giovannoni J, **Siepel A**, Tanksley SD. Sequencing and comparative analysis of a conserved syntenic segment in the Solanaceae. To appear, *Genetics*.
- 3 Kosiol C, Vinar T, da Fonseca RR, Hubisz MJ, Bustamante CD, Nielsen R, **Siepel A**. Patterns of positive selection in six mammalian genomes. To appear, *PLoS Genetics*.
- 4 **Siepel A**, Diekhans M, Brejova B, Langton L, Stevens M, Comstock CLG, Davis C, Ewing B, Oommen S, Lau C, Yu H-C, Li J, Roe BA, Green P, Gerhard DS, Temple G, Haussler D, Brent MR. Targeted discovery of novel human exons by comparative genomics. *Genome Res*. 17:1763-1773, 2007.
- 5 Miller W, Rosenbloom K, Hardison RC, Hou M, Taylor J, Raney B, Burhans R, King DC, Baertsch R, Blankenberg D, Kosakovsky Pond SL, Nekrutenko A, Giardine B, Harris RS, Tyekucheva S, Diekhans M, Pringle TH, Murphy WJ, Lesk A, Weinstock GM, Lindblad-Toh K, Gibbs RA, Lander ES, **Siepel A**, Haussler D, Kent WJ. 28-way vertebrate alignment and conservation track in the UCSC genome browser. *Genome Res*. 17:1797-1808, 2007.
- 6 ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447:799–816, 2007.
- 7 Margulies EH, Cooper GM, Asimenos G, Thomas DJ, Dewey CN, **Siepel A**, ..., Green ED, and Sidow A [42 authors]. Analyses of deep mammalian sequence alignments and constraint predictions for 1% of the human genome. *Genome Res* 17(6):760–774, 2007.
- 8 Kininis M, Chen BS, Diehl AG, Isaacs GD, Zhang T, **Siepel AC**, Clark AG, Kraus WL. Genomic analyses of transcription factor binding, histone acetylation, and gene expression reveal mechanistically distinct classes of estrogen-regulated promoters. *Mol Cell Biol* 27(14):5090-104, 2007.
- 9 Rhesus Macaque Genome Sequencing and Analysis Consortium. Evolutionary and biomedical insights from the rhesus macaque genome. *Science*, 13(316):222–234, 2007.
- 10 Kuhn RM, Karolchick D, Zweig AS, Trumbower H, Thomas DJ, Thakkapallayil A, Sugnet CW, Stanke M, Smith KE, **Siepel A**, Rosenbloom KR, Rhead B, Raney BJ, Pohl A, Pedersen JS, Hsu F, Hinrichs AS, Harte RA, Diekhans M, Clawson H, Bejerano G, Barber GP, Baertsch R, Haussler D, and Kent WJ. The UCSC genome browser database: update 2007. *Nucleic Acids Res* 35:D668-D673, 2007.
- 11 Pollard KS, Salama SR, King B, Kern AD, Dreszer T, Katzman S, **Siepel A**, Pedersen JS, Bejerano G, Baertsch R, Rosenbloom KR, Kent J, and Haussler D. Forces shaping the fastest evolving regions in the human genome. *PLoS Genetics*, 2(10):e168, 2006.
- 12 Pollard KS, Salama SR, Lambert N, Lambot M-A, Coppens S, Pedersen JS, Katzman S, King B, Onodera C, **Siepel A**, Kern AD, Dehay C, Igel H, Ares M, Vanderhaeghen P, and Haussler D. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* 443:167–172, 2006.
- 13 Bejerano G, Lowe CB, Ahituv N, King B, **Siepel A**, Salama SR, Rubin EM, Kent WJ, and Haussler D. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* 441:87–90, 2006.

