

Steven L. Salzberg

McKusick-Nathans Institute of Genetic Medicine
Johns Hopkins School of Medicine
733 North Broadway, MRB 459, Baltimore, MD 20742
Phone: 410-614-6112 Email: salzberg@jhu.edu

Education

Ph.D. Computer Science 1989, Harvard University, Cambridge, MA
M.Phil. 1984, M.S. 1982, Computer Science, Yale University, New Haven, CT
B.A. *cum laude* English 1980, Yale University

Research Areas: Genomics, bioinformatics, genome assembly, gene finding, sequence analysis algorithms.

Academic and Professional Experience

2011-present	Professor, Department of Medicine and the McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University. Joint appointments as Professor in the Department of Biostatistics, Bloomberg School of Public Health, and in the Department of Computer Science, Whiting School of Engineering.
2012-present	Director, Center for Computational Biology, Johns Hopkins University.
2005-2011	Director, Center for Bioinformatics and Computational Biology, University of Maryland Institute for Advanced Computer Studies
2005-2011	Horvitz Professor, Department of Computer Science, University of Maryland.
1997-2005	Senior Director of Bioinformatics (2000-2005), Director of Bioinformatics (1998-2000), and Investigator (1997-2005), The Institute for Genomic Research (TIGR).
1999-2006	Research Professor, Departments of Computer Science and Biology, Johns Hopkins University
1989-1999	Associate Professor (1996-1999), Assistant Professor (1989-1996), Department of Computer Science, Johns Hopkins University. On leave 1997-99.
1988-1989	Associate in Research, Graduate School of Business Administration, Harvard University. Consultant to Ford Motor Co. of Europe and to N.V. Bekaert (Kortrijk, Belgium).
1985-1987	Research Scientist and Senior Knowledge Engineer, Applied Expert Systems, Inc., Cambridge, MA. Designed expert systems for financial services companies.

Awards, honors, and editorial boards

- Winner, Benjamin Franklin Award for Open Access in the Life Sciences, 2013
- Fellow, American Association for the Advancement of Science, elected 2004
- Fellow, International Society for Computational Biology, elected 2013
- Winner, Robert G. Balles Prize in Critical Thinking, 2013 (for my *Forbes* science column)
- Board of Scientific Counselors, National Center for Biotechnology Information, NIH, 2003-2008
- Member, International Advisory Committee, International Nucleotide Sequence Database Collaboration (comprising GenBank, EMBL, and DDBJ), 2005-present
- Member, Biodata Management and Analysis Study Section, NIH, 2007-2011
- Co-winner (with Ben Langmead, Cole Trapnell, and Mihai Pop) of *Genome Biology* Award for Best Research Paper of 2009
- Named as one of BioMed Central's "Hot 100 authors," December 2007
- Chosen for ISIHighlyCited.com, an online directory of the most highly cited researchers of the past 20 years

- Ranked among scientists for “The Hottest Research of 1999-2000,” based on number of highly-cited papers (tied for 2nd in list of 13), by the Institute for Scientific Information.
- NIH Career Award, 1996-1998
- 2009-present: Editorial Advisory Board, *Genome Biology*, BioMed Central Publishers
- 2009-present: Editorial Board, *Genome Research*, Cold Spring Harbor Laboratory Press
- 2003–present: Editorial Board, *BMC Biology*, BioMed Central Publishers
- 2008-present: Editorial Board, *Biology Direct*, BioMed Central Publishers
- 2008-present: Member, *Faculty of 1000*, Bioinformatics section
- 2002–present: Associate Editor, *Journal of Computational Biology*, Mary Ann Liebert Publishers
- 2008-2012: Board of Reviewers, *PLoS Currents Influenza*, Public Library of Science
- 2006-present: Editorial Board, *PLoS ONE*, Public Library of Science
- 2001–2006: Associate Editor, *Bioinformatics*, Oxford University Press
- 2003–present: Editorial Board, *BMC Genomics*, BioMed Central Publishers
- 2003–present: Editorial Board, *BMC Bioinformatics*, BioMed Central Publishers
- 2005-present: Editorial Board, *Evolutionary Bioinformatics*, Libertas Academica Ltd
- 1999–2003: Editorial Board, *Gene*, Elsevier Science Publishers.
- 2002–2006: Editorial Board, *Applied Bioinformatics*, Open Mind Journals Ltd
- 1997–2000: Editorial Board, *Pattern Analysis and Applications*, Springer-Verlag Publishers
- 1995–2002: Editorial Board, *Artificial Intelligence Review*, Kluwer Academic Publishers

Articles in Refereed Journals and Book Chapters (h-index: 107)

1. D.E. Wood and S.L. Salzberg. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* **15**:R46 (2014).
2. D.B. Neale, J.L. Wegrzyn, K.A. Stevens, A. Zimin, D. Puiu, M. Crepeau, C. Cardeno, M. Koriabine, A.E. Holtz-Morris, J.D. Liechty, P.J. Martínez-García, H.A. Vasquez-Gross, B.Y. Lin, J.J. Zieve, W.M. Dougherty, S. Fuentes-Soriano, L. Wu, D. Gilbert, G. Marçais, M. Roberts, C. Holt, M. Yandell, J.M. Davis, K.E. Smith, J.F.D. Dean, W.W. Lorenz, R.W. Whetten, R. Sederoff, N. Wheeler, P.E. McGuire, D. Main, C.A. Loopstra, K. Mockaitis, P.J. deJong, J.A. Yorke, S.L. Salzberg, and C.H. Langley. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology* **15**:R59 (2014). <http://genomebiology.com/2014/15/3/R59>
3. A. Zimin, K.A. Stevens, M. Crepeau, A. Holtz-Morris, M. Koriabine, G. Marçais, D. Puiu, M. Roberts, J.L. Wegrzyn, P.J. de Jong, D.B. Neale, S.L. Salzberg, J.A. Yorke, and C.H. Langley. Sequencing and assembly of the 22-Gb loblolly pine genome. *Genetics* **196**:875-890 (2014).
4. S.L. Salzberg, M. Perteau, J.A. Fahrner, and N. Sobreira. DIAMUND: Direct comparison of genomes to detect mutations. *Human Mutation* **35** (2014), 283-288.
5. V.G. Martinson, T. Magoc, H. Koch, S.L. Salzberg, and N.A. Moran. Genomic features of a bumble bee symbiont reflect its host environment. *Applied and Environmental Microbiology*, to appear (2014).
6. J.L. Wegrzyn, J.D. Liechty, K.A. Stevens, L.-S. Wu, C.A. Loopstra, H. Vasquez-Gross, W.M. Dougherty, B.Y. Lin, J.J. Zieve, P.J. Martínez-García, C. Holt, M. Yandell, A. Zimin, J.A. Yorke, M. Crepeau, D. Puiu, S.L. Salzberg, P. de Jong, K. Mockaitis, D. Main, C.H. Langley, D.B. Neale. Unique features of the loblolly pine (*Pinus taeda* L.) megagenome revealed through sequence annotation. *Genetics* **196**:891-909 (2014).
7. D. Kim, G. Perteau, C. Trapnell, H. Pimentel, R. Kelley, and S.L. Salzberg. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**:R36 (2013).
8. A. Zimin, G. Marçais, D. Puiu, M. Roberts, S.L. Salzberg, and J.A. Yorke. The MaSuRCA genome assembler. *Bioinformatics* **29**:21 (2013), 2669-2677.
9. L.D. Florea and S.L. Salzberg. Genome-guided transcriptome assembly in the age of next-generation sequencing. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **10**:5 (2013), 1234-1240.

10. B. Anton, Y.-C. Chang, P. Brown, L.L. Faller, ...[...], S.L. Salzberg, D. Vitkup, S. Letovsky, D Segre, C. DeLisi, R.J. Roberts, M. Steffen, and S. Kasif. The COMBREX project: design, methodology, and initial results. *PLoS Biology* 11:8 (2013): e1001638.
11. L. Florea, L. Song, and S.L. Salzberg. Thousands of exon skipping events differentiate among splicing patterns in sixteen human tissues. *F1000Research* 2013, **2**:188 (doi: 10.12688/f1000research.2-188.v1).
12. T. Magoc, S. Pabinger, S. Canzar, X. Liu, Q. Su, D. Puiu, L.J. Tallon, and S.L. Salzberg. GAGE-B: An Evaluation of Genome Assemblers for Bacterial Organisms. *Bioinformatics* **29**:14 (2013), 1718-1725.
13. J.T. Leek and S.L. Salzberg. Sequestration: inadvertently killing biomedical research to score political points. *Genome Biology* 2013, **14**:109.
14. J.L. Wegrzyn, B.Y. Lin, J.J. Zieve, W.M. Dougherty, P.J. Martínez-García, M. Koriabine, A. Holtz-Morris, P. deJong, M. Crepeau, C.H. Langley, D.Puiu, S.L. Salzberg, D.B. Neale, K.A. Stevens. Insights into the Loblolly Pine Genome: Characterization of BAC and Fosmid Sequences. *PLoS ONE*, 8(9): e72439 (2013).
15. T. Magoc, D. Wood, and S.L. Salzberg. EDGE-pro: Estimated Degree of Gene Expression in Bacterial Genomes. *Evolutionary Bioinformatics* 9 (2013) 127-136.
16. B. Langmead and S.L. Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9 (2012), 357-359.
17. The Heliconius Genome Consortium. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**, 94–98 (2012). doi:10.1038/nature11041.
18. S.L. Salzberg. The perils of gene patents. *Clinical Pharmacology & Therapeutics* 91:6 (2012), 969-971.
19. C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D.R. Kelley, H. Pimental, S.L. Salzberg, J.L. Rinn, and L. Pachter. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* 7:3 (2012), 562-578.
20. S.L. Salzberg A.M. Phillippy, A. Zimin, D. Puiu, T. Magoc, S. Koren, T. Treangen, M.C. Schatz, A.L. Delcher, M. Roberts, G. Marçais, M. Pop, and J.A. Yorke. GAGE: a critical evaluation of genome assemblies and assembly algorithms. *Genome Research*, 22:3 (2012), 557-567.
21. T.J. Treangen and S.L. Salzberg. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Reviews Genetics* 13, 36-46 (2012).
22. D.E. Wood, H. Lin, A. Levy-Moonshine, R. Swaminathan, Y.-C. Chang, B.P. Anton, L. Osman, M. Steffen, S. Kasif, and S.L. Salzberg. Thousands of missed genes found in bacterial genomes and their analysis with COMBREX. *Biology Direct* 2012, **7**:37.
23. A.V. Zimin, D.R. Kelley, M. Roberts, S.L. Salzberg, and J.A. Yorke. Mis-assembled segmental duplications in two versions of the *Bos taurus* genome. *PLoS ONE* 7(8): e42680 (2012).
24. D.R. Kelley, B. Liu, A.L. Delcher, M. Pop, and S.L. Salzberg. Gene prediction with Glimmer for metagenomic sequences augmented by classification and clustering. *Nucleic Acids Research* 40:1 (2012), e9. doi:10.1093/nar/gkr1067.
25. M.C. Schatz, A.M. Phillippy, D.D. Sommer, A.L. Delcher, D. Puiu, G. Narzisi, S.L. Salzberg, and Mihai Pop. Hawkeye & AMOS: Visualizing and assessing the quality of genome assemblies. *Briefings in Bioinformatics*, 2011, doi:10.1093/bib/bbr074.
26. D. Kim and S.L. Salzberg. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biology* (2011), 12:R72. **Highly accessed**.
27. T. Magoč and S.L. Salzberg. FLASH: Fast Length Adjustment of Short Reads to Improve Genome Assemblies. *Bioinformatics* 27:21 (2011), 2957-2963.
28. M. Pertea, G.M. Pertea, and S.L. Salzberg. Detection of lineage-specific evolutionary changes among primate species. *BMC Bioinformatics* 2011, 12:274. **Highly accessed**.
29. A.J. Bogdanove, R. Koebnik, H. Lu, A. Furutani, S.V. Angiuoli, P.B. Patil, M. Van Sluys, R.P. Ryan, D.F. Meyer, S.-W. Han, G. Aparna, M. Rajaram, A.L. Delcher, A.M. Phillippy, D. Puiu, M.C. Schatz, M. Shumway, D.D. Sommer, C. Trapnell, F. Benahmed, G. Dimitrov, R. Madupu, D. Radune, S. Sullivan, G. Jha, H. Ishihara, S.-W. Lee, A. Pandey, V. Sharma, M. Sriariyanun, B. Szurek, C.M. Vera-Cruz, K.S. Dorman, P.C. Ronald, V. Verdier, J.M. Dow, R.V. Sonti, S. Tsuge, V. Brendel, P.D. Rabinowicz, J.E.

