

Serafim Batzoglou

Email: serafim@cs.stanford.edu
Tel: 650-723-3334
Fax: 650-725-1449
<http://robotics.stanford.edu/~serafim>

Clark Center S266,
318 Campus Drive,
Stanford University,
Stanford, CA 94305

Research Area

Computational Biology.

Ph.D. Thesis Title

Computational Genomics: Mapping, Comparison, and Annotation of Genomes

Educational Background

2000	Ph.D., Computer Science, Massachusetts Institute of Technology Thesis Advisor: Bonnie Berger
1996	MEng., Electrical Engineering and Computer Science, Massachusetts Institute of Technology
1996	S.B. in Computer Science, Massachusetts Institute of Technology
1996	S.B. in Mathematics, Massachusetts Institute of Technology

Professional Employment

Assistant Professor, Stanford University, Department of Computer Science	Sep. 2001 – present
Research Scientist, MIT/Whitehead Institute Center for Genome Research	Apr. 2000 – Aug. 2001, Summers 1999, 1998
Research Assistant, Sandia National Laboratories	Summer 1997

Awards

2006	Best Paper Award, ISMB Conference
2004	Best Paper Award, Joint 12 th ISMB and 3 rd ECCB Conferences
2004	Sloan Foundation Fellowship
2004	NSF CAREER Award, Methods for Comparative Genomics
2003	Top 100 Young Technology Innovators, MIT's Technology Review Magazine
2001	James H. Clark Faculty Scholar, School of Engineering, Stanford University
1998	Merck/MIT Graduate Fellowship
1997	Program in Mathematics and Molecular Biology Graduate Fellowship
1995	Ranked 2 nd in class of '95 in the Mathematics major, MIT

Corporate Scientific Advisory Boards

NextBio, Los Altos, CA	July 2005 – present
23andMe, Mountain View, CA	October 2006 – present

Professional Service

Conference Organizing

Steering Committee	RECOMB Research in Computational Molecular Biology	2006 – present
Chair	RECOMB DNA Sequencing, Technologies and Computation	2002 – 2004
Organizing Committee	RECOMB Research in Computational Molecular Biology	2007
Program Committee Member	RECOMB	2002, '04, '05, '07
Session Chair and Program Committee Member	ISMB Intelligent Systems in Molecular Biology	2005
Program Committee Member and Stanford sponsor	CSB IEEE Bioinformatics Conference	2002 – 2005
Session Chair	PSB Pacific Symposium on Biocomputing	2003
Program Committee Member	SDM SIAM International Conference on Data Mining	2002
Program Committee Member	ICDM IEEE International Conference on Data Mining	2002

Other Services

Panel Member	SciDAC Computational Biology Panel	2005
Editorial Board Member	Genome Research Journal	2004–present
Panel Member (2 times)	Genomics and Bioinformatics, NIH	
Panel Member (4 times)	Biological Databases and Informatics, NSF	
Panel Member	Cancer Diagnostic and Treatment, NIH	

Teaching

Instructor	CS273 Algorithms for Structure and Motion in Biology <i>*With Leo Guibas and Jean-Claude Latombe</i> <i>**With Jean-Claude Latombe</i>	Spring 2004*, Spring 2005**
Instructor	CS161 Data Structures and Algorithms <i>with Leo Guibas</i>	Winter 2004
Instructor	CS374 Algorithms in Biology	Fall 2002* – 2006
Instructor	CS262 Computational Genomics	Spring 2002*, 2003 Winter 2004, 2005, 2006
Instructor	CS154 Automata and Complexity Theory	Winter 2002, 2003
Teaching Assistant	Distributed Algorithms, MIT Dept of EECS Instructor: Nancy Lynch	Fall 1998, 1999
Author	Solutions Manual for Discrete Algorithms, <i>with Victor Boyko</i> Morgan Kauffman	1998

*Designed the course during this quarter

Thesis Supervision

Primary Supervisor, Completed/Defended Theses

Michael Brudno	Ph.D., Computer Science	Defense: Apr. 2004	Completed: Aug. 2004
	<i>Title: Algorithms for Comparative Genomics</i>		
Yueyi Liu	Ph.D., Biomedical Informatics	Defense: Apr. 2004	Completed: Sep. 2005
Chuong B. Do	Masters, Computer Science		June 2004
	<i>Title: ProbCons: Probabilistic Consistency-based Multiple Sequence Alignment</i>		
	• Best Master's Thesis 2004, Department of Computer Science		