- 14 Pedersen JK, Bejerano G, **Siepel A**, Rosenbloom K, Lindblad-Toh K, Lander ES, Kent J, Miller W, and Haussler D. Identification and classification of conserved RNA secondary structures in the human genome. *PLoS Computational Biology* 2(4):e33, 2006.
- 15 Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F, Hillman-Jackson J, Kuhn RM, Pedersen JS, Pohl A, Raney BJ, Rosenbloom KR, **Siepel A**, Smith KE, Sugnet CW, Sultan-Qurraie A, Thomas DJ, Trumbower H, Weber RJ, Weirauch M, Zweig AS, Haussler D, and Kent WJ. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res* 34:D590–598, 2006.
- 16 **Siepel A**, Bejerano G, Pedersen JS, Hinrichs A, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, Weinstock GM, Wilson RK, Gibbs RA, Kent WJ, Miller W, and Haussler D. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 15:1034–1050, 2005.
- 17 Bejerano G, **Siepel AC**, Kent WJ, and Haussler D. Computational screening of conserved genomic DNA in search of functional noncoding elements. *Nature Methods* 2(7):535–545, 2005.
- 18 **Siepel A** and Haussler D. Phylogenetic hidden Markov models. In R. Nielsen, ed., *Statistical Methods in Molecular Evolution*, Springer, 2005.
- 19 International Chicken Genome Sequencing Consortium. Sequencing and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695–716, 2004.
- 20 International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 431:931–945, 2004.
- 21 ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 306:636–640, 2004.
- 22 **Siepel A** and Haussler D. Combining phylogenetic and hidden Markov models in biosequence analysis. *J. Comput. Biol.* 11:413–428, 2004.
- 23 **Siepel A** and Haussler D. Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Mol. Biol. Evol.* 21:468–488, 2004.
- 24 Thomas JW, Touchman JW, ..., **Siepel A**, ..., and Green ED [71 authors]. Comparative analyses of multi-species sequences from targeted genomic regions. *Nature*, 424:788–793, 2003.
- 25 **Siepel A**. An algorithm to enumerate sorting reversals for signed permutations. *J. Comput. Biol.*, 10:575–597, 2003.
- 26 **Siepel A**, Farmer A, Tolopko A, Zhuang M, Mendes P, Beavis W, and Sobral B. ISYS: A decentralized, component-based approach to the integration of heterogeneous bioinformatics resources. *Bioinformatics*, 17:83–94, 2001.
- 27 **Siepel A**, Tolopko A, Farmer A, Steadman P, Schilkey F, Perry BD, and Beavis W. An integration platform for heterogeneous bioinformatics software components. *IBM Systems Journal*, 40:570–591, 2001.
- 28 Skupski MP, Booker M, Farmer A, Harpold M, Huang W, Inman J, Kiphart D, Kodira C, Root S, Schilkey F, Schwertfeger J, **Siepel A**, Stamper D, Thayer N, Thompson R, Wortman J, Zhuang JJ, Harger C. The Genome Sequence DataBase: towards an integrated functional genomics resource. *Nucleic Acids Res.* 27:35–38, 1999.
- 29 Harger C, Skupski M, Bingham J, Farmer A, Hoisie S, Hrabec P, Kiphart D, Krakowski L, McLeod M, Schwertfeger J, Seluja G, **Siepel A**, Singh G, Stamper D, Steadman P, Thayer N, Thompson R, Wargo P, Waugh M, Zhuang JJ, Schad PA. The Genome Sequence DataBase (GSDB): improving data quality and data access. *Nucleic Acids Res.* 26:21–26, 1998.
- 30 Harger C, Skupski M, Allen E, Clark C, Crowley D, Dickinson E, Easley D, Espinosa-Lujan A, Farmer A, Fields C, Flores L, Harris L, Keen G, Manning M, McLeod M, O’Neill J, Pumilia M, Reinert R, Rider D, Rohrlich J, Romero Y, Schwertfeger J, Seluja G, **Siepel A**, Schad PA. The Genome Sequence DataBase version 1.0 (GSDB): from low pass sequences to complete genomes. *Nucleic Acids Res.* 25:18–23, 1997.
- 31 **Siepel AC**, Halpern AL, Macken C, and Korber B. A computer program designed to screen rapidly for HIV Type 1 intersubtype recombinant sequences. *AIDS Res. Hum. Retroviruses* 11:1413–1416, 1995.