- Leach, F.F. White, and S.L. Salzberg. Two new complete genome sequences offer insight into host and tissue specificity of plant pathogenic *Xanthomonas spp.* *J. Bacteriology* 193:19 (2011), 5450-64.
30. J. Enk, R. Debruyne, A. Devault, C.E. King, T. Treangen, D. O'Rourke, S.L. Salzberg, D. Fisher, R. MacPhee, and H. Poinar. Complete Columbian mammoth mitogenome suggests interbreeding with woolly mammoths. *Genome Biology* (2011), 12:R51. **Highly accessed.**
 31. A. Brady and S. Salzberg. PhymmBL expanded: confidence scores, custom databases, parallelization and more. *Nature Methods* **8**, 367 (2011). doi:10.1038/nmeth0511-367.
 32. L. Florea, A. Souvorov, T.S. Kalbfleisch, and S.L. Salzberg (2011). Genome assembly has a major impact on gene content: A comparison of annotation in two *Bos taurus* assemblies. *PLoS ONE* 6(6), e21400. doi:10.1371/journal.pone.0021400.
 33. S.V. Angiuoli, J.C. Dunning Hotopp, S.L. Salzberg and H. Tettelin. Improving pan-genome annotation using whole genome multiple alignment. *BMC Bioinformatics* 2011, 12:272. **Highly accessed.**
 34. D.A. Rasko, P. Worsham, T.G. Abshire, S.T. Stanley, J.D. Bannan, M.R. Wilson, R. Langham, R.S. Decker, L. Jiang, T.D. Read, A.M. Phillippy, S.L. Salzberg, M. Pop, M.N. Van Ert, L.J. Kenefic, P.S. Keim, C.M. Fraser-Liggett, and J. Ravel. *Bacillus anthracis* comparative genome analysis in support of the Amerithrax investigation. *Proc. Natl. Acad. Sci. USA* 102:12(2011), 5027-32.
 35. V Shulaev, DJ Sargent, RN Crowhurst, TC Mockler, O Folkerts, AL Delcher, P Jaiswal, K Mockaitis, A Liston, SP Mane, P Burns, TM Davis, JP Slovin, N Bassil, RP Hellens, C Evans, T Harkins, C Kodira, B Desany, OR Crasta, RV Jensen, AC Allan, TP Michael, JC Setubal, J-M Celton, DJG Rees, KP Williams, SH Holt, JJR Rojas, M Chatterjee, B Liu, H Silva, L Meisel, A Adato, SA Filichkin, M Troggio, R Viola, T-L Ashman, H Wang, P Dharmawardhana, J Elser, R Raja, HD Priest, DW Bryant Jr, SE Fox, SA Givan, LJ Wilhelm, S Naithani, A Christoffels, DY Salama, J Carter, EL Girona, A Zdepski, W Wang, RA Kerstetter, W Schwab, SS Korban, J Davik, A Monfort, B Denoyes-Rothan, P Arus, R Mittler, B Flinn, A Aharoni, JL Bennetzen, SL Salzberg, AW Dickerman, R Velasco, M Borodovsky, RE Veilleux, and KM Folta. The genome of woodland strawberry (*Fragaria vesca*). *Nature Genetics* 43(2011), 109-116.
 36. S.V. Angiuoli and S.L. Salzberg. Mugsy: Fast multiple alignment of closely related whole genomes. *Bioinformatics* (2011), 27(3), 334-342.
 37. R.J. Roberts, Y.-C. Chang, Z. Hu, J.N. Rachlin, B.P. Anton, R.M. Pokrzywa, H.-P. Choi, L.L. Faller, J. Guleria, G. Housman, N. Klitgord, V. Mazumdar, M.G. McGettrick, L. Osmani, R. Swaminathan, K.R. Tao, S. Letovsky, D. Vitkup, D. Segrè, S.L. Salzberg, C. Delisi, M. Steffen, and S. Kasif. COMBEX: a project to accelerate the functional annotation of prokaryotic genomes. *Nucleic Acids Research* (2011) January; 39(Database issue):D11-D14. doi: 10.1093/nar/gkq1168.
 38. S.L. Salzberg and M. Pertea. Do-it-yourself genetic testing. *Genome Biology* (2010), 11:404.
 39. D.R. Kelley, M.C. Schatz, and S.L. Salzberg. Quake: quality-aware detection and correction of sequencing errors. *Genome Biology* (2010), 11:R116. <http://genomebiology.com/2010/11/11/R116/>.
 40. D.R. Kelley and S.L. Salzberg. Clustering metagenomic sequences with interpolated Markov models. *BMC Bioinformatics* 11:544 (2010). <http://www.biomedcentral.com/1471-2105/11/544>
 41. M.C. Schatz, B. Langmead, and S.L. Salzberg. Cloud computing and the DNA data race. *Nature Biotechnology* **28**, 691-693 (2010) doi:10.1038/nbt0710-691.
 42. M.C. Schatz, A.L. Delcher, and S.L. Salzberg. Assembly of large genomes using second-generation sequencing. *Genome Research* **20**, 1165-1173 (2010).
 43. C. Trapnell, B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. van Baren, S.L. Salzberg, B.J. Wold, and L. Pachter. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* **28**, 511-515 (2 May 2010).
 44. T.R. Miller, A.L. Delcher, S.L. Salzberg, E. Saunders, J.C. Detter, and R.U. Halden. Genome sequence of the dioxin-mineralizing bacterium *Sphingomonas wittichii* RW1. *J. Bacteriology* 192 (2010), 6101-2.
 45. S.L. Salzberg. Recent advances in RNA sequence analysis. *F1000 Biology Reports* 2010, **2**:64.
 46. M. Pertea and S.L. Salzberg. Between a chicken and a grape: estimating the number of human genes. *Genome Biology* (2010), 11:206, doi:10.1186/gb-2010-11-5-206.
 47. R.A. Dalloul, J.A. Long, A.V. Zimin, L. Aslam, K. Beal, L. Blomberg, D.W. Burt, O. Crasta, R.P.M.A. Crooijmans, K. Cooper, R.A. Coulombe, S. De, M.E. Delany, J.B. Dodgson, J.J. Dong, C. Evans, P.

- Flicek, L. Florea, O. Folkerts, M.A.M. Groenen, T.T. Harkins, J. Herrero, S. Hoffmann, H.-J. Megens, A. Jiang, P. de Jong, P. Kaiser, H. Kim, K.-W. Kim, S. Kim, D. Langenberger, M.-K. Lee, T. Lee, S. Mane, G. Marcais, M. Marz, A.P. McElroy, T. Modise, M. Nefedov, C. Notredame, I.R. Paton, W.S. Payne, G. Pertea, D. Prickett, D. Puiu, D. Qioa, E. Raineri, S.L. Salzberg, M.C. Schatz, C. Scheuring, C.J. Schmidt, S. Schroeder, E.J. Smith, J. Smith, T.S. Sonstegard, P.F. Stadler, H. Tafer, Z. Tu, C.P. Van Tassell, A.J. Vilella, K. Williams, J.A. Yorke, L. Zhang, H.-B. Zhang, X. Zhang, Y. Zhang, and K.M. Reed. Multi-platform next generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *PLoS Biology* (2010), 8(9): e1000475. doi:10.1371/journal.pbio.1000475.
48. X. Deng, A.M. Phillippy, Z. Li, S.L. Salzberg, and W. Zhang. Probing the pan-genome of *Listeria monocytogenes*: new insights into intraspecific niche expansion and genomic diversification. *BMC Genomics* (2010), 11:500.
49. S.L. Salzberg. Mind the gaps. *Nature Methods* 7:2 (2010), 105-106.
50. D.R. Kelley and S.L. Salzberg. Detection and correction of false segmental duplications caused by genome mis-assembly. *Genome Biology* (2010), 11:R28. doi:10.1186/gb-2010-11-3-r28
51. B. Langmead, M.C. Schatz, J. Lin, M. Pop, and S.L. Salzberg. Searching for SNPs with cloud computing. *Genome Biology* (2009) 10:R134. doi:10.1186/gb-2009-10-11-r134. **Highly accessed.**
52. A. Brady and S.L. Salzberg. Phymm and PhymmBL: Metagenomic phylogenetic classification with interpolated Markov models. *Nature Methods*, 6:9 (2009), 673 – 676.
53. A.V. Zimin, A.L. Delcher, L. Florea, D.A. Kelley, M.C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C.P. Van Tassell, T.S. Sonstegard, G. Marçais, M. Roberts, P. Subramanian, J.A. Yorke, and S.L. Salzberg. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biology* (2009), 10:R42. **Highly accessed.**
54. C. Kingsford, N. Nagarajan and S.L. Salzberg. 2009 Swine-Origin Influenza A (H1N1) Resembles Previous Influenza Isolates. *PLoS ONE* 4:7 (2009), e6402. (doi:10.1371/journal.pone.0006402)
55. M. Berriman, B.J. Haas, P.T. LoVerde, R.A. Wilson, G.P. Dillon, G.C. Cerqueira, S.T. Mashiyama, B. Al-Lazikani, L.F. Andrade, P.D. Ashton, M.A. Aslett, D.C. Bartholomeu, G. Blandin, C.R. Caffrey, A. Coghlan, R.Coulson, T.A. Day, A. Delcher, R. DeMarco, A. Djikeng, T. Eyre, J.A. Gamble, E. Ghedin, Y. Gu, C. Hertz-Fowler, H. Hirai, Y. Hirai, R. Houston, A. Ivens, D.A. Johnston, D. Lacerda, C.D. Macedo, P. McVeigh, Z. Ning, G. Oliveira, J.P. Overington, J. Parkhill, M. Pertea, R.J. Pierce, A.V. Protasio, M.A. Quail, M.-A. Rajandream, J. Rogers, M. Sajid, S.L. Salzberg, M. Stanke, A.R. Tivey, O. White, D.L. Williams, J. Wortman, W. Wu, M. Zamanian, A. Zerlotini, C.M. Fraser-Liggett, B.G. Barrell, and N.M. El-Sayed. The genome of the blood fluke *Schistosoma mansoni*. *Nature* 460 (2009), 352-358.
56. B. Langmead, C. Trapnell, M. Pop, and S.L. Salzberg. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* (2009) 10:R25. **Winner of Genome Biology Award for Best Research Paper of 2009.**
57. C. Trapnell, L. Pachter, and S.L. Salzberg. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* 25 (2009), 1105-1111. **Selected by Bioinformatics as the best Next Generation Sequencing paper in the journal (2008-2009).**
58. Cole Trapnell and Steven L. Salzberg. How to map billions of short reads onto genomes. *Nature Biotechnology* 27:5 (2009), 455-457.
59. A.M. Phillippy, X. Deng, W. Zhang, and S.L. Salzberg. Efficient oligonucleotide probe selection for pan-genomic tiling arrays. *BMC Bioinformatics* 10:293 (2009). (doi:10.1186/1471-2105-10-293)
60. A.M. Phillippy, K. Ayanbule, N.J. Edwards, and S.L. Salzberg. Insignia: a DNA signature search web server for diagnostic assay development. *Nucleic Acids Research* 37 (2009), W229-W234.
61. S.L. Salzberg, D. Puiu, D.D. Sommer, V. Nene, and N.H. Lee. The genome sequence of *Wolbachia* endosymbiont of *Culex quinquefasciatus* JHB. *J. Bacteriology* 191:5 (2009). 1725.
62. J. Ravel, L. Jiang, S.T. Stanley, M.R. Wilson, R.S. Decker, T.D. Read, P. Worsham, P.S. Keim, S.L. Salzberg, C.M. Fraser-Liggett, and D.A. Rasko. The complete genome sequence of *Bacillus anthracis* Ames “Ancestor.” *J. Bacteriology* 191:1 (2009), 445-446.