Primary Supervisor, Current Theses

Eugene Davydov	Ph.D., Computer Science	6 th year
Omkar Deshpande	Ph.D., Computer Science	5 th year
George Asimenos	Ph.D., Computer Science	4 th year
Jason Flannick	Ph.D., Computer Science	4 th year
Andreas Sundquist	Ph.D., Computer Science	4 th year
Antal Novak	Ph.D., Computer Science	3 rd year
Samuel Gross	Ph.D., Computer Science	3 rd year
Eugene Fratkin	Ph.D., Computer Science	3 rd year

Co-Supervisor, Completed Thesis

Balaji Srinivasan	Ph.D., Electrical Engineering,	April 2006
	Primary Advisor: Harley McAdams	
	<i>Title: "Protein Interaction Networks: Integration, Analysis, and Alignment"</i>	

Co-Supervisor, Current Thesis

Leticia Britos	Ph.D., Developmental Biology	3 rd year
Chuong B. Do	Ph.D., Computer Science,	3 rd year

Thesis Committee Chair

Kerrin Small	Genetics	<u>Thesis Defense</u> August 2006
Dave Johnson	Genetics	March 2005
Shelley J. Force Aldred	Genetics	February 2005
Daniel Rene Richards	Genetics	November 2004
Eric Stone	Statistics	April 2004
David Vengerov	Management Science and Engineering	February 2003

Thesis Reader

Gregory Cooper	Genetics	July 2006
Brian Naughton	Biomedical Informatics	May 2006
Serge Saxonov	Biomedical Informatics	February 2005
Itay Lotan	Computer Science	June 2004
Jorge Moraleda	Electrical Engineering	June 2003
Jeffrey Chang	Biomedical Informatics	April 2002
Soumya Raychaudhuri	Biomedical Informatics	March 2002

Thesis Orals

Eugene Nudelman	Computer Science	May 2005
Hoav Shacham	Computer Science	June 2005

Other Students

Research

Greg Goldgof	CURIS, Computer Science	Summer 2006
Ophelia Venturelli	Undergraduate Research, Biology	Fall 2006
Olga Russakovsky	CURIS & Masters Student	Summer 2005 – present
Marina Sirota	CURIS, Computer Science & RA	Summer 2003--Spring 2006
Daniel Woods	Research Assistant, Computer Science	Spring 2005 – Spring 2006
Serge Saxonov	Ph.D. Research, Biomedical Informatics	Fall 2003 – Spring 2006
Leticia Britos	Ph.D. Rotation, Developmental Biology	Spring 2005
Daniel Wellesley Holbert	CURIS, Computer Science	Summer 2004–Spring 2005
Stephen Dean Guo	Undergraduate Research, Computer Science	Spring 2004 – Spring 2005
Su-In Lee	Ph.D., Electrical Engineering	Winter 2002–Summer 2004
Mukund Sundararajan	Research Assistant, Computer Science	Winter – Summer 2004
Gauhar Wadhwa	Research Assistant, Computer Science	Fall 2003 – Summer 2004
Mahathi Mahabhashyam	Research Assistant, Computer Science	Fall 2003 – Spring 2004
Lucinda Southworth	Ph.D. Rotation, Biomedical Informatics	Spring 2004
Rong Xu	Research Assistant, Computer Science	Fall 2003
Aswath Manohar	Undergraduate Research, Computer Science	Summer 2003
Nick Webb	Ph.D. Rotation, Genetics	Spring 2003
Sanket Malde	Masters Research, Computer Science	Fall 2002 – Spring 2003
Allison Waugh	Ph.D. Rotation, Biomedical Informatics	Fall – Winter 2002
Mike Kim	Undergraduate Research, Computer Science	Spring – Summer 2002

Independent Study: Jinhui Pan (Summer 2002); Garima Sahai (Summer 2002); Senthil Arun Govindasamy Singaravelu (Spring 2003 – Summer 2004), Vincent Joseph Doris CS/BMI (Winter 2005)