Refereed Conference Papers

- 32 Zhang Y, Song GT, Vinar T, Green ED, **Siepel A**, Miller W. Reconstructing the evolutionary history of complex human gene clusters. In *Proc. 12th Int'l Conf. on Research in Computational Molecular Biology*, 2008.
- 33 **Siepel A**, Pollard KS, and Haussler D. New methods for detecting lineage-specific selection. In *Proc. 10th Int'l Conf. on Research in Computational Molecular Biology*, pp. 190–205, 2006.
- 34 Jovic V, Jovic N, Meek C, Geiger D, **Siepel A**, Haussler D, and Heckerman D. Efficient approximations for learning phylogenetic HMM models from data. In *Intelligent Systems for Molecular Biology (ISMB) 2004* and *Bioinformatics* 20:i161–168, 2004.
- 35 **Siepel A** and Haussler D. Computational identification of evolutionarily conserved exons. In *Proc. 8th Int'l Conf. on Research in Computational Molecular Biology*, pp. 177–186, 2004.
- 36 **Siepel A** and Haussler D. Combining phylogenetic and hidden Markov models in biosequence analysis. In *Proc. 7th Int'l Conf. on Research in Computational Molecular Biology*, 2003.
- 37 Moret BME, **Siepel AC**, Tang J, and Liu T. Inversion medians outperform breakpoint medians in phylogeny reconstruction from gene-order data. In *Proc. 2nd Int'l Workshop on Algorithms in Bioinformatics*, Lecture Notes in Computer Science, vol. 2452, 2002.
- 38 **Siepel A**. An algorithm to find all sorting reversals. In *Proc. 6th Int'l Conf. on Research in Computational Molecular Biology*, 2002.
- 39 **Siepel AC** and Moret BME. Finding an optimal inversion median: experimental results. In *Proc. 1st Int'l Workshop on Algorithms in Bioinformatics*, Lecture Notes in Computer Science, vol. 2149, 2001.
- 40 Sobral BW, Mangalam H, **Siepel A**, Mendes P, Pecherer R, and McLaren G. Bioinformatics for rice resources. *Novartis Found Symp* 236:59–81, 2001.

TEACHING

- “Topics in Computational Genomics” (BTRY 720), Spring 2006 and 2007
- “Computational Genomics” (BTRY 484/684), Fall 2006 and 2007
- “Probabilistic Graphical Models” (BTRY 479/679; CS 4782/6782), Fall 2008
- “Comparative Genomics/Molecular Evolution,” Three-day module at Summer Institute in Statistical Genetics, Seattle, Washington, June 2008 (with Spencer Muse)

STUDENT SUPERVISION

Post-doctoral Researchers

- Bronislava Brejová, 2006–present
- Tomáš Vinař, 2006–present
- Carolin Kosiol (joint with C.D. Bustamante), 2006–present

Ph.D. Students

- Adam Diehl, 2006–present
- Andre Luis Martins, 2008–present

Undergraduate Students

- Daniel Sussman, Summer 2007–present
- Alexandra Denby, Fall 2006–Spring 2008
- Alison Marklein, Summer 2006
- Tytus Mak, Spring 2006

THESIS COMMITTEES

Current

- Xu Wang (Ph.D., Genetics and Development, qualified 2007)
- Helgi Ingólfsson (Ph.D., Tri-Institute Computational Biology and Medicine program, qualified 2007)
- Samuel Arbesman (Ph.D., Computational Biology, qualified 2007)
- Jeremiah Degenhardt (Ph.D., Computational Biology)
- Chun-Nam Yu (Ph.D., Computer Science)