63. M. Pertea, K. Ayanbule, M. Smedinghoff, and S.L. Salzberg. OperonDB: a comprehensive database of predicted operons in microbial genomes. *Nucleic Acids Research* 2009 37(Database issue):D479-D482; doi:10.1093/nar/gkn784.
64. S.L. Salzberg, D.D. Sommer, D. Puiu, and V.T. Lee. Gene-boosted assembly of a novel bacterial genome from very short reads. *PLoS Computational Biology*, 4:9 (2008), e1000186. doi:10.1371/journal.pcbi.1000186.
65. J.M. Carlton, J.H. Adams, J.C. Silva, S.L. Bidwell, H. Lorenzi, E. Caler, J. Crabtree, S.V. Angiuoli, E.F. Merino, P. Amedeo, Q. Cheng, R.M.R. Coulson, B.S. Crabb, H.A. del Portillo, K. Essien, T.V. Feldblyum, C. Fernandez-Becerra, P.R. Gilson, A.H. Gueye, X. Guo, S. Kang'a, T.W.A. Kooij, M. Korsinczky, E.V.-S. Meyer, Vish Nene, I. Paulsen, O. White, S.A. Ralph, Q. Ren, T.J. Sargeant, S.L. Salzberg, C.J. Stoeckert, S.A. Sullivan, M.M. Yamamoto, S.L. Hoffman, J.R. Wortman, M.J. Gardner, M.R. Galinski, J.W. Barnwell, and C.M. Fraser-Liggett. Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature* 455 (2008), 757-763.
66. E.V. Shakirov, S.L. Salzberg, M. Alam, and D.E. Shippen. Analysis of *Carica papaya* Telomeres and Telomere-Associated Proteins: Insights into the Evolution of Telomere Maintenance in Brassicales. *Tropical Plant Biology* (2008) doi:10.1007/s12042-008-9018-x.
67. H. Lu, P. Patil, M. Van Sluys, F.F. White, R.P. Ryan, J.M. Dow, P. Rabinowicz, S.L. Salzberg, J.E. Leach, R. Sonti, V. Brendel, and A.J. Bogdanove. Acquisition and Evolution of Plant Pathogenesis-Associated Gene Clusters and Candidate Determinants of Tissue-Specificity in *Xanthomonas*. *PLoS ONE* 3:11 (2008), e3828. doi:10.1371/journal.pone.0003828.
68. Daniela Puiu and Steven L. Salzberg. Re-assembly of the genome of *Francisella tularensis* subsp. *holarctica* OSU18. *PLoS ONE* 3:10 (2008): e3427. doi:10.1371/journal.pone.0003427.
69. C. Kingsford and S.L. Salzberg. What are decision trees? *Nature Biotechnology* 26:9 (2008), 1011-1013.
70. N. Nagarajan, R. Navajas-Pérez, M. Pop, M. Alam, R. Ming, A.H. Paterson, and S.L. Salzberg. Genome-wide analysis of repetitive elements in papaya. *Tropical Plant Biology* (2008) doi:10.1007/s12042-008-9015-0.
71. S.L. Salzberg. The contents of the syringe. *Nature* 454 (2008), 160-162.
72. S.L. Salzberg, D.D. Sommer, M.C. Schatz, A.M. Phillippy, P.D. Rabinowicz, S. Tsuge, A. Furutani, H. Ochiai, A.L. Delcher, D. Kelley, R. Madupu, D. Puiu, D. Radune, M. Shumway, C. Trapnell, G.r Aparna, G. Jha, A. Pandey, P.B Patil, H. Ishihara, D.F. Meyer, B. Szurek, V. Verdier, R. Koebnik, J.M. Dow, R.P. Ryan, H. Hirata, S. Tsuyumu, S.W. Lee, Y.-S. Seo, M. Sriariyanum, P.C. Ronald, R.V. Sonti, M. Van Sluys, J.E. Leach, F.F. White and A.J. Bogdanove. Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* PXO99A. *BMC Genomics* 9:204 (2008). Highly accessed.
73. R. Ming, Sh. Hou, Y. Feng, Q. Yu, A. Dionne-Laporte, J.H. Saw, P. Senin, W. Wang, B.V. Ly, K.L.T. Lewis, S.L. Salzberg, L. Feng, M.R. Jones, R.L. Skelton, J.E. Murray, C. Chen, W. Qian, J. Shen, P. Du, M. Eustice, E. Tong, H. Tang, E. Lyons, R.E. Paull, T.P. Michael, K. Wall, D.W. Rice, H. Albert, M.-L. Wang, Y.J. Zhu, M. Schatz, N. Nagarajan, R.A. Acob, P. Guan, A. Blas, C.M. Wai, C.M. Ackerman, Y. Ren, C. Liu, Jianmei Wang, Jianping Wang, J.-K. Na, E.V. Shakirov, B. Haas, J. Thimmapuram, D. Nelson, X. Wang, J.E. Bowers, A.R. Gschwend, A.L. Delcher, R. Singh, J.Y. Suzuki, S. Tripathi, K. Neupane, H. Wei, B. Irikura, M. Paidi, N. Jiang, W. Zhang, G. Presting, A. Windsor, R. Navajas-Pérez, M.J. Torres, F.A. Feltus, B. Porter, Y. Li, A.M. Burroughs, M.-C. Luo, L. Liu, D.A. Christopher, S.M. Mount, P.H. Moore, T. Sugimura, J. Jiang, M.A. Schuler, V. Friedman, T. Mitchell-Olds, D.E. Shippen, C.W. dePamphilis, J.D. Palmer, M. Freeling, A.H. Paterson, D. Gonsalves, L. Wang, and M. Alam. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* 452 (2008), 991-996.
74. M. Pop and S.L. Salzberg. Bioinformatics challenges of new sequencing technology. *Trends in Genetics* 24:3 (2008), 142-149.
75. B.J. Haas, S.L. Salzberg, W. Zhu, M. Pertea, J.E. Allen, J. Orvis, O. White, C.R. Buell, and J.R. Wortman. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biology* (2008), 9:R7, <http://genomebiology.com/2008/9/1/R7>. Highly accessed.

76. *Drosophila* 12 Genomes Consortium. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450 (2007), 203-218.
77. E. Ghedin, S. Wang, D. Spiro, E. Caler, Q. Zhao, J. Crabtree, J.E. Allen, A.L. Delcher, D.B. Guiliano, D. Miranda-Saavedra, S.V. Angiuoli, T. Creasy, P. Amedeo, B. Haas, N.M. El-Sayed, J.R. Wortman, T. Feldblyum, L. Tallon, M. Schatz, M. Shumway, H. Koo, S.L. Salzberg, S. Schobel, M. Pertea, M. Pop, O. White, G.J. Barton, C.K.S. Carlow, M. Crawford, J. Daub, M.W. Dimmic, C.F. Estes, J.M. Foster, M. Ganatra, W.F. Gregory, N.M. Johnson, J. Jin, R. Komuniecki, I. Korf, S. Kumar, S. Laney, B.-W. Li, W. Li, T.H. Lindblom, S. Lustigman, D. Ma, C.V. Maina, D.M.A. Martin, J.P. McCarter, L. McReynolds, M. Mitreva, T.B. Nutman, J. Parkinson, J.M. Peregrín-Alvarez, C. Poole, Q. Ren, L. Saunders, A.E. Sluder, K. Smith, M. Stanke, T.R. Unnasch, J. Ware, A.D. Wei, G. Weil, D.J. Williams, Y. Zhang, S.A. Williams, C. Fraser-Liggett, B. Slatko, M.L. Blaxter, and A.L. Scott. Draft Genome of the Filarial Nematode Parasite *Brugia malayi*. *Science* 317:5845 (2007), 1756-60.
78. C. Kingsford, A.L. Delcher, and S.L. Salzberg. A unified model explaining the offsets of overlapping and near-overlapping prokaryotic genes. *Molecular Biology and Evolution*, 24:9 (2007), 2091-2098.
79. S.L. Salzberg, C. Kingsford, G. Cattoli, D.J. Spiro, D.A. Janies, M.M. Aly, I.H. Brown, E. Couacy-Hymann, G.M. De Mia, D.H. Dung, A. Guercio, T. Joannis, A.S. Maken Ali, A. Osmani, I. Padalino, M.D. Saad, V. Savić, N.A. Sengamalay, S. Yingst, J. Zaborsky, O. Zorman-Rojs, E. Ghedin, and I. Capua. Genome analysis linking recent European and African influenza (H5N1) viruses. *Emerging Infectious Diseases* 13:5 (2007), 713-718 (<http://www.cdc.gov/EID/content/13/5/713.htm>).
80. A.M. Phillippy, J.A. Mason, K. Ayanbule, D.D. Sommer, E. Taviani, A. Huq, R.R. Colwell, I.T. Knight, and S.L. Salzberg. Comprehensive DNA signature discovery and validation. *PLoS Computational Biology* 3:5 (2007), e98.
81. A.L. Delcher, K.A. Bratke, E.C. Powers, and S.L. Salzberg. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:6 (2007), 673-679.
82. B.J. Haas and S.L. Salzberg. Finding repeats in genome sequences. In *Bioinformatics – From Genomes to Therapies*, Volume 1: Molecular Sequences and Structures (T. Lengauer, ed.). Weinheim, Germany: Wiley-VCH, 2007, 197-234.
83. V. Nene, J.R. Wortman, D. Lawson, B. Haas, C. Kodira, Z. Tu, B. Loftus, Z. Xi, K. Megy, M. Grabherr, Q. Ren, E.M. Zdobnov, N.F. Lobo, K.S. Campbell, S.E. Brown, M.F. Bonaldo, J. Zhu, S.P. Sinkins, D.G. Hogenkamp, P. Amedeo, P. Arensburger, P.W. Atkinson, S. Bidwell, J. Biedler, E. Birney, R.V. Bruggner, J. Costas, M.R. Coy, J. Crabtree, M. Crawford, B. deBruyn, D. DeCaprio, K. Eiglmeier, E. Eisenstadt, H. El-Dorry, W.M. Gelbart, S.L. Gomes, M. Hammond, L.I. Hannick, J.R. Hogan, M.H. Holmes, D. Jaffe, J.S. Johnston, R.C. Kennedy, H. Koo, S. Kravitz, E.V. Kriventseva, D. Kulp, K. LaButti, E. Lee, S. Li, D.D. Lovin, C. Mao, E. Mauceli, C.F.M. Menck, J.R. Miller, P. Montgomery, A. Mori, A.L. Nascimento, H.F. Naveira, C. Nusbaum, S. O'Leary, J. Orvis, M. Pertea, H. Quesneville, K.R. Reidenbach, Y.-H. Rogers, C.W. Roth, J.R. Schneider, M. Schatz, M. Shumway, M. Stanke, E.O. Stinson, J.M.C. Tubio, J.P. VanZee, S. Verjovski-Almeida, D. Werner, O. White, S. Wyder, Q. Zeng, Q. Zhao, Y. Zhao, C.A. Hill, A.S. Raikhel, M.B. Soares, D.L. Knudson, N.H. Lee, J. Galagan, S.L. Salzberg, I.T. Paulsen, G. Dimopoulos, F.H. Collins, B. Birren, C.M. Fraser-Liggett, and D.W. Severson. Genome Sequence of *Aedes aegypti*, a Major Arbovirus Vector. *Science* 316:5832 (2007), 1718-1723.
84. M. Pertea, S.M. Mount, and S.L. Salzberg. A computational survey of candidate exon splicing enhancer motifs in the model plant *Arabidopsis thaliana*. *BMC Bioinformatics* (2007), 8:159.
85. S.L. Salzberg. Genome re-annotation: a wiki solution? *Genome Biology* 2007, 8:102. **Highly accessed.**
86. D.D. Sommer, A.L. Delcher, S.L. Salzberg, and M. Pop. Minimus: A fast, lightweight genome assembler. *BMC Bioinformatics* 8:64 (2007). **Highly accessed**
87. J.M. Carlton, R.P. Hirt, J.C. Silva, A.L. Delcher, M. Schatz, Q. Zhao, J.R. Wortman, S.L. Bidwell, U.C.M. Alsmark, S. Besteiro, T. Sicheritz-Ponten, C.J. Noel, J.B. Dacks, P.G. Foster, C. Simillion, Y. Van de Peer, D. Miranda-Saavedra, G.J. Barton, G.D. Westrop, S. Müller, D. Dessi, P.L. Fiori, Q. Ren, I. Paulsen, H. Zhang, F.D. Bastida-Corcuera, A. Simoes-Barbosa, M.T. Brown, R.D. Hayes, M. Mukherjee, C.Y. Okumura, R. Schneider, A.J. Smith, S. Vanacova, M. Villalvazo, B.J. Haas, M. Pertea, T.V. Feldblyum, T.R. Utterback, C.-L. Shu, K. Osoegawa, P.J. de Jong, I. Hrdy, L. Horvathova, Z. Zubacova, P. Dolezal,

- S.-B. Malik, J.M. Logsdon Jr., K. Henze, A. Gupta, C.C. Wang, R.L. Dunne, J.A. Upcroft, P. Upcroft, O. White, S.L. Salzberg, P. Tang, C.-H. Chiu, Y.-S. Lee, T.M. Embley, G.H. Coombs, J.C. Mottram, J. Tachezy, C.M. Fraser-Liggett, and P.J. Johnson. Draft Genome Sequence of the Sexually Transmitted Pathogen *Trichomonas vaginalis*. *Science* 315 (2007), 207-212.
88. C.L. Kingsford, K. Ayanbule, and S.L. Salzberg. Rapid, accurate computational discovery of rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biology* (2007), 8:R22.
89. M.C. Schatz, A.M. Phillippy, B. Shneiderman, and S.L. Salzberg. Hawkeye: An interactive visual analytics tool for genome assembly. *Genome Biology* (2007), 8:R34. **Highly accessed**
90. J.E. Allen, W.M. Majoros, M. Pertea, and S.L. Salzberg. JIGSAW, GeneZilla, and GlimmerHMM: puzzling out the features of human genes in the ENCODE regions. *Genome Biology* 7 (2006). Suppl 1:S9.
91. J.E. Allen and S.L. Salzberg. A phylogenetic generalized hidden Markov model for predicting alternatively spliced exons. *Algorithms for Molecular Biology* 1:14 (2006). **Highly accessed.**
92. J. Quackenbush and S.L. Salzberg. It is time to end the patenting of software. *Bioinformatics* 22:12 (2006), 1416-1417.
93. J.A. Eisen, R.S. Coyne, M. Wu, D. Wu, M. Thiagarajan, J.R. Wortman, J.H. Badger, Q. Ren, P. Amedeo, K.M. Jones, L.J. Tallon, A.L. Delcher, S.L. Salzberg, J.C. Silva, Brian J. Haas, William H. Majoros, Maryam Farzad, Jane M. Carlton, R.K. Smith Jr., J. Garg, R.E. Pearlman, K.M. Karrer, L. Sun, G. Manning, N.C. Elde, A.P. Turkewitz, D.J. Asai, D.E. Wilkes, Y. Wang, H. Cai, K. Collins, B.A. Stewart, S.R. Lee, K. Wilamowska, Z. Weinberg, W.L. Ruzzo, D. Wloga, J. Gaertig, J. Frankel, C.-C. Tsao, M.A. Gorovsky, P.J. Keeling, R.F. Waller, N.J. Patron, J.M. Cherry, N.A. Stover, C.J. Krieger, C. del Toro, H.F. Ryder, S.C. Williamson, R.A. Barbeau, E.P. Hamilton, and E. Orias. Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biology* 4:9 (2006), e286.
94. R.L. Malek, H.-Y. Wang, A.E. Kwitek, A.S. Greene, N. Bhagabati, G. Borchardt, L. Cahill, T. Currier, B. Frank, X. Fu, M. Hasinoff, E. Howe, N. Letwin, T.V. Luu, A. Saeed, H. Sajadi, S.L. Salzberg, R. Sultana, M. Thiagarajan, J. Tsai, K. Veratti, J. White, J. Quackenbush, H.J. Jacob and N.H. Lee. Physiogenomic resources for rat models of heart, lung and blood disorders. *Nature Genetics* 38 (2006), 234-239.
95. Elodie Ghedin, Naomi A. Sengamalay, Martin Shumway, Jennifer Zaborosky, Tamara Feldblyum, Vik Subbu, David J. Spiro, Jeff Sitz, Hean Koo, Pavel Bolotov, Dmitry Dernovoy, Tatiana Tatusova, Yiming Bao, Kirsten St. George, Jill Taylor, David J. Lipman, Claire M. Fraser, Jeffery K. Taubenberger, and Steven L. Salzberg. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature* 437 (2005), 1162-1166.
96. Steven L. Salzberg and James A. Yorke. Beware of Mis-assembled Genomes. *Bioinformatics* 21:24 (2005), 4320-21.
97. W.C. Nierman, A. Pain, M.J. Anderson, J.R. Wortman, H.S. Kim, J. Arroyo, M. Berriman, K. Abe, D.B. Archer, C. Bermejo, J. Bennett, P. Bowyer, D. Chen, M. Collins, R. Coulsen, R. Davies, P.S. Dyer, M. Farman, Nadia Fedorova, Natalie Fedorova, T.V. Feldblyum, R. Fischer, N. Fosker, A. Fraser, J.L. García, M.J. García, A. Goble, G.H. Goldman, K. Gomi, S. Griffith-Jones, R. Gwilliam, B. Haas, H. Haas, D. Harris, H. Horiuchi, J. Huang, S. Humphray, J. Jiménez, N. Keller, H. Khouri, K. Kitamoto, T. Kobayashi, S. Konzack, R. Kulkarni, T. Kumagai, A. Lafton, J.-P. Latgé, W. Li, A. Lord, C. Lu, W.H. Majoros, G.S. May, B.L. Miller, Y. Mohamoud, M. Molina, M. Monod, I. Mouyna, S. Mulligan, L. Murphy, S. O'Neil, I. Paulsen, M.A. Peñalva, M. Pertea, C. Price, B.L. Pritchard, M.A. Quail, E. Rabinowitsch, N. Rawlins, M.-A. Rajandream, U. Reichard, H. Renault, G.D. Robson, S.R. de Córdoba, J.M. Rodríguez-Peña, C.M. Ronning, S. Rutter, S.L. Salzberg, M. Sanchez, J.C. Sánchez-Ferrero, D. Saunders, K. Seeger, R. Squares, S. Squares, M. Takeuchi, F. Tekaiia, G. Turner, C.R. Vazquez de Aldana, J. Weidman, O. White, J. Woodward, J.-H. Yu, C. Fraser, J.E. Galagan, K. Asai, M. Machida, N. Hall, B. Barrell, and D.W. Denning. Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* 438 (2005), 1151-1156.
98. Edward C. Holmes, Elodie Ghedin, Naomi Miller, Jill Taylor, Yiming Bao, Kirsten St. George, Bryan T. Grenfell, Steven L. Salzberg, Claire M. Fraser, David J. Lipman, and Jeffery K. Taubenberger. Whole-