Publications

Refereed Journal Articles

1. Phuong TM, Do CB, Edgar RC, Batzoglou S. Multiple alignment of protein sequences with repeats and rearrangements. *Nucleic Acids Research*, *in press*.
2. Naughton B, Fratkin E, Batzoglou S, Brutlag DL. A graph-based motif detection algorithm models complex nucleotide dependencies in transcription factor binding sites. *Nucleic Acids Research*, *in press*.
3. Flannick J, Novak A, Srinivasan BS, McAdams HH, Batzoglou S. Graemlin: general and robust alignment of multiple large interaction networks. *Genome Research*, 16:1169-1181, 2006.
4. Davydov E, Batzoglou S. A computational model for RNA multiple structural alignment. *Theoretical Computer Science, Special Issue on Combinatorial Pattern Matching*, *in press*.
5. Galagan JE, Calvo SE, Cuomo C, Ma L-J, Wortman J, Batzoglou S, Lee S-I, Baştürkmen M, Spevak CC, Clutterbuck J, Kapitonov V, Jurka J, Scuzzocchio C, Farman M, Butler J, Purcell S, Harris S, Braus GH, Draht O, Busch S, D'Enfert C, Bouchier C, Goldman GH, Bell-Pedersen D, Griffiths-Jones S, Doonan JH, Yu J, Vienken K, Pain A, Freitag M, Selker EU, Archer DB, Peñalva MA, Oakley BR, Momany M, Tanaka T, Kumagai T, Asai K, Machida M, Nierman WC, Denning DW, Caddick M, Hynes M, Paoletti M, Fischer R, Miller B, Dyer P, Sachs MS, Osmani SA, Birren B. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature*, 438: 1105–1115, 2005.
6. Flannick J, Batzoglou S. Using multiple alignments to improve seeded local alignment algorithms. *Nucleic Acids Research*, 33(14): 4563–4577, 2005.
7. Cooper GM, Stone EA, Asimenos G, NISC Comparative Sequencing Program, Green ED, Batzoglou S, Sidow A. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Research* 15: 901–913, 2005.
8. Do CB, Mahabhashyam MS, Brudno M, Batzoglou S. ProbCons: probabilistic consistency-based multiple sequence alignment. *Genome Research* 15:330–340, 2005.
9. The ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 306:636–640, 2004.
10. Liu Y, Wei L, Batzoglou S, Brutlag DL, Liu JS, Liu XS. A suite of web-based programs to search for transcriptional regulatory motifs. *Nucleic Acids Research* 32:W204 - W207, 2004.
11. Rat Genome Sequencing Project Consortium (RGSP). Genome sequence of the Brown Norway Rat yields insights into mammalian evolution. *Nature* 428:493–521, 2004.
12. Brudno M, Poliakov A, Salamov A, Cooper GM, Sidow A, Rubin EM, Solovyev V, Batzoglou S, Dubchak I. Automated whole-genome multiple alignment of Rat, Mouse, and Human. *Genome Research* 14:685–692, 2004.
13. Cooper GM, Brudno M, Stone ES, Dubchak I, Batzoglou S, Sidow A. Characterization of evolutionary rates and constraints in three mammalian genomes. *Genome Research* 14:539–548, 2004.
14. Liu Y, Liu XS, Wei L, Altman RB, Batzoglou S. Eukaryotic regulatory element conservation and their identification using comparative genomics. *Genome Research* 14:451-458, 2004.
15. Shan N, Couronne O, Pennacchio LA, Brudno M, Batzoglou S, Joy S, Bethel W, Rubin EM, Hamann B, Dubchak I. Phylo-VISTA: an interactive visualization tool for multiple DNA sequence alignments. *Bioinformatics* 19: 1575-1577, 2003.
16. Lee S-I, Batzoglou S. Application of independent component analysis to microarrays. *Genome Biology* 4:R76, 2003.