Completed

- Wendy Wong (Ph.D., Biometry, 2006)

OTHER SERVICE Genomics consortia: Orangutan Sequencing and Analysis Consortium (2008–present), Rhesus Macaque Sequencing and Analysis Consortium (2006–2007), ENCyclopedia Of DNA Elements (ENCODE) Consortium (2004–2007), Mammalian Gene Collection (MGC) project (2004–present), International Human Genome Sequencing Consortium (2004), International Chicken Genome Sequencing Consortium (2003–2004)

Graduate fields at Cornell: Computational Biology, Biometry, Computer Science, and Genetics & Development (2006–present)

Committees at Cornell: Graduate student admissions committee, field of Computational Biology, 2006–2008; Undergraduate Biology Curriculum Task Force, 2007–2008; Advisory Committee for Cornell Institute for Cell and Molecular Biology, 2007–present

Developer of freely available software package called PHAST (PHYlogenetic Analysis with Space/Time models), 2003–present (~200 downloads)

Contributor to UC Santa Cruz Genome Browser (conservation, ExoniPhy, and DLESS tracks), 2003–present

RESEARCH
SUPPORT

Awarded

Project Title: “CAREER: Models of evolution and discovery of functional elements in mammals and *Drosophila*”

PI: Adam Siepel, Cornell; Source: NSF (DBI-0644111)

Amount: \$645,870; Period Covered: 03/01/07–02/29/12

Microsoft Research New Faculty Fellowship

Awarded to: Adam Siepel, Cornell

Amount: \$200,000 (lump sum award)

David and Lucile Packard Foundation Fellowship

Awarded to: Adam Siepel, Cornell

Amount: \$825,000; Period Covered: 11/01/07–10/31/12

Project Title: “Evolutionary genomics and population genetics of pathogenic streptococci”

PI: Michael Stanhope, Cornell

co-PIs: Adam Siepel, Carlos D. Bustamante, Cornell; Robert Burne, U. Florida.

Source: NIH/NIAID (1 R01 AI073368-01)

Amount: \$2,291,320; Period Covered: 07/01/08–06/30/13

Project Title: “What made us human?”

PI: Katherine S. Pollard, UC Davis; co-PI: Adam Siepel, Cornell

Source: NIH/NIGMS (1 R01 GM082901-01); Period Covered: 09/01/08–08/31/12

Completed

Project Title: “Prediction of novel human protein-coding genes using cross-species alignments and phylogenetic hidden Markov models”

PI: Adam Siepel, Cornell; Source: NIH/NCI (subcontract 22XS013A)

Subcontract of: “Development of a web interface for Mammalian Gene Collection program data” (PI: David Haussler, UC Santa Cruz)

Amount: \$105,000 [subaward only]; Period Covered: 05/15/06–05/14/07

Extended 5/15/07–11/14/07 in the amount of \$54,075.

Project Title: “Feasibility of identifying novel single exon genes by comparative genomics”

PI: Adam Siepel, Cornell; Source: NIH/NCI

Subcontract of: “Development of a web interface for Mammalian Gene Collection program data” (PI: David Haussler, UC Santa Cruz)

Amount: \$39,810 [subaward only]; Period Covered: 02/01/06–03/31/06

INVITED
CONFERENCE
LECTURES

“Targeted discovery of novel human exons by comparative genomics,” Plenary lecture for Genome-scale Pattern Analysis in the Post-ENCODE Era, ISMB 2008 Special Interest Group, Toronto, Canada, July, 2008.

“Comparative mammalian genomics: models of evolution and detection of functional elements,” Biomedical Engineering Society (BMES) annual meeting, Chicago, IL, October, 2006.

“Searching for novel functional elements in mammalian genomes by comparative genomics,” Bioinformatics

2006, Aarhus, Denmark, June, 2006.