- Genome Analysis of Human Influenza A Virus Reveals Multiple Persistent Lineages and Reassortment among Recent H3N2 Viruses. *PLoS Biology* 3:9 (2005), e300.
99. The FANTOM Consortium [224 authors]. The Transcriptional Landscape of the Mammalian Genome. *Science* 309 (2005), 1559-1563.
 100. Steven L. Salzberg, Julie C. Dunning Hotopp, Arthur L. Delcher, Mihai Pop, Douglas R. Smith, Michael B. Eisen, and William C. Nelson. Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species. *Genome Biology* (2005), 6:R23. [Highly accessed](#).
 101. N.M. El-Sayed, P.J. Myler, G. Blandin, M. Berriman, J. Crabtree, G. Aggarwal, E. Caler, H. Renauld, E.A. Worthey, C. Hertz-Fowler, E. Ghedin, C. Peacock, D.C. Bartholomeu, B.J. Haas, A.-N. Tran, J.R. Wortman, U.C.M. Alsmark, S. Angiuoli, A. Anupama, J. Badger, F. Bringaud, E. Cadag, J.M. Carlton, G. Cerqueira, T. Creasy, A.L. Delcher, A. Djikeng, T.M. Embley, C. Hauser, A.C. Ivens, S.K. Kummerfeld, J.B. Pereira-Leal, D. Nilsson, J. Peterson, S.L. Salzberg, J. Shallom, J.C. Silva, J. Sundaram, S. Westenberger, O. White, S.E. Melville, J.E. Donelson, B. Andersson, K.D. Stuart, and N. Hall. Comparative genomics of trypanosomatid parasitic protozoa. *Science* 309:5733 (2005), 404-409.
 102. M. Berriman, E. Ghedin, C. Hertz-Fowler, G. Blandin, H. Renauld, D.C. Bartholomeu, N.J. Lennard, E. Caler, N.E. Hamlin, B. Haas, U. Böhme, L. Hannick, M.A. Aslett, J. Shallom, L. Marcello, L. Hou, B. Wickstead, U.C.M. Alsmark, C. Arrowsmith, R.J. Atkin, A.J. Barron, F. Bringaud, K. Brooks, M. Carrington, I. Cherevach, T.-J. Chillingworth, C. Churcher, L.N. Clark, C.H. Corton, A. Cronin, R.M. Davies, J. Doggett, A. Djikeng, T. Feldblyum, M.C. Field, A. Fraser, I. Goodhead, Z. Hance, D. Harper, B.R. Harris, H. Hauser, J. Hostetler, A. Ivens, K. Jagels, D. Johnson, J. Johnson, K. Jones, A.X. Kerhornou, H. Koo, N. Larke, S. Landfear, C. Larkin, V. Leech, A. Line, A. Lord, A. MacLeod, P.J. Mooney, S. Moule, D.M.A. Martin, G.W. Morgan, K. Mungall, H. Norbertczak, D. Ormond, G. Pai, C.S. Peacock, J. Peterson, M.A. Quail, E. Rabinowitsch, M.-A. Rajandream, C. Reitter, S.L. Salzberg, M. Sanders, S. Schobel, S. Sharp, M. Simmonds, A.J. Simpson, L. Tallon, C.M.R. Turner, A. Tait, A.R. Tivey, S. Van Aken, D. Walker, D. Wanless, S. Wang, B. White, O. White, S. Whitehead, J. Woodward, J. Wortman, M.D. Adams, T. M. Embley, K. Gull, E. Ullu, J.D. Barry, A.H. Fairlamb, F. Opperdoes, B.G. Barrell, J.E. Donelson, Neil Hall, C.M. Fraser, S.E. Melville, and N.M. El-Sayed. The genome of the African trypanosome *Trypanosoma brucei*. *Science* 309:5733 (2005), 416-422.
 103. N.M. El-Sayed, P.J. Myler, D.C. Bartholomeu, D. Nilsson, G. Aggarwal, A.-N. Tran, E. Ghedin, E.A. Worthey, A.L. Delcher, G. Blandin, S.J. Westenberger, E. Caler, G.C. Cerqueira, C. Branche, B. Haas, A. Anupama, E. Arner, L. Åslund, P. Attipoe, E. Bontempi, F. Bringaud, P. Burton, E. Cadag, D.A. Campbell, M. Carrington, J. Crabtree, H. Darban, J.F. da Silveira, P. de Jong, K. Edwards, P.T. Englund, G. Fazelina, T. Feldblyum, M. Ferella, A.C. Frasch, K. Gull, D. Horn, L. Hou, Y. Huang, E. Kindlund, M. Klingbeil, Sindy Kluge, Hean Koo, Daniela Lacerda, Mariano J. Levin, H. Lorenzi, T. Louie, C.R. Machado, R. McCulloch, A. McKenna, Y. Mizuno, J.C. Mottram, S. Nelson, S. Ochaya, K. Osoegawa, G. Pai, M. Parsons, M. Pentony, U. Pettersson, M. Pop, J.L. Ramirez, J. Rinta, L. Robertson, S.L. Salzberg, D.O. Sanchez, A. Seyler, R. Sharma, J. Shetty, A.J. Simpson, E. Sisk, M.T. Tammi, R. Tarleton, S. Teixeira, S. Van Aken, C. Vogt, P.N. Ward, B. Wickstead, J. Wortman, O. White, C.M. Fraser, K.D. Stuart, and B. Andersson. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 309:5733 (2005), 409-415.
 104. M.J. Gardner, R. Bishop, T. Shah, E.P. de Villiers, J.M. Carlton, N. Hall, Q. Ren, I.T. Paulsen, A. Pain, M. Berriman, R.J. Wilson, S. Sato, S.A. Ralph, D.J. Mann, Z. Xiong, S.J. Shallom, J. Weidman, L. Jiang, J. Lynn, B. Weaver, A. Shoaibi, A.R. Domingo, D. Wasawo, J. Crabtree, J.R. Wortman, B. Haas, S.V. Angiuoli, T.H. Creasy, C. Lu, B. Suh, J.C. Silva, T.R. Utterback, T.V. Feldblyum, M. Pertea, J. Allen, W.C. Nierman, E. Taracha, S.L. Salzberg, O.R. White, H.A. Fitzhugh, S. Morzaria, J.C. Venter, C.M. Fraser, and V. Nene. Genome sequence of *Theileria parva*, a bovine pathogen that transforms lymphocytes. *Science*, 309:5731 (2005), 134-137.
 105. S. Subramanian, A. Kuspa, S.L. Salzberg, L. Aravind, H.A. Lewin, J. Quackenbush, W.R. Taylor, R.B. Altman, and P. Clote (Eds.). *Encyclopedia of Genetics, Genomics, Proteomics, and Bioinformatics: Bioinformatics*, Volume 7. West Sussex, England: John Wiley & Sons Ltd., 2005.

106. Jonathan E. Allen and Steven L. Salzberg. JIGSAW: integration of multiple sources of evidence for gene prediction. *Bioinformatics* 21:18 (2005), 3596-3603.
107. C.R. Buell, Q. Yuan, S. Ouang, J. Liu, W. Zhu, A. Wang, R. Maiti, B. Haas, J. Wortman, M. Pertea, K.M. Jones, M. Kim, L. Overton, T. Tsitrin, D. Fadrosch, J. Bera, B. Weaver, S. Jin, S. Johri, M. Reardon, K. Gansberger, J. Hill, K. Moffat, L. Tallon, S. Van Aken, M. Lewis, T. Utterback, T. Feldblyum, V. Zismann, S. Iobst, J. Hsiao, A. de Vazeilles, S.L. Salzberg, O. White, C.M. Fraser, Y. Yu, H. Kim, T. Rambo, J. Currie, K. Collura, F. Wei, D. Kudrna, W. Nelson, R. Jetty, M. Luo, D. Henry, R. Oates, M. Palmer, G. Pries, C. Sasaki, J. Simmons, M. de la Bastide, L. Spiegel, L. Nascimento, E. Huang, R. Preston, T. Zutavern, L. Palmer, A. O'Shaughnessy, S. Dike, P. Minx, H. Cordum, W. Jin, H.-R. Lee, J. Jiang, C. Soderlund, R. Wilson, W.R. McCombie, and R.A. Wing. Sequence, annotation, and analysis of synteny between rice chromosome 3 and diverged grass species. *Genome Research* 15 (2005), 1284-91.
108. W.H. Majoros, M. Pertea, and S.L. Salzberg. Efficient implementation of a generalized pair hidden Markov model for comparative gene finding. *Bioinformatics* 21 (2005), 1782-88.
109. Steven L. Salzberg, Deanna Church, Michael DiCuccio, Eugene Yaschenko, and James Ostell. The Genome Assembly Archive: a new public resource. *PLoS Biology* 9:2 (2004), 1273-1275.
110. The ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 306:5696 (2004), 636-640.
111. N.Ward, Ø. Larsen, J. Sakwa, L. Bruseth, H. Khouri, A.S. Durkin, G. Dimitrov, L. Jiang, D. Scanlan, K. Kang, M. Lewis, K.E. Nelson, B. Methé, M. Wu, J.F. Heidelberg, I.T. Paulsen, D. Fouts, J. Ravel, H. Tettelin, Q. Ren, T. Read, R. DeBoy, R. Seshadri, S.L. Salzberg, H.B. Jensen, N.K. Birkeland, W. Nelson, R.J. Dodson, S.H. Grindhaug, I. Holt, I. Eidhammer, I. Jonassen, S. Vanaken, T. Utterback, T.V. Feldblyum, J. Lillehaug, C.M. Fraser, and J.A. Eisen. Genomic insights into methanotrophy: The complete genome sequence of *Methylococcus capsulatus* (Bath). *PLoS Biology* 2:10 (2004), e303.
112. B.J. Loftus, E. Fung, P. Roncaglia, D. Rowley, P. Amedeo, D. Bruno, J. Vamathevan, M. Miranda, I.J. Anderson, J.A. Fraser, J.E. Allen, I.E. Bosdet, M.R. Brent, R. Chiu, T.L. Doering, M.J. Donlin, C.A. D'Souza, D.S. Fox, V. Grinberg, J. Fu, M. Fukushima, B.J. Haas, J.C. Huang, G. Janbon, S.J.M. Jones, M.I. Krzywinski, J.K. Kwon-Chung, K.B. Lengeler, R. Maiti, M.A. Marra, R.E. Marra, C.A. Mathewson, T.G. Mitchell, M. Pertea, F.R. Riggs, S.L. Salzberg, J.E. Schein, A. Shvartsbeyn, H. Shin, C.A. Specht, B.B. Suh, A. Tenney, T.R. Utterback, B.L. Wickes, N.H. Wye, J.W. Kronstad, J.K. Lodge, J. Heitman, R.W. Davis, C.M. Fraser, and R.W. Hyman. The Genome of the Basidiomycetous Yeast and Human Pathogen *Cryptococcus neoformans*. *Science* 307:5713 (2005), 1321-1324.
113. W.H. Majoros, Mihaela Pertea, Arthur L. Delcher, and Steven L. Salzberg. Efficient decoding algorithms for generalized hidden Markov model gene finders. *BMC Bioinformatics* (2005), 6:16.
114. Tom R. Slezak and Steven L. Salzberg. Bioinformatics methods for microbial detection and forensic diagnostic design. In *Microbial Forensics* (Roger G. Breeze, Bruce Budowle, and Steven E. Schutzer, Eds.). New York: Academic Press, 2005, 313-353.
115. W.H. Majoros and S.L. Salzberg. An empirical analysis of training protocols for probabilistic gene finders. *BMC Bioinformatics* (2004) 5, 206.
116. B.J. Haas, A.L. Delcher, J.R. Wortman, and S.L. Salzberg. DAGchainer: A tool for mining segmental genome duplications and synteny. *Bioinformatics* 20:18 (2004), 3643-6.
117. M. Pop, A. Phillippy, A.L. Delcher, and S.L. Salzberg. Comparative genome assembly. *Briefings in Bioinformatics* 5:3 (2004), 237-248.
118. P. Gajer, M. Schatz, and S.L. Salzberg. Automated correction of genome sequence errors. *Nucleic Acids Research*, 32:2 (2004), 562-569.
119. William H. Majoros, Mihaela Pertea, and Steven L. Salzberg. TigrScan and GlimmerHMM: two open-source ab initio eukaryotic gene finders. *Bioinformatics* 20 (2004), 2878-79.
120. Benjamin P. Berman, Barret D. Pfeiffer, Todd R. Laverty, Steven L. Salzberg, Gerald M. Rubin, Michael B. Eisen, and Susan E. Celniker. Binding-site clustering and the optimal use of comparative sequence data in identifying developmental enhancers in *Drosophila*. *Genome Biology*, 5:R61 (2004). Highly accessed.