17. Brudno M, Chapman M, Göttgens B, Batzoglou S, Morgenstern B. Fast and sensitive multiple alignment of large genomic sequences. *BMC Bioinformatics* 4, 66, 2003.
18. Khambata-Ford S, Liu Y, Gleason C, Dickson M, Altman RB, Batzoglou S, Myers RM. Identification of promoter regions in the human genome by using a retroviral plasmid library-based functional reporter gene assay. *Genome Research* 13:1765-1774, 2003.
19. Taher L, Rinner O, Garg S, Sczyrba A, Brudno M, Batzoglou S, Morgenstern B. AgenDA: homology-based gene prediction. *Bioinformatics* 19:1575-1577, 2003.
20. Cooper GM, Brudno M, NISC Comparative Sequencing Program, Green ED, Batzoglou S, Sidow A. Quantitative estimates of sequence divergence for comparative analyses of mammalian genomes. *Genome Research* 13:813-820, 2003.
21. Brudno M, Do C, Cooper GM, Kim MF, Davydov E, NISC Comparative Sequencing Program, Green ED, Sidow A, Batzoglou S. LAGAN and Multi-LAGAN: Efficient tools for large-scale multiple alignment of genomic DNA. *Genome Research* 13: 721-731, 2003.
22. Batzoglou S, Jaffe D, Stanley K, Butler J, Gnerre S, Mauceli E, Berger B, Mesirov JP, Lander ES. ARACHNE: A whole genome shotgun assembler. *Genome Research* 12:177-189, 2002.
23. Lander ES et al. Initial sequencing and analysis of the human genome. *Nature* 409:860-921, 2001.
24. Batzoglou S, Pachter L, Mesirov JP, Berger B, Lander ES. Human and mouse gene structure: comparative analysis and application to exon prediction. *Genome Research* 10:950-958, 2000.
25. Batzoglou S, Mesirov JP, Berger B, Lander ES. Sequencing a genome by walking with clone-ends: A mathematical analysis. *Genome Research* 9:1163-1174, 1999.
26. Pachter L, Batzoglou S, Spitkovsky VI, Banks E, Lander ES, Kleitman DJ, Berger B. A dictionary based approach to gene annotation. *Journal of Computational Biology* 6:419-430, 1999.
27. Agarwala R, Batzoglou S, Dancik V, Decatur SE, Hannenhalli S, Farach M, Muthukrishnan M, Skiena S. Local rules for protein folding on a triangular lattice and generalized hydrophobicity in the HP model. *Journal of Computational Biology* 4: 275--296, 1997.

Invited Papers and Book Chapters

1. Srinivasan BS, Do CB, Batzoglou S. RECOMB 2006: evidence for intelligent (algorithm) design. *Genome Biology*, in press.
2. Edgar RC, Batzoglou S. Multiple Sequence Alignment. *Current Opinion in Structural Biology* 16: 368-373, 2006.
3. Batzoglou S, Algorithmic Challenges in Mammalian Genome Sequence Assembly. Special Review, *Encyclopedia of Genomics, Proteomics, and Bioinformatics*. Michael Dunn, Lynn Jorde, Peter Little, and Shankar Subramaniam (Editors), *in press*.
4. Batzoglou S, The many faces of sequence alignment. *Current Opinion in Bioinformatics Issue, Briefings in Bioinformatics*. Russ Altman (Editor), 6(1):6-22, 2005.
5. Batzoglou S and Pachter L. Gene Regulation, Session Introduction. *Pacific Symposium on Biocomputing* 8:3-4, 2003.

Refereed Conference Papers

1. Gross SS, Russakovsky O, Do CB, Batzoglou S. Training conditional random fields for maximum parse accuracy. *To appear in NIPS 2006*.