“Lineage-specific conserved elements and evolutionary turnover in ENCODE regions,” Biology of Genomes Conference, Cold Spring Harbor, NY, May, 2006.

“Identification of functional elements in mammalian genomes by phylogenomics,” First International Conference on Phylogenomics, Sainte-Adele, Quebec, March, 2006.

“Comparative analysis of complete mammalian genomes,” Biology of Genomes Conference, Cold Spring Harbor, NY, May, 2004.

INVITED
SEMINARS AND
WORKSHOPS

“Phylogenomics of mammals: finding genes and identifying positive selection,” Cornell University, “Tree Thinking at Cornell” Symposium, Laboratory of Ornithology, Ithaca, NY, May, 2008.

“Comparative genomics of animals and plants,” University of California, Davis, Evolutionary Genetics Seminar Series, Davis, CA, May, 2008.

“Comparative genomics of animals and plants,” Stanford University, Department of Developmental Biology, Stanford, CA, May, 2008.

“Comparative genomics of animals and plants,” University of Washington, Department of Genome Sciences, Seattle, WA, March, 2008.

“Computational methods for the detection of positive and lineage-specific selection from genomic sequence data,” University of Washington, Department of Statistics, Seattle, WA, March, 2008.

“Computational methods for the detection of positive and lineage-specific selection from genomic sequence data,” Microsoft Research, Redmond, WA, March, 2008.

“Casting light on the dark matter of the human proteome,” University of Toronto, Centre for Cellular and Biomolecular Research, Toronto, Canada, October, 2007.

“Casting light on the dark matter of the human proteome,” University of Copenhagen, Center for Comparative Genomics, Copenhagen, Denmark, June, 2007.

“Computational genomics: decoding the evolution and function of the human genome,” 3rd Barbados Workshop on Computational Gene Regulation, McGill University Bellairs Research Institute, Holetown, Barbados, April, 2007.

“Use of Phylo-HMMs in motif finding and other problems,” NESCent workshop on *cis*-regulatory molecular evolution, Durham, NC, March, 2007.

“Searching for novel functional elements in mammalian genomes by comparative genomics,” Cornell University, Department of Molecular Biology and Genetics, September, 2006.

“Searching for novel functional elements in mammalian genomes by comparative genomics,” Workshop on Computational and Statistical Genomics, Banff International Research Station, Banff, Canada, July, 2006.

“Reconstructing protein-coding genes at the DNA level,” 1st Barbados Workshop on Paleogenomics, McGill University Bellairs Research Institute, Holetown, Barbados, April, 2006.

“New methods for detecting lineage-specific selection,” McGill University, Centre for Bioinformatics, March, 2006.

“New methods for detecting lineage-specific selection,” Penn State University, Center for Comparative Genomics and Bioinformatics, February, 2006.

“Identifying protein-coding genes and other functional elements by comparative genomics,” Cornell University, Department of Entomology, January, 2006.

“Comparative mammalian genomics: models of evolution and detection of functional elements,” Penn State University, Department of Biochemistry and Molecular Biology, April, 2005.

“Comparative mammalian genomics: models of evolution and detection of functional elements,” New York University, Department of Biology and Courant Institute, March, 2005.

“Comparative mammalian genomics: models of evolution and detection of functional elements,” UC Berkeley, Department of Computer Science, February, 2005.

“Comparative mammalian genomics: models of evolution and detection of functional elements,” Cornell University, Department of Computer Science and Department of Biological Statistics and Computational Biology, February, 2005.

“Comparative mammalian genomics: models of evolution and detection of functional elements,” Penn State

University, Department of Statistics, February, 2005.

“Comparative mammalian genomics: models of evolution and detection of functional elements,” University of Washington, Department of Genome Sciences, January, 2005.

“Phylogenetic HMMs for gene finding and other applications.” Stanford University, Department of Computer Science, October, 2003.