121. Stefan Kurtz, Adam Phillippy, Arthur L. Delcher, Michael Smoot, Martin Shumway, Corina Antonescu, and Steven L. Salzberg. Versatile and Open Software for Comparing Large Genomes. *Genome Biology*, 5:R12 (2004). Highly accessed.
122. E. Ghedin, F. Bringaud, J. Peterson, P. Myler, M. Berriman, A. Ivens, B. Andersson, E. Bontempi, J. Eisen, S. Angiuoli, D. Wanless, A. Von Arx, L. Murphy, N. Lennard, S. Salzberg, M.D. Adams, O. White, N. Hall, K. Stuart, C.M. Fraser and N.M.A. El-Sayed. Gene syteny and evolution of genome architecture in trypanosomatids. *Molecular and Biochemical Parasitology* 134:2 (2004), 183-91.
123. M. Pop, D. Kosack, and S.L. Salzberg. Hierarchical scaffolding with Bambus. *Genome Research* 14 (2004), 149-159.
124. J.E. Allen, M. Pertea, and S.L. Salzberg. Computational gene prediction using multiple sources of evidence. *Genome Research* 14 (2004), 142-148.
125. A. Pain, J. Woodward, M.A. Quail, M.J. Anderson, R. Clark, M. Collins, N. Fosker, A. Fraser, D. Harris, N. Larke, L. Murphy, S. Humphray, S. O'Neil, M. Pertea, C. Price, E. Rabinowitsch, M.A. Rajandream, S. Salzberg, D. Saunders, K. Seeger, S. Sharp, T. Warren, D.W. Denning, B. Barrell, and N. Hall. Insight into the genome of *Aspergillus fumigatus*: analysis of a 922kb region encompassing the nitrate assimilation gene cluster. *Fungal Genetics and Biology* 41:4 (2004), 443-53.
126. S.L. Salzberg and A.L. Delcher. Tools for gene finding and whole genome comparison. In C.M. Fraser, T.D. Read, and K.E. Nelson (Eds.), *Microbial Genomes*. Totowa, NJ: Humana Press, 2004, 19-32.
127. S.L. Salzberg. Yeast rises again. *Nature* 423 (2003), 233-234.
128. S.L. Salzberg, E. Birney, S. Eddy, and O. White. Unrestricted free access works and must continue. Letter to *Nature* 422 (2003), 801.
129. N.M.A. El-Sayed, E. Ghedin, J. Song, A. MacLeod, F. Bringaud, C. Larkin, D. Wanless, J. Peterson, L. Hou, S. Taylor, A. Tweedie, N. Biteau, H.G. Khalak, X. Lin, T. Mason, L. Hannick, E. Caler, G. Blandin, D. Bartholomeu, A.J. Simpson, S. Kaul, H. Zhao, G. Pai, S. Van Aken, T. Utterback, B. Haas, H.L. Koo, L. Umayam, B. Suh, C. Gerrard, V. Leech, R. Qi, S. Zhou, D. Schwartz, T. Feldblyum, S.L. Salzberg, A. Tait, C.M.R. Turner, E. Ullu, O. White, S. Melville, M.D. Adams, C.M. Fraser, and J.E. Donelson. The sequence and analysis of *Trypanosoma brucei* chromosome II. *Nucleic Acids Research* 31:16 (2003), 4856-4863.
130. B.J. Haas, A.L. Delcher, S.M. Mount, J.R. Wortman, R.K. Smith, Jr., L.I. Hannick, R. Maiti, C.M. Ronning, D.B. Rusch, C.D. Town, S.L. Salzberg, and O. White. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research* 31:19 (2003), 5654-5666.
131. The Rice Chromosome 10 Sequencing Consortium. In-depth view of structure, activity, and evolution of rice chromosome 10. *Science* 300 (2003), 1566-1569.
132. T.D. Read, S.N. Peterson, N. Tourasse, L.W. Baillie, I.T. Paulsen, K.E. Nelson, H. Tettelin, D.E. Fouts, J.A. Eisen, S.R. Gill, E.K. Holtzapple, O.A. Økstad, E. Helgason, J. Rilstone, M. Wu, J.F. Kolonay, M.J. Beanan, R.J. Dodson, L.M. Brinkac, M. Gwinn, R.T. DeBoy, R. Madpu, S.C. Daugherty, A.S. Durkin, D.H. Haft, W.C. Nelson, J.D. Peterson, M. Pop, H.M. Khouri, D. Radune, J.L. Benton, Y. Mahamoud, L. Jiang, I.R. Hance, J.F. Weidman, K.J. Berry, R.D. Plaut, A.M. Wolf, K.L. Watkins, W.C. Nierman, A. Hazen, R. Cline, C. Redmond, J.E. Thwaite, O. White, S.L. Salzberg, B. Thomason, A.M. Friedlander, T.M. Koehler, P.C. Hanna, A.-B. Kolstø, and C.M. Fraser. The genome sequence of *Bacillus anthracis* Ames and comparison to closely-related bacteria. *Nature* 423 (2003), 81-86.
133. N. Volfovsky, B.J. Haas., and S.L. Salzberg. Computational discovery of internal micro-exons. *Genome Research* 13:6 (2003), 1216-1221.
134. W.H. Majoros, M. Pertea, C. Antonescu, and S.L. Salzberg. GlimmerM, Exonomy, and Unveil: Three *ab initio* eukaryotic gene finders. *Nucleic Acids Research*, 31 (2003), 3601-3604.
135. T.D. Read, G.S.A. Myers, R.C. Brunham, W.C. Nelson, I.T. Paulsen, J. Heidelberg, E. Holtzapple, H. Khouri, N.B. Federova, H.A. Carty, L.A. Umayam, D.H. Haft, J. Peterson, M.J. Beanan, O. White, S.L. Salzberg, R.-C. Hsia, G. McClarty, R.G. Rank, P.M. Bavoil, and C.M. Fraser. Genome sequence of *Chlamydia caviae* (*Chlamydia psittaci* GPIC): examining the role of niche-specific genes in the evolution of the Chlamydiaceae. *Nucleic Acids Research* 31:8 (2003), 2134-2147.

136. M.D. Ermolaeva, M. Wu, J.A. Eisen, and S.L. Salzberg. The age of the *Arabidopsis thaliana* genome duplication. *Plant Molecular Biology*, 51:6 (2003), 859-866.
137. A.L. Delcher, S.L. Salzberg, and A.M. Phillippy. Using MUMmer to identify similar regions in large sequence sets. In A. Baxevanis, D. Davison, R. Page, G. Petsko, G. Stormo, and L. Stein (eds.), *Current Protocols in Bioinformatics*, (2003), Chapter 10:Unit 10.3. New York: John Wiley and Sons.
138. M. Pertea and S.L. Salzberg. Using GlimmerM to find genes in eukaryotic genomes. In A. Baxevanis, D. Davison, R. Page, G. Petsko, G. Stormo, and L. Stein (eds.), *Current Protocols in Bioinformatics*, (2003), Unit 4.4. New York: John Wiley and Sons.
139. M.J. Gardner, N. Hall, E. Fung, O. White, M. Berriman, R. Hyman, J. Carlton, A. Pain, K. Nelson, S. Bowman, I. Paulsen, K. James, J.A. Eisen, K. Rutherford, S.L. Salzberg, A. Craig, S. Kyes, M.-S. Chan, V. Nene, S. Shallom, B. Suh, J. Peterson, S. Angiuoli, M. Pertea, J. Allen, J. Selengut, D. Haft, M. Mather, A. Vaidya, D. Martin, A. Fairlamb, M. Fraunholz, D. Roos, S. Ralph, G.I. McFadden, L.M. Cummings, M. Subramanian, C. Mungall, J.C. Venter, D.J. Carucci, S.L. Hoffman, C. Newbold, R.W. Davis, C.M. Fraser, and B.Barrell. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419 (2002), 498-511.
140. M.J. Gardner, S. Shallom, J.M. Carlton, S.L. Salzberg, V. Nene, A. Shoaibi, A. Ciecko, J. Lynn, M. Rizzo, B. Weaver, B. Jarrahi, M. Brenner, B. Parvizi, L. Tallon, A. Moazzez, D. Granger, C. Fujii, C. Hansen, T. Feldblyum, J. Peterson, B. Suh, S. Angiuoli, M. Pertea, J. Allen, J. Selengut, O. White, L.M. Cummings, H.O. Smith, M.D. Adams, J. C. Venter, D.J. Carucci, S.L. Hoffman, and C.M. Fraser. Sequence of *Plasmodium falciparum* chromosomes 2, 10, 11, and 14. *Nature* 419 (2002), 531-534.
141. R.A. Holt, G.M. Subramanian, A. Halpern, G.G. Sutton, R. Charlab, D.R. Nusskern, P. Wincker, A.G. Clark, J.M.C. Ribeiro, R. Wides, S.L. Salzberg, B. Loftus, M. Yandell, W.H. Majoros, D.B. Rusch, Z. Lai, C.L. Kraft, J.F. Abril, V. Anthouard, P. Arensburger, P.W. Atkinson, H. Baden, V. de Berardinis, D. Baldwin, V. Benes, J. Biedler, C. Blass, R. Bolanos, D. Boscus, M. Barnstead, S. Cai, A. Center, K. Chatuverdi, G.K. Christophides, M.A. Chrystal, M. Clamp, A. Cravchik, V. Curwen, A. Dana, A. Delcher, I. Dew, C. A. Evans, M. Flanigan, A. Grundschober-Freimoser, L. Friedli, Z. Gu, P. Guan, R. Guigo, M. E. Hillenmeyer, S. L. Hladun, J. R. Hogan, Y. S. Hong, J. Hoover, O. Jaillon, Z. Ke, C. Kodira, E. Kokoza, A. Koutsos, I. Letunic, A. Levitsky, Y. Liang, J. Lin, N. F. Lobo, J. R. Lopez, J. A. Malek, T. C. McIntosh, S. Meister, J. Miller, C. Mobarry, E. Mongin, S. D. Murphy, D. A. O'Brochta, C. Pfannkoch, R. Qi, M.A. Regier, K. Remington, H. Shao, M.V. Sharakhova, C.D. Sitter, J. Shetty, T.J. Smith, R. Strong, J. Sun, D. Thomasova, L.Q. Ton, P. Topalis, Z. Tu, M.F. Unger, B. Walenz, A. Wang, J. Wang, M. Wang, X. Wang, K.J. Woodford, J.R. Wortman, M. Wu, A. Yao, E.M. Zdobnov, H. Zhang, Q. Zhao, S. Zhao, S.C. Zhu, I. Zhimulev, M. Coluzzi, A. della Torre, C.W. Roth, C. Louis, F. Kalush, R.J. Mural, E.W. Myers, M.D. Adams, H.O. Smith, S. Broder, M.J. Gardner, C.M. Fraser, E. Birney, P. Bork, P.T. Brey, J.C. Venter, J. Weissenbach, F.C. Kafatos, F.H. Collins, and S.L. Hoffman. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 298 (2002), 129-149.
142. E.M. Zdobnov, C. von Mering, I. Letunic, D. Torrents, M. Suyama, R.R. Copley, G.K. Christophides, D. Thomasova, R.A. Holt, G.M. Subramanian, H.-M. Mueller, G. Dimopoulos, J.H. Law, M.A. Wells, E. Birney, R. Charlab, A.L. Halpern, E. Kokoza, C.L. Kraft, Z. Lai, S. Lewis, C. Louis, C. Barillas-Mury, D. Nusskern, G.M. Rubin, S.L. Salzberg, G.G. Sutton, P. Topalis, R. Wides, P. Wincker, M. Yandell, F.H. Collins, J. Ribeiro, W.M. Gelbart, F.C. Kafatos, and P. Bork. Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* 298 (2002), 149-159.
143. J.M. Carlton, S.V. Angiuoli, B.B. Suh, T.W. Kooij, J.C. Silva, M. Pertea, M.D. Ermolaeva, J.E. Allen, J.D. Selengut, H.L. Koo, J.D. Peterson, M. Pop, D.S. Kosack, M.F. Shumway, S.L. Bidwell, S.J. Shallom, S.E. van Aken, S.B. Riedmuller, T.V. Feldblyum, J.K. Cho, J. Quackenbush, M. Sedegah, A. Shoaibi, L.M. Cummings, L.Florens, J.R. Yates III, J.D. Raine, R.E. Sinden, M.A. Harris, D.A. Cunningham, P.R. Preiser, L.W. Bergman, A.B. Vaidya, L.H. van Lin, C.J. Janse, A.P. Waters, H.O. Smith, O.R. White, S.L. Salzberg, J.C. Venter, C.M. Fraser, S.L. Hoffman, M.J. Gardner, and D.J. Carucci. Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature* 419 (2002), 512-519.