2. Do CB, Woods DA, Batzoglou S. CONTRAfold: RNA Secondary Structure Prediction without Physics-Based Models. *Special Issue on the Proceedings of the ISMB2006, Bioinformatics, in press. Best Paper Award.*
3. Fratkin E, Naughton B, Brutlag DL, Batzoglou S. MotifCut: Finding Regulatory Motifs with Maximum Density Subgraphs. *Special Issue on the Proceedings of the ISMB2006, Bioinformatics, in press.*
4. Do CB, Gross SS, Batzoglou S. CONTRAlign: Discriminative Training for Protein Sequence Alignment. *Proceedings of the Tenth Annual International Conference on Computational Molecular Biology, (RECOMB 2006), pp. 160–164.*
5. Srinivasan B, Novak A, Flannick J, Batzoglou S, McAdams H. Integrated Protein Interaction Networks for 11 Microbes. *Proceedings of the Tenth Annual International Conference on Computational Molecular Biology, (RECOMB 2006), pp. 1–14.*
6. Manohar A, Batzoglou S. TreeRefiner: a tool for refining a multiple alignment on a phylogenetic tree. *CSB 2005, pp. 111-119.*
7. Sundararajan M, Brudno M, Small K, Sidow A, Batzoglou S. Chaining algorithms for alignment of draft sequence. *Proceedings of the 4th Workshop on Algorithms in Bioinformatics (WABI2004).*
8. Do CB, Brudno M, Batzoglou S. ProbCons: Probabilistic consistency-based multiple alignment of amino acid sequences. *Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI2004).*
9. Davydov E, Batzoglou S. A computational model for RNA multiple structural alignment. *Combinatorial Pattern Matching 2004.*
10. Lee S-I, Batzoglou S. ICA-based clustering of genes from microarray expression data. In *Advances in Neural Information Processing Systems (NIPS)*, Vancouver, British Columbia, Canada, December 2003.
11. Brudno M, Malde S, Poliakov A, Do C, Couronne O, Dubchak I, Batzoglou S. Glocal alignment: finding rearrangements during alignment. *Special Issue on the Proceedings of the ISMB 2003, Bioinformatics 19: 54i-62i, 2003.*
12. Batzoglou S, Pachter L, Mesirov JP, Berger B, Lander ES. Human and mouse gene structure: comparative analysis and application to exon prediction. *Proceedings of the Fourth Annual International Conference on Computational Molecular Biology, (RECOMB 2000) p.46-53.*
13. Batzoglou S, Mesirov JP, Berger B, Lander ES. Sequencing a genome by walking with clone-ends: A mathematical analysis. *Proceedings of the Fourth Annual International Conference on Computational Molecular Biology, (RECOMB 2000) p.45.*
14. Pachter L, Batzoglou S, Spitkovsky VI, Beebe W, Lander ES, Berger B, Kleitman DJ. A dictionary based approach to gene annotation. *Proceedings of the Third Annual International Conference on Computational Molecular Biology, (RECOMB 1999), 285-294.*
15. Batzoglou S, Istrail S. Physical mapping with repeated probes: The hypergraph superstring problem. *Lecture Notes in Computer Science 1645:66 ff, 1999, Special Issue on Combinatorial Pattern Matching 1999.*
16. Batzoglou S, Berger B, Kleitman DJ, Lander ES, Pachter L. Recent developments in computational gene recognition. *Documenta Mathematica, Extra Volume ICM I, 649-658, 1998.*
17. Agarwala R, Batzoglou S, Dancik V, Decatur SE, Hannenhalli S, Farach M, Muthukrishnan S, Skiena S. Local rules for protein folding on a triangular lattice and generalized hydrophobicity in the HP model. *Proceedings of the 8th Annual ACM-SIAM Symposium on Discrete Algorithms, (SODA '97) 390-399.*
18. Agarwala R, Batzoglou S, Dancik V, Decatur SE, Hannenhalli S, Farach M, Muthukrishnan S, Skiena S. Local rules for protein folding on a triangular lattice and generalized hydrophobicity in the HP model.

Proceedings of the First Annual International Conference on Computational Molecular Biology, (RECOMB 1997) 1-2.

19. Decatur S, Batzoglou S. Protein folding in the hydrophobic-polar model on the 3D triangular lattice. *Proceedings of the 6th Annual MIT Student Workshop on Computing Technology*, 1997.

Conference Contributed Talk Abstracts

1. Novak AF, Flannick JA, Srinivasan B, McAdams HH, Batzoglou S. NUKE: fast and scalable multiple alignment of protein interaction networks. *CSHL Conference Genome Informatics, October 28-November 1, 2005, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2005.*
2. Srinivasan B, Novak AF, Flannick JA, Batzoglou S, McAdams HH. Integrated protein interaction networks for 230 microbes. In *BCATS 2005 Symposium Proceedings*, p. 26, 2005.
3. Fratkin E, Naughton B, Brutlag D, Batzoglou S. Motif finding in DNA sequences using maximum density subgraphs. *2nd Moscow Conference in Computational Molecular Biology (MCCMB)*, 2005.
4. Asimenos G, Cooper GM, Holbert D, Sidow A, Batzoglou S. A reference mammalian whole-genome alignment. *The Biology of Genomes, CSHL*, May 2005.
5. Dubchak I, Brudno M, Poliakov A, Kislyuk A, Sundararajan M, Batzoglou S. Glocal (global/local) alignment methods for comparison of DNA sequences and whole genome assemblies utilized in VISTA tools. *2nd RECOMB Comparative Genomics Satellite Workshop*, 16–19 October, Bertinoro (Forli), Italy, 2004.
6. Liu Y, Batzoglou S, Kim SK. Global identification of *Caenorhabditis elegans* regulatory motifs. In *BCATS 2004 Symposium Proceedings*, p. 26, 2004.
7. Do CB, Brudno M, Batzoglou S. ProbCons: Probabilistic consistency-based multiple alignment of amino acid sequences. *Intelligent Systems in Molecular Biology and European Conference on Computational Biology (ISMB/ECCB) 2004. Best Paper Award.*
8. Do CB, Brudno M, Batzoglou S. Probabilistic consistency-based multiple alignment of proteins. In *BCATS 2003 Symposium Proceedings*, 2003.
9. Liu Y, Liu XS, Stuart JM, Kim SK, Batzoglou S. Predicting the Activity of Transcription Factor Binding Motifs. In *BCATS 2003 Symposium Proceedings*, 2003.
10. Brudno M, Malde S, Poliakov A, Do CB, Couronne O, Dubchak I, Batzoglou S. Glocal alignment: finding rearrangements during alignment. *Joint CSHL/Wellcome Trust Conference Genome Informatics May 7-11 Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2003.*
11. Brudno M, Poliakov A, Couronne O, Do CB, Batzoglou S, Dubchak I. Multiple alignment of whole genomes: a pipeline approach. *Joint CSHL/Wellcome Trust Conference Genome Informatics May 7-11 Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2003.*
12. Brudno M, Do CB, Cooper GM, Kim M, Davydov E. NISC Comparative Sequencing Program, Green ED, Sidow A, Batzoglou S. Multiple genomic sequence alignment. *Advances in Genome Biology and Technology (AGBT) Marco Island, Florida, February 2003.*
13. Brudno M, Do CB, Kim M, Batzoglou S. Multiple genomic sequence alignment. In *BCATS 2002 Symposium Proceedings*, p.23, 2002.
14. Brudno M, Kim M, Batzoglou S. Multiple alignment of genomic sequences. *Joint CSHL/Wellcome Trust Conference Genome Informatics September 4–8 p.6.* Wellcome Trust Genome Campus, Hinxton, United Kingdom, 2002.
15. Couronne O, Bray N, Khatib F, Dubchak I, Batzoglou S, Pachter L. Comparative whole genome shotgun assembly. *Abstracts of papers presented at the Joint CSHL/Wellcome Trust Conference Genome Sequencing and Biology May 7–11 p.63, p.123.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2002.
16. Jaffe D, Batzoglou B, Stanley K, Butler J, Gnerre S, Mauceli E, Berger B, Mesirov JP, Lander ES. ARACHNE: A whole genome shotgun assembler. *Abstracts of papers presented at the Joint CSHL/Wellcome Trust Conference Genome Informatics August 8–12 p.10.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2001.

17. Batzoglou S, Jaffe D, Stanley K, Berger B, Mesirov JP, Lander ES. ARACHNE: A whole genome shotgun assembler. *Abstracts of papers presented at the 2001 meeting on Genome Sequencing and Biology* May 9–13 p.199. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2001.
18. Batzoglou S, Jaffe D, Stanley K, Berger B, Mesirov JP, Lander ES. ARACHNE: A whole-genome shotgun assembler. *Advances in Genome Biology and Technology (AGBT)* Marco Island, Florida, February 2001.
19. Batzoglou S., Istrail S. Physical mapping with repeated probes: the hypergraph superstring problem. In Ninth SIAM Conference on Discrete Mathematics, Toronto, Canada, July 1998.