144. I.T. Paulsen, R. Seshadri, K.E. Nelson, J.A. Eisen, J.F. Heidelberg, T.D. Read, R.J. Dodson, L. Umayam, L.M. Brinkac, M.J. Beanan, S.C. Dougherty, R.T. Deboy, R. Madupu, W.C. Nelson, B. Ayodeji, M. Kraul, J. Shetty, J. Malek, S.E. Van Aken, S. Riedmuller, H. Tettelin, S.R. Gill, O. White, S.L. Salzberg, D.L. Hoover, L.E. Lindler, S.M. Halling, S.M. Boyle, and C.M. Fraser. The *Brucella suis* genome reveals fundamental similarities between animal and plant pathogens and symbionts. *Proc. Natl. Acad. Sci. USA* 99 (2002), 13148-13153.
145. T.D. Read, S.L. Salzberg, M. Pop, M. Shumway, L. Umayam, L. Jiang, E. Holtzapple, J. Busch, K.L. Smith, J.M. Schupp, D. Solomon, P. Keim, and C.M. Fraser. Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. *Science* 296 (2002), 2028-2033.
146. R.D. Fleischmann, D. Alland, J.A. Eisen, L. Carpenter, O. White, J. Peterson, R. DeBoy, R. Dodson, M. Gwinn, D. Haft, E. Hickey, J.F. Kolonay, W.C. Nelson, L.A. Umayam, M. Ermolaeva, S.L. Salzberg, A. Delcher, T. Utterback, J. Weidman, H. Khouri, J. Gill, A. Mikula, W. Bishai, W.R. Jacobs, J.C. Venter, and C.M. Fraser. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J. Bacteriology* 184:19 (2002), 5479-5490.
147. C.M. Fraser, J.A. Eisen, K.E. Nelson, I.T. Paulsen, and S.L. Salzberg. The value of complete microbial genome sequencing (you get what you pay for). *J. Bacteriology*, 184:23 (2002), 6403-6405.
148. E. Willerslev, T. Mourier, A.J. Hansen, B. Christensen, I. Barnes, and S.L. Salzberg. Contamination in the draft of the human genome masquerades as lateral gene transfer. *DNA Sequence* 13:2 (2002), 75-76.
149. M. Pop, S.L. Salzberg, and M. Shumway. Genome sequence assembly: algorithms and issues. *IEEE Computer* 35:7 (2002), 47-54.
150. A.L. Delcher, A. Phillippy, J. Carlton, and S.L. Salzberg. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Research* 30:11 (2002), 2478-2483.
151. B.J. Haas, N. Volfovsky, C.D. Town, M. Troukhan, N. Alexandrov, K.A. Feldmann, R.B. Flavell, O. White, and S.L. Salzberg. Full-length messenger RNA sequences greatly improve genome annotation. *Genome Biology* 3:6 (2002), research0029.1-0029.12. Highly accessed.
152. R. J. Mural, M.D. Adams, E.W. Myers, H.O. Smith, G.L.G. Miklos, R. Wides, A. Halpern, P.W. Li, G. Sutton, J. Nadeau, S.L. Salzberg, R.A. Holt, C.D. Kodira, F. Lu, L. Chen, Z. Deng, C.C. Evangelista, W. Gan, T.J. Heiman, J. Li, Z. Li, G.V. Merkulov, N.V. Milshina, A.K. Naik, R. Qi, B.C. Shue, A. Wang, J. Wang, X. Wang, X. Yan, J. Ye, S. Yooseph, Q. Zhao, L. Zheng, S.C. Zhu, K. Biddick, R. Bolanos, A.L. Delcher, I.M. Dew, D. Fasulo, M.J. Flanigan, D.H. Huson, S.A. Kravitz, J.R. Miller, C.M. Mobarry, K. Reinert, K.A. Remington, Q. Zhang, X.H. Zheng, D.R. Nusskern, Z. Lai, Y. Lei, W. Zhong, A. Yao, P. Guan, R. Ji, Z. Gu, Z. Wang, F. Zhong, C. Xiao, C. Chiang, M. Yandell, J.R. Wortman, P.G. Amanatides, S.L. Hladun, E.C. Pratts, J.E. Johnson, K.L. Dodson, K.J. Woodford, C.A. Evans, B. Gropman, D.B. Rusch, E. Venter, M. Wang, T.J. Smith, J.T. Houck, D.E. Tompkins, C. Haynes, D. Jacob, S.H. Chin, D.R. Allen, C.E. Dahlke, R. Sanders, K. Li, X. Liu, A.A. Levitsky, W.H. Majoros, Q. Chen, A.C. Xia, J.R. Lopez, M.T. Donnelly, M.H. Newman, A. Glodek, C.L. Kraft, M. Nodell, F. Ali, H. An, D. Baldwin-Pitts, K.Y. Beeson, S. Cai, M. Carnes, A. Carver, P.M. Caulk, A. Center, Y. Chen, M. Cheng, M.D. Coyne, M. Crowder, S. Danaher, L.B. Davenport, R. Desilets, S.M. Dietz, L. Doup, P. Dullaghan, S. Ferriera, C.R. Fosler, H.C. Gire, A. Gluecksmann, J.D. Gocayne, J. Gray, B. Hart, J. Haynes, J. Hoover, T. Howland, C. Ibegwam, M. Jalali, D. Johns, L. Kline, D.S. Ma, S. MacCawley, A. Magoon, F. Mann, D. May, T.C. McIntosh, S. Mehta, L. Moy, M.C. Moy, B. J. Murphy, S.D. Murphy, K.A. Nelson, Z. Nuri, K.A. Parker, A.C. Prudhomme, V.N. Puri, H. Qureshi, J.C. Raley, M.S. Reardon, M.A. Regier, Y.C. Rogers, D.L. Romblad, J. Schutz, J.L. Scott, R. Scott, C.D. Sitter, M. Smallwood, A.C. Sprague, E. Stewart, R.V. Strong, E. Suh, K. Sylvester, R. Thomas, N.N. Tint, C. Tsonis, G. Wang, G. Wang, M.S. Williams, S.M. Williams, S.M. Windsor, K. Wolfe, M.M. Wu, J. Zaveri, K. Chaturvedi, A.E. Gabrielian, Z. Ke, J. Sun, G. Subramanian, and J.C. Venter. A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome. *Science* 296 (2002), 1661-1671.
153. M. Pertea and S.L. Salzberg. Computational gene finding in plants. *Plant Molecular Biology* 48:1-2 (2002), 39-48.

154. V. Nene, R. Bishop, J. Quackenbush, M. Pertea, S.L. Salzberg, E. Taracha, S. Morzaria, C.M. Fraser, and M. Gardner. Genomics of *Theileria parva*. In *Theileria* (D.A.E. Dobbelaere and D.J. McKeever, Eds.). Boston: Kluwer Academic Publishers, 2002.
155. S.L. Salzberg, O. White, J. Peterson, and J.A. Eisen. Microbial genes in the human genome: lateral transfer or gene loss? *Science* 292 (2001), 1903–1906.
156. J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans, R.A. Holt, J.D. Gocayne, P. Amanatides, R.M. Ballew, D.H. Huson, J.R. Wortman, Q. Zhang, C.D. Kodira, X.H. Zheng, L. Chen, M. Skupski, G. Subramanian, P.D. Thomas, J. Zhang, G.L.G. Miklos, C. Nelson, S. Broder, A.G. Clark, J. Nadeau, V.A. McKusick, N. Zinder, A.J. Levine, R.J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A.L. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A.E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T.J. Heiman, M.E. Higgins, R.-R. Ji, Z. Ke, K.A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G.V. Merkulov, N. Milshina, H.M. Moore, A.K. Naik, V.A. Narayan, B. Neelam, D. Nusskern, D.B. Rusch, S.L. Salzberg, W. Shao, B. Shue, J. Sun, Z.Y. Wang, A. Wang, X. Wang, J. Wang, M.-H. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S.C. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Amy Carver, A. Center, M. Cheng, L. Curry, S.L. Davenport, R. Desilets, S.K. Dodson, L. Doup, S. Ferreira, N. Garg, A. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hostin, J. Houck, T. Howland, C.J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S.T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Cynthia Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Rodriguez, Y.-H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. N.N. Tint, S. Tse, C. Vech, G.J. Wetter, S.M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J.F. Abril, R. Guigó, M. J. Campbell, K.B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J.P. Caulk, Y.-H. Chiang, M. Coyne, C. Dahlke, A. Deslattes Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, and X. Zhu. The Sequence of the Human Genome. *Science* 291 (2001), 1304–1351.
157. S. Kennedy, W.V. Ng, S.L. Salzberg, L. Hood, and S. DasSarma. Understanding the adaptation of *Halobacterium* species NRC-1 to its extreme environment through computational analysis of its genome sequence. *Genome Research* 11:10 (2001), 1641–1650.
158. B.E. Suzek, M.D. Ermolaeva, M. Schreiber, and S.L. Salzberg. A probabilistic method for identifying start codons in bacterial genomes. *Bioinformatics* 17:12 (2001), 1123–1130.
159. N. Volfovsky, B.J. Haas, and S.L. Salzberg. A clustering method for repeat analysis in DNA sequences. *Genome Biology* 2:8 (2001), research 0027.1–11. **Highly accessed.**
160. Q. Yuan, J. Quackenbush, R. Sultana, M. Pertea, S.L. Salzberg, and C.R. Buell. Rice bioinformatics: analysis of rice sequence data and leveraging the data to other plant species. *Plant Physiology* 125:3 (2001), 1166–74.
161. H. Tettelin, K.E. Nelson, I.T. Paulsen, J.A. Eisen, T.D. Read, S. Peterson, J. Heidelberg, R.T. DeBoy, D.H. Haft, R.J. Dodson, A.S. Durkin, M.L. Gwinn, J.F. Kolonay, W.C. Nelson, J.D. Peterson, L.A. Umayam, O. White, S.L. Salzberg, M.R. Lewis, D. Radune, E. Holtzapple, H. Khouri, A.M. Wolf, T.R. Utterback, C.L. Hansen, L.A. McDonald, T.V. Feldblyum, S. Angiuoli, T. Dickinson, E.K. Hickey, I.E. Holt, B.J. Loftus, F. Yang, H.O. Smith, J.C. Venter, B.A. Dougherty, D.A. Morrison, S.K. Hollingshead, and C.M. Fraser.

- Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* 293 (2001), 498–506.
162. W.C. Nierman, T.V. Feldblyum, M.T. Laub, I.T. Paulsen, K.E. Nelson, J.A. Eisen, J.F. Heidelberg, M.R.K. Alley, N. Ohta, J.R. Maddock, I. Potocka, W.C. Nelson, A. Newton, C. Stephens, N.D. Phadke, B. Ely, R.T. DeBoy, R.J. Dodson, A.S. Durkin, M.L. Gwinn, D.H. Haft, J.F. Kolonay, J. Smit, M.B. Craven, H. Khouri, J. Shetty, K. Berry, T. Utterback, K. Tran, A. Wolf, J. Vamathevan, M.D. Ermolaeva, O. White, S.L. Salzberg, J.C. Venter, L. Shapiro, and C.M. Fraser. Complete genome sequence of *Caulobacter crescentus*. *Proc. Natl. Acad. Sci. USA* 98:7 (2001), 4136–4141.
 163. M. Bevan, K. Mayer, O. White, J.A. Eisen, D. Preuss, T. Bureau, S.L. Salzberg, and H. Mewes. Sequence and analysis of the Arabidopsis genome. *Current Opinion Plant Biology* 4:2 (2001), 105–110.
 164. M. Ermolaeva, O. White, and S.L. Salzberg. Prediction of operons in microbial genomes. *Nucleic Acids Research*, 29:5 (2001), 1216–1221.
 165. M. Pertea, X. Lin, and S.L. Salzberg. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Research*, 29:5 (2001), 1185–1190.
 166. F. Liang, I.E. Holt, G. Pertea, S. Karamycheva, S.L. Salzberg, and J. Quackenbush. Gene index analysis of the human genome estimates approximately 120,000 genes. *Nature Genetics* 25:2 (2000), 239–240. Correction [to 56,000 genes]: *Nature Genetics* 26:4 (2000), 501.
 167. J.A. Eisen, J.F. Heidelberg, O. White, S.L. Salzberg, Evidence for symmetric chromosomal inversions around the replication origin in bacteria, *Genome Biology* 1:6 (2000): research0011.1–0011.9.
 168. The Arabidopsis Genome Initiative (143 authors). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408 (2000), 796–815.
 169. European Union Chromosome 3 Arabidopsis Sequencing Consortium, The Institute for Genomic Research, and Kazusa DNA Research Institute. Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*. *Nature* 408 (2000), 820–823.
 170. Theologis, J.R. Ecker, C.J. Palm, N.A. Federspiel, S. Kaul, O. White, J. Alonso, H. Altfati, R. Araujo, C.L. Bowman, S.Y. Brooks, E. Buehler, A. Chan, Q. Chao, H. Chen, R.F. Cheuk, C.W. Chin, M.K. Chung, L. Conn, A.B. Conway, A.R. Conway, T.H. Creasy, K. Dewar, P. Dunn, P. Egtu, T.V. Feldblyum, J. Feng, B. Fong, C.Y. Fujii, J.E. Gill, J.L. Hunter, J. Jenkins, C. Johnson-Hopson, S. Khan, E. Khaykin, C.J. Kim, H.L. Koo, I. Kremenetskaia, D.B. Kurtz, A. Kwan, B. Lam, S. Langin-Hooper, A. Lee, J.M. Lee, C.A. Lenz, J.H. Li, Y. Li, X. Lin, S.X. Liu, Z.A. Liu, J.S. Iuros, R. Maiti, A. Marziali, J. Militscher, M. Miranda, M. Nguyen, W.C. Nierman, B.I. Osborne, G. Pai, J. Peterson, P.K. Pham, M. Rizzo, T. Rooney, D. Rowley, H. Sakano, S.L. Salzberg, J.R. Schwartz, P. Shinn, A.M. Southwick, H. Sun, L.J. Tallon, G. Tambunga, M. J. Toriumi, C.D. Town, T. Utterback, S. Van Aken, M. Vaysberg, V.S. Vysotskaia, M. Walker, D. Wu, G. Yu, C.M. Fraser, J.C. Venter, R.W. Davis. Sequence and analysis of chromosome 1 of the plant *Arabidopsis thaliana*. *Nature* 408, 816–820 (2000).
 171. J.F. Heidelberg, J.A. Eisen, W.C. Nelson, R.A. Clayton, M.L. Gwinn, R.J. Dodson, D.H. Haft, E.K. Hickey, J.D. Peterson, L. Umayam, S.R. Gill, K.E. Nelson, T.D. Read, H. Tettelin, D. Richardson, M.D. Ermolaeva, J. Vamathevan, S. Bass, H. Qin, H. Dragoi, P. Sellers, L. McDonald, T. Utterback, R.D. Fleishmann, W.C. Nierman, O. White, S.L. Salzberg, H.O. Smith, R. R. Colwell, J.J. Mekalanos, J.C. Venter, and C.M. Fraser. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 406, 477–483 (2000).
 172. C.M. Fraser, J.A. Eisen, and S.L. Salzberg. Microbial genome sequencing. *Nature* 406 (2000), 799–803.
 173. F. Liang, I.E. Holt, G. Pertea, S. Karamycheva, S.L. Salzberg, and J. Quackenbush. An optimized protocol for analysis of EST sequences. *Nucleic Acids Research* 28:18 (2000), 3657–3665.
 174. M.D. Ermolaeva, H. Khalak, O. White, H.O. Smith, and S.L. Salzberg. Prediction of transcription terminators in bacterial genomes. *Journal of Molecular Biology* (2000), 301, 27–33.
 175. M. Pertea, S.L. Salzberg, and M.J. Gardner. Finding genes in Plasmodium falciparum chromosome 3. *Nature* 404 (2000), 34.
 176. T.D. Read, R.C. Brunham, C. Shen, S.R. Gill, J.F. Heidelberg, O. White, E.K. Hickey, J. Peterson, T. Utterback, K. Berry, S. Bass, K. Linher, J. Weidman, H. Khouri, B. Craven, C. Bowman, R. Dodson, M.

- Gwinn, W. Nelson, R. DeBoy, J. Kolonay, G. McClarty, S.L. Salzberg, J. Eisen, C. M. Fraser. Genome sequences of *Chlamydia trachomatis* MoPn and *C. pneumoniae* AR39. *Nucleic Acids Research*, 28:6 (2000), 1397–1406.
177. X. Lin, S. Kaul, S. Rounsley, T.P. Shea, M.-I. Benito, C.D. Town, C.Y. Fujii, T. Mason, C.L. Bowman, M. Barnstead, T. Feldblyum, C.R. Buell, K.A. Ketchum, C.M. Ronning, H. Koo, K. Moffat, L. Cronin, M. Shen, G. Pai, S. Van Aken, L. Umayam, L. Tallon, J. Gill, M.D. Adams, A.J. Carrera, T.H. Creasy, H.M. Goodman, C.R. Somerville, G. Copenhagen, D. Preuss, W.C. Nierman, O. White, J.A. Eisen, S. Salzberg, C. Fraser, and J.C. Venter. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* 402 (1999), 761–768.
 178. H. Tettelin, N.J. Saunders, J. Heidelberg, A.C. Jeffries, K.E. Nelson, J.A. Eisen, K.A. Ketchum, D.W. Hood, J.F. Peden, R.J. Dodson, W.C. Nelson, M.L. Gwinn, R. DeBoy, J.D. Peterson, E.K. Hickey, D.H. Haft, S.L. Salzberg, O. White, R.D. Fleischmann, B.A. Dougherty, T. Mason, A. Ciecko, D.S. Parksey, E. Blair, H. Cittone, E.B. Clark, M.D. Cotton, T.R. Utterback, H. Khouri, H. Qin, J. Vamathevan, J. Gill, V. Scarlato, V. Masignani, M. Pizza, G. Grandi, L. Sun, H.O. Smith, C.M. Fraser, E.R. Moxon, R. Rappuoli, and J.C. Venter. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* 287 (2000), 1809–1815.
 179. H. Tettelin, D. Radune, S. Kasif, H. Khouri, and S.L. Salzberg. Optimized multiplex PCR: Efficiently closing a whole-genome shotgun sequencing project. *Genomics* 62:3 (1999), 500–507.
 180. O. White, J.A. Eisen, J.F. Heidelberg, E.K. Hickey, J.D. Peterson, R.J. Dodson, D.H. Haft, M.L. Gwinn, W.C. Nelson, D.L. Richardson, K.S. Moffat, H. Qin, L. Jiang, W. Pamphile, M. Crosby, M. Shen, J.J. Vamathevan, P. Lam, L. McDonald, T.R. Utterback, C. Zalewski, K.S. Makarova, L. Aravind, M.J. Daly, K.W. Minton, R.D. Fleischmann, K.A. Ketchum, K.E. Nelson, S.L. Salzberg, H.O. Smith, J.C. Venter, and C.M. Fraser. Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. *Science* 286 (1999), 1571–1577.
 181. H.O. Smith, M.L. Gwinn, and S.L. Salzberg. DNA uptake signal sequences in naturally transformable bacteria. *Research in Microbiology*, 150 (1999), 603–616.
 182. A.L. Delcher, D. Harmon, S. Kasif, O. White, and S.L. Salzberg. Improved microbial gene identification with Glimmer. *Nucleic Acids Research*, 27 (1999), 4636–4641.
 183. S.L. Salzberg, M. Pertea, A.L. Delcher, M.J. Gardner, and H. Tettelin. Interpolated Markov models for eukaryotic gene finding. *Genomics*, 59 (1999), 24–31.
 184. S.L. Salzberg. Gene discovery in DNA sequences. *IEEE Intelligent Systems*, 14:6 (1999), 44–48.
 185. A.L. Delcher, S. Kasif, R.D. Fleischmann, J. Peterson, O. White, and S.L. Salzberg. Alignment of whole genomes. *Nucleic Acids Research*, 27:11 (1999), 2369–2376.
 186. K.E. Nelson, R.A. Clayton, S.R. Gill, M.L. Gwinn, R.J. Dodson, D.H. Haft, E.K. Hickey, J.D. Peterson, W.C. Nelson, K.A. Ketchum, L. McDonald, T.R. Utterback, J.A. Malek, K.D. Linher, M.M. Garrett, A.M. Stewart, M.D. Cotton, M.S. Pratt, C.A. Phillips, D. Richardson, J. Heidelberg, G.G. Sutton, R.D. Fleischmann, J.A. Eisen, O. White, S.L. Salzberg, H.O. Smith, J.C. Venter and C.M. Fraser. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399 (1999), 323–329.
 187. M.J. Gardner, H. Tettelin, D.J. Carucci, L.M. Cummings, L. Aravind, E.V. Koonin, S. Shallom, T. Mason, K. Yu, C. Fujii, J. Pederson, K. Shen, J. Jing, C. Aston, Z. Lai, D.C. Schwartz, M. Pertea, S.L. Salzberg, L. Zhou, G.G. Sutton, R. Clayton, O. White, H.O. Smith, C.M. Fraser, M.D. Adams, J.C. Venter, and S.L. Hoffman. Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science* 282 (1998), 1126–1132.
 188. S.L. Salzberg, A.J. Salzberg, A.R. Kerlavage, and J.-F. Tomb. Skewed oligomers and origins of replication. *Gene* 217:1-2 (1998), 57–67.
 189. S.L. Salzberg, A.L. Delcher, K. Fasman, and J. Henderson. A Decision Tree System for Finding Genes in DNA. *Journal of Computational Biology* 5:4 (1998), 667–680.
 190. C.M. Fraser, S.J. Norris, G.M. Weinstock, O. White, G.G. Sutton, R.A. Clayton, R.J. Dodson, M.L. Gwinn, E. Hickey, K.A. Ketchum, E. Sodergren, J. Hardham, M. McLeod, S.L. Salzberg, H. Khalak, J. Weidman, J.K. Howell, M. Chidambaram, T. Utterback, L. Watthey, L. McDonald, P. Artiach, C.

- Bowman, S. Garland, C. Fujii, M.D. Cotton, K. Horst, K. Roberts, B. Hatch, H.O. Smith, and J.C. Venter. Complete genomic sequence of *Treponema pallidum*, the syphilis spirochete. *Science* 281 (1998), 375–388.
191. S.L. Salzberg, A.L. Delcher, S. Kasif, and O. White. Microbial gene identification using interpolated Markov models. *Nucleic Acids Research*, 26:2 (1998), 544–548.
 192. S.L. Salzberg. A tutorial introduction to computation for biologists. In: S.L. Salzberg, D. Searls, and S. Kasif (eds.), *Computational Methods in Molecular Biology*, pp. 11–27. Amsterdam: Elsevier Science B.V., 1998.
 193. K. Fasman and S.L. Salzberg. An introduction to biological sequence analysis. In S.L. Salzberg, D. Searls, and S. Kasif (eds.), *Computational Methods in Molecular Biology*, pp. 29–42. Amsterdam: Elsevier Science B.V., 1998.
 194. S.L. Salzberg. Decision trees and Markov chains for gene finding. In S.L. Salzberg, D. Searls, and S. Kasif (eds.), *Computational Methods in Molecular Biology*, pp. 187–203. Amsterdam: Elsevier Science B.V., 1998.
 195. S. Kasif, S.L. Salzberg, D. Waltz, J. Rachlin, and D. Aha. A probabilistic framework for memory-based reasoning. *Artificial Intelligence*, 104 (1998), 287–311.
 196. C.M. Fraser, S. Casjens, W. Huang, G. Sutton, R. Clayton, R. Lathigra, O. White, K. Ketchum, R. Dodson, E. Hickey, M. Gwinn, B. Dougherty, J.-F. Tomb, R. Fleischmann, D. Richardson, J. Peterson, A. Kerlavage, J. Quackenbush, S.L. Salzberg, M. Hanson, R. van Vugt, N. Palmer, M. Adams, J. Gocayne, J. Weidman, T. Utterback, L. Wathley, L. McDonald, P. Artiach, C. Bowman, S. Garland, C. Fujii, M. Cotton, K. Horst, K. Roberts, B. Hatch, H.O. Smith, and J.C. Venter. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* 390 (1997), 580–586.
 197. J. Henderson, S.L. Salzberg, and K. Fasman. Finding genes in DNA with a hidden Markov model. *Journal of Computational Biology* 4:2 (1997), 127–141.
 198. S.L. Salzberg. A method for identifying splice sites and translational start sites in eukaryotic mRNA. *Computer Applications in the Biosciences (CABIOS)* 13:4 (1997), 365–376.
 199. S.L. Salzberg. On comparing classifiers: Pitfalls to avoid and a recommended approach. *Data Mining and Knowledge Discovery*, 1:3 (1997), 317–327.
 200. J. Sheppard and S.L. Salzberg. A teaching strategy for memory-based control. *Artificial Intelligence Review* 11:5 (1997), 343–370.
 201. E. Arkin, P. Belleville, J. Mitchell, D. Mount, K. Romanik, S.L. Salzberg, and D. Souvaine. Testing simple polygons. *Computational Geometry: Theory and Applications* 8:2 (1997), 97–114.
 202. D. Heath, S. Kasif, and S.L. Salzberg. Committees of decision trees. In B. Gorayska and J. Mey (Eds.), *Cognitive Technology: In Search of a Humane Interface* (pp. 305–317). Amsterdam: Elsevier Science B.V., 1996.
 203. D. Heath, S. Kasif, R. Kosaraju, S.L. Salzberg, and G. Sullivan. Learning nested concept classes with limited storage. *Journal of Theoretical and Experimental Artificial Intelligence* 8 (1996), 129–147.
 204. S.L. Salzberg. Locating protein coding regions in human DNA using a decision tree algorithm. *Journal of Computational Biology* 2:3 (1995), 473–485.
 205. S.L. Salzberg, R. Chandar, H. Ford, S. Murthy, and R. White. Decision trees for automated identification of cosmic ray hits in Hubble Space Telescope images. *Publications of the Astronomical Society of the Pacific* 107 (1995), 1–10.
 206. S.L. Salzberg, A.L. Delcher, D. Heath, and S. Kasif. Best-case results for nearest-neighbor learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17:6 (1995), 599–608.
 207. K. Romanik and S.L. Salzberg. Testing orthogonal shapes. *Computational Geometry: Theory and Applications* 5 (1995), 33–49.
 208. S.K. Murthy, S. Kasif, and S.L. Salzberg. A system for induction of oblique decision trees. *Journal of Artificial Intelligence Research* 2:1 (1994), 1–32.
 209. S. Cost and S.L. Salzberg. A weighted nearest neighbor algorithm for learning with symbolic features. *Machine Learning* 10:1 (1993), 57–78.