Conference Poster Abstracts

1. Gross SS, Do CB, Batzoglou S. De novo gene prediction using a semi-Markov conditional random field. RECOMB 2006.
2. Do CB, Woods DA, Batzoglou S. CONTRAfold: RNA secondary structure prediction without physics-based models. RECOMB 2006.
3. Flannick J, Batzoglou S. Using multiple alignments to improve seeded local alignment algorithms. *CSHL Conference Genome Informatics, October 28-November 1, 2005, Cold Spring Harbor Laboratory*, Cold Spring Harbor, New York, 2005.
4. Srinivasan B, Novak AF, Flannick JA, Batzoglou S, McAdams HH. Integrated protein interaction networks for 230 microbes. *CSHL Conference Genome Informatics, October 28-November 1, 2005, Cold Spring Harbor Laboratory*, Cold Spring Harbor, New York, 2005. **Best Poster Award.**
5. Sundquist A, Ronaghi M, Tang H, Pevzner P, Batzoglou S. A strategy for whole-genome sequencing and assembly with high-throughput, short-read technologies. *CSHL Conference Genome Informatics, October 28-November 1, 2005, Cold Spring Harbor Laboratory*, Cold Spring Harbor, New York, 2005.
6. Do CB, Gross SS, Edgar RC, Batzoglou S. CONTRAlign: a discriminative framework for protein sequence alignment. In *BCATS 2005 Symposium Proceedings*, p. 45, 2005. **Best Poster Award.**
7. Gross SS, Do CB, Batzoglou S. CONTRAST: de novo gene prediction using a semi-Markov conditional random field. In *BCATS 2005 Symposium Proceedings*, p. 82, 2005.
8. Sundquist A, Ronaghi M, Tang H, Pevzner P, Batzoglou S. A strategy for whole-genome sequencing and assembly with high-throughput, short-read technologies. In *BCATS 2005 Symposium Proceedings*, p. 38, 2005.
9. Cooper GM, Stone EA, Asimenos G, NISC Comparative Sequencing Program, Green ED, Batzoglou S, Sidow A. Characterization of the effects of purifying selection in a sample of the human genome. *The Biology of Genomes*, CSHL, May 2005.
10. Davydov E, Batzoglou S. A computational model for RNA multiple structural alignment. In *BCATS 2004 Symposium Proceedings*, p. 42, 2004.
11. Brudno M, Sundararajan M, Poliakov A, Kislyuk A, Dubchak I, Batzoglou S. Constructing synteny maps for whole-genome alignments. *Joint CSHL/Wellcome Trust Conference Genome Informatics* September 22–26, 2004.
12. Do CB, Brudno M, Batzoglou S. LAGAN2: Probabilistic global alignment of DNA under multiple conservation models. *Intelligent Systems in Molecular Biology (ISMB)* 2003.
13. Lee S, Batzoglou S. Discovering biological processes from microarray data using independent component analysis. *Joint CSHL/Wellcome Trust Conference Genome Informatics* May 7-11 Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2003.
14. Do C, Brudno M, Batzoglou S. The draft problem: genomic sequence alignment reconsidered. In *BCATS 2002 Symposium Proceedings*, p.41, 2002.

15. Taher L, Rinner O, Garg S, Brudno M, Batzoglou S, Morgenstern B. AGenDA: A WWW server for gene recognition by comparative sequence analysis. In *Thomas Lengauer, Hans-Peter Lenhof, Ruth Christmann (editors) European Conference on Computational Biology 2002, Poster Abstracts* pp. 236–238, 2002.
16. Vinson J, Jaffe D, Stange-Thomann N, Galagan J, Batzoglou S, Nusbaum C, Birren B, Zody M, Mesirov J, Sidow A, Lander ES. Highly polymorphic genomes: a challenge for assembly and an opportunity for comparative genomics. *Joint CSHL/Wellcome Trust Conference Genome Informatics* September 4–8 p.86. Wellcome Trust Genome Campus, Hinxton, United Kingdom, 2002.
17. Vinson J, Jaffe D, Sidow A, Batzoglou S, Butler J, Nusbaum C, Birren B, Stange-Thomann N, Zody M, Mesirov J, and Lander ES. Assembly of the highly polymorphic genome of *Ciona savignyi*. *Abstracts of papers presented at the Joint CSHL/Wellcome Trust Conference Genome Sequencing and Biology* May 7–11 p.291. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2002.
18. Nusbaum C, Endrizzi M, Calvo S, Foley K, Stange-Thomann N, Sachs M, Kinsey J, Staben C, Jaffe D, Batzoglou S, Galagan J, and Birren B. Sequencing the Neurospora Genome. *Abstracts of papers presented at the 2001 meeting on Genome Sequencing and Biology* May 9-13 p.182. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2001.

Other Manuscripts

1. Serafim Batzoglou. Computational Genomics: Mapping, Comparison, and Annotation of Genomes. Ph.D. Dissertation, Department of Electrical Engineering and Computer Science, MIT, June 2000.
2. Istrail S, Hurd A, Lippert RA, Walenz B, Batzoglou S, Conway JH, Peyerl FW. Prediction of Self-Assembly of Energetic Tiles and Dominos: Experiments, Mathematics and Software. *Sandia Labs Technical Report*, March 2000.
3. Serafim Batzoglou. DNA Computing and Molecular Self-Assembly. Area examination, Department of Electrical Engineering and Computer Science, MIT, December 1999.
4. Serafim Batzoglou and Victor Boyko. Discrete Algorithms: a Solutions Manual. Morgan Kauffman, 1998.

Selected Invited Presentations

Invited Seminars

Models and algorithms for Genomic Sequences, Proteins, and Networks of Protein Interactions	Princeton University	Nov. 2006
	Washington University	Nov. 2006
	Carnegie Mellon University	Jan. 2007
	MIT	May 2006
	Broad Institute	May 2006
Aligning Protein Sequences and Interaction Networks	Cornell University	Nov. 2005
	Microsoft Research	Sep. 2005
Some Algorithms for Assembly and Alignment	University College of London	July 2005
	3 rd Bertinoro CompBio Meeting	June 2005
Algorithms for Multiple Alignment of Genomes and proteins	Carnegie Mellon University	Oct. 2004
	University of Southern California	Apr. 2004
	University of Pennsylvania	Sep. 2003

Genomics and Computational Biology Seminar Series, UC Berkeley	Alignments, Motifs, and Microarrays	Apr. 2003
MIT Mathematics/LCS Bioinformatics Seminar Series, MIT	Alignments, Motifs, and Microarrays	Mar. 2003
Biostatistics Department Seminar Series Biostatistics Department, UCSF	Alignments, Motifs, and Microarrays	Mar. 2003
Broad Area Colloquium, Dept. of Computer Science, Stanford	Arachne: a whole genome shotgun assembler	Oct. 2001
Seminar Series in Bioinformatics, Boston University	Human and mouse gene structure: comparative analysis and gene prediction	Jan. 2001

Invited Talks

Algorithmic Biology, UCSD	Models and Algorithms for Biosequences and Networks of Protein Interactions	Nov. 2006
ORFeome Meeting: ORFeomes and Systems	Models and Algorithms for Biosequences and Networks of Protein Interactions	Nov. 2006
Joint Genome Institute, Host: Inna Dubchak	Sequencing, Assembly, and Alignment	Nov. 2004
Stanford/UC Berkeley Computer Science Day	Multiple alignment of genomic sequences	Feb. 2003
Advances in Genome Biology and Technology Conference, Marco Island, Florida	Multiple alignment of genomic sequences	Feb. 2003
Whitehead Institute/MIT Center for Genome Research, Cambridge, MA	Multiple alignment of genomic sequences	May 2002
Computer Forum Annual Meeting Stanford University	Computational Genomics	Mar. 2002
Genetics Lecture Series Stanford University	Computational Genomics	Feb. 2002
8 th Annual Research Directors Conference, MIT Series on Technology and Corporation	Human and mouse gene structure: comparative analysis and gene prediction	Apr. 2001
Advances in Genome Biology and Technology Conference, Marco Island, Florida	Arachne: a whole genome shotgun assembler	Feb. 2001
DIMACS Workshop on whole genome comparisons, DIMACS Center Sanger Center, Cambridge, UK	Human and mouse gene structure: comparative analysis and gene prediction	Feb. 2001
Sanger Center, Cambridge, UK	Arachne: a whole genome shotgun assembler	Oct. 2000
Sanger Center, Cambridge, UK	Human and mouse gene structure: comparative analysis and gene prediction	Oct. 2000
MIT Department of Brain and Cognitive Science	Human and mouse gene structure: comparative analysis and gene prediction	Oct. 2000