210. D. Aha and S.L. Salzberg. Learning to catch: Applying nearest neighbor algorithms to dynamic control tasks. In P. Cheeseman & R.W. Oldford (Eds.), *Selecting Models from Data: Artificial Intelligence and Statistics IV*. New York: Springer-Verlag, 1993.
211. S.L. Salzberg and S. Cost. Predicting protein secondary structure with a nearest-neighbor algorithm. *Journal of Molecular Biology* 227 (1992), 371–374.
212. S.L. Salzberg. A nearest hyperrectangle learning method. *Machine Learning* 6 (1991), 251–276.
213. S.L. Salzberg and M. Watkins. Managing Information for Concurrent Engineering: Challenges and Barriers. *Research in Engineering Design* 2:1 (1990), 35–52.
214. S.L. Salzberg. Pinpointing Good Hypotheses with Heuristics. In W. Gale (ed.), *Artificial Intelligence and Statistics*, Reading, MA: Addison-Wesley, 1986, 133–158.
215. M.J. Fischer and S.L. Salzberg. Finding a Majority Among N Votes. *Journal of Algorithms* 3 (1982), 375–379.

Refereed conference publications (acceptance rate < 25%)

216. S.K. Murthy and S.L. Salzberg. On Lookahead for Decision Tree Induction. *Proc. 14th Internatl. Joint Conf. Artificial Intelligence (IJCAI-95)* (pp. 1025–1031), Montreal, Canada, 1995.
217. T. Fulton, S. Kasif, and S.L. Salzberg. An Efficient Algorithm for Finding Multi-way Splits in Decision Trees. *Proc. 12th Internatl. Conf. Machine Learning*, (pp. 244–251), Tahoe City, California, 1995. Morgan Kaufmann Publishers.
218. J. Sheppard and S.L. Salzberg. Combining Memory-Based Reasoning with Genetic Algorithms. In *Proc. 6th Internatl. Conf. Genetic Algorithms* (pp. 452–459), L.J. Eshelman (ed.). Morgan Kaufmann Publishers.
219. S.K. Murthy and S.L. Salzberg. Decision Tree Induction: How Effective is the Greedy Heuristic? *Proc. First Internatl. Conf. Knowledge Discovery and Data Mining (KDD-95)* (pp. 222–227), Montreal, Canada, 1995.
220. J. Rachlin, S. Kasif, S.L. Salzberg, and D. Aha. Towards a Better Understanding of Memory-Based Reasoning Systems. *Proc. 11th Internatl. Conf. on Machine Learning* (pp. 242–250). New Brunswick, NJ, 1994.
221. D. Heath, S. Kasif, and S.L. Salzberg. Learning Oblique Decision Trees. *Proc. 13th Internatl. Joint Conf. on Artificial Intelligence (IJCAI-93)* (pp. 1002–1007). Chambéry, France, 1993.
222. S.K. Murthy, S. Kasif, S.L. Salzberg, and R. Beigel. OC1: Randomized Induction of Oblique Decision Trees. *Proc. 11th Natl. Conf. on Artificial Intelligence (AAAI-93)* (pp. 322–327), Washington, D.C., 1993.
223. S.L. Salzberg. Distance Metrics for Instance-Based Learning. *Methodologies for Intelligent Systems: 6th Internatl. Symposium (ISMIS '91)*, Z. Ras and M. Zemankova (eds.), pp. 399–408. New York: Springer-Verlag, 1991.
224. S.L. Salzberg, A.L. Delcher, D. Heath, and S. Kasif. Learning with a Helpful Teacher. *Proc. 12th Internatl. Joint Conf. on Artificial Intelligence (IJCAI-91)* (pp. 705–711). Sydney, Australia, 1991.
225. Heath, D, S. Kasif, R. Kosaraju, S.L. Salzberg, and G. Sullivan. Learning Nested Concept Classes with Limited Storage. *Proc. 12th Internatl. Joint Conf. On Artificial Intelligence (IJCAI-91)* (pp. 777–782). Sydney, Australia, 1991.
226. S.L. Salzberg. Heuristics for Inductive Learning. *Proc. 9th Internatl. Joint Conf. on Artificial Intelligence (IJCAI-85)*, Los Angeles, CA, 1985, 602–609.
227. S.L. Salzberg and D. Atkinson. Learning by Building Causal Explanations. *Proc. 6th European Conf. on Artificial Intelligence*, Pisa, Italy, 1984.
228. S.L. Salzberg. Generating Hypotheses to Explain Prediction Failures. *Proc. 3rd Natl. Conf. on Artificial Intelligence (AAAI-83)*, Washington, D.C., 1983, 352–355.

Other publications

229. D.B. Neale, C.H. Langley, S.L. Salzberg, and J.L. Wegrzyn. Open access to tree genomes: the path to a better forest. *Genome Biology* 2013, 14:120. doi:10.1186/gb4102.

230. Steven L. Salzberg. NIH funding: It does support innovators. Letter to *Nature* 493 (3 January 2013), 26.
231. Steven Salzberg. A 'Triumph' of Hype Over Reality. *The Atlantic*, online Forum, June 16, 2011.
<http://www.theatlantic.com/life/archive/2011/06/a-triumph-of-hype-over-reality/240464/>.
232. Steven L. Salzberg. No one should own your genes: patents on human genes stifle science and innovation. Editorial, *The Baltimore Sun*, November 10, 2010.
233. Steven L. Salzberg. Creationism Slips Into a Peer-Reviewed Journal. *Reports of the National Center for Science Education* 28:3, May-June 2008, 12-19.
234. Steven Salzberg, Elodie Ghedin, and David Spiro. Shared data are key to beating threat from flu. Letter to *Nature* 440 (30 Mar 2006), 605.
235. Steven Salzberg. Bird flu, Bush, evolution – and us. Editorial, *The Philadelphia Inquirer*, p. A16, Nov. 2, 2005.
236. Steven Salzberg. Journal Club review, *Nature* 438 (10 Nov 2005), 133.
237. J. Dunning Hotopp, W.C. Nelson and S.L. Salzberg. Response to “New names for old strains? *Wolbachia* wSim is actually wRi, I. Iturbe-Ormaetxe, M. Riegler, and S.L. O’Neill. *Genome Biology* 2005, 6:401. Online at <http://genomebiology.com/2005/6/7/401>.
238. Journal Editors and Authors group. Statement on Scientific Publication and Security. *Science* 299:5610 (21 February 2003), 1149. Published simultaneously in: *Nature* 421, 771 (20 February 2003) and *Proc. Natl. Acad. Sci.* (18 February 2003).
239. S.L. Salzberg and J.A. Eisen. Response to “Lateral Gene Transfer or Viral Colonization,” (letter by V. DeFilippis and L.P. Villarreal), *Science* 293 (2001), 1048.
240. S.L. Salzberg. Review of *Computational Molecular Biology: An Algorithmic Approach* by P.A. Pevzner. *Quarterly Review of Biology* 76:4 (2001), 485-6.
241. S.L. Salzberg. Review of *C4.5: Programs for Machine Learning* by J.R. Quinlan. *Machine Learning* 16 (1994), 235–240.
242. S.L. Salzberg. Machine learning moves out of the lab. *AI Expert* 3:2 (1988), 44-54.
243. S.L. Salzberg. Real world knowledge representation. *AI Expert* 2:8 (1987), 32-41.

Books

1. S.L. Salzberg and T. Warnow (editors). *Algorithms in Bioinformatics: 9th International Workshop, WABI 2009*. ISBN: 978-3-642-04240-9. Berlin: Springer-Verlag, 2009.
2. S.L. Salzberg, D. Searls, and S. Kasif (editors). *Computational Methods in Molecular Biology*. Amsterdam: Elsevier Science B.V., 1998.
3. S.L. Salzberg. *Learning with Nested Generalized Exemplars*. Norwell, MA: Kluwer Academic Publishers, 1990.

Blogs

- <http://www.forbes.com/sites/stevensalzberg> (A featured writer at *Forbes* magazine, a major business-oriented publisher, since January 2010)
- <http://genome.fieldofscience.com> (since 2007)

Research Grants (PI is Salzberg unless otherwise indicated)

4/1/2014-3/31/2018 National Institutes of Health, NIGMS, “Bioinformatics software for analyzing microbial genomes,” R01 GM083873-10, \$972,000 (\$600,000 direct costs).

9/21/2011-8/31/2014 National Institutes of Health, NHGRI, "Computational gene modeling and genome sequence assembly", R01 HG006677, \$1,910,827 (\$469,084 direct costs, year 1).

9/1/2013-8/31/2017 National Science Foundation, "Sequencing of the *Aegilops tauschii* genome." NSF IOS-1238231, \$897,564 (\$170,740 direct costs, year 1; \$554,052 direct costs in JHU subcontract. Overall award: \$8,959,701). **PI:** Jan Dvorak, UC Davis.

7/6/2011-4/30/2014 National Institutes of Health, NHGRI, "Alignment Software for Second-Generation Sequencing," R01 HG006102, \$2,122,212 (\$1,459,353 direct costs).

2/1/2011-1/31/2016 U.S. Dept. of Agriculture, National Institute of Food and Agriculture, "Loblolly Pine Genome Project, Award 2011-67009-30030, \$1,981,737. (Overall award: \$14,994,424) **PI:** David Neale, UC Davis.

3/25/08-2/28/13 National Institutes of Health, "Bioinformatics software for analyzing microbial genomes," R01 GM083873, \$1,107,000 (\$738,000 direct costs).

2/1/09-1/31/12 U.S. Department of Agriculture National Research Initiative, "Assembly and Annotation of Agricultural Genomes," 2008-04049, \$906,098.

6/1/07-9/20/11 National Institutes of Health, "Computational gene modeling and genome sequence assembly," R01 LM006845, \$2,398,178.

7/1/09-5/31/2012 National Institutes of Health, "Computational gene modeling and genome sequence assembly," R01 LM006845-Supplement, \$250,000 (\$167,333 direct costs). This ARRA supplement adds support for one postdoctoral fellow to R01 LM006845.

9/24/08-7/31/11 National Institutes of Health, "Assembly and analysis software for exploring the human microbiome," R01 HG004885, \$891,000. **PI:** Mihai Pop.

10/1/09-9/30/2011 National Institutes of Health, "SciBay: A New Methodology for Scientific Collaboration and Gene Function Determination," RC2-GM092602, \$305,995 (subcontract). **PI:** Martin Steffen, Boston University.

4/1/09-3/31/14 National Institutes of Health, "Genomic Sequencing Centers for Infectious Diseases," HHSN2722009000009C, \$471,186 (subcontract). **PI:** Claire Fraser-Liggett, Univ. of Maryland School of Medicine.

12/29/06-1/30/09 Department of Homeland Security, "Bioinformatics tools for Rapid Pathogen Detection and Analysis," award NBCH207002, \$4,973,420.

4/1/09-3/31/11 National Science Foundation, "Algorithms for the analysis of data from massively-parallel genome sequencing." IIS-0844494, \$379,919. **PI:** Mihai Pop.

6/1/2008-5/31/2009 Burroughs Wellcome Fund (subcontract to University of Pittsburgh), "Brugia Malayi: Proposal to Complete Assembly and Annotation of the Genome," \$40,000. **PI:** Elodie Ghedin.

5/4/05-9/3/06 Homeland Security Advanced Research Projects Agency (HSARPA), "Bioinformatics tools for Rapid Pathogen Detection and Analysis," award W81XWH-05-2-0051, \$2,094,428.

1/1/07-12/31-09 Henry M. Jackson Foundation (Naval Medical Research Center), "Assembly and gene finding algorithms for genome sequences generated by pyrosequencing," \$1,154,879. **PI:** Mihai Pop.

9/30/02-9/29/06 National Institutes of Health, "Computational gene modeling and genome sequence assembly," R01 LM06845-04, \$2,667,000.

9/1/03-2/29/08 National Institutes of Health, "Bioinformatics software for analyzing microbial genomes," R01 LM007938-01, \$779,500 (500,000 direct).

7/1/04 – 6/30/09 National Institutes of Health (NIAID), "Bioinformatics Resource Centers for Biodefense and Emerging/Re-emerging Infectious Diseases," Contract HHSN266200400038C, \$21,078,000. **PI:** Owen White.

7/1/04 – 9/30/06 National Science Foundation, "Sequencing of Xanthomonas pathovars enabling a broad comparative genomics approach to understanding vascular and non-vascular bacterial diseases of dicots and monocots, MCB-0412260, \$400,376 (320,301 direct). Subcontract to Iowa State (full award was \$736,336), **PI:** Adam Bogdanove.

09/30/03-09/29/08 National Institutes of Health, "Network for Large-scale Sequencing of Microbial Genomes," NIH-N01-AI-30, \$65,042,611. **PI:** C.M. Fraser.

11/10/03-10/31/06 National Institutes of Health, "TIGR/JTC Large Scale Sequencing Production Center," U54-HG003068-01, \$27,200,000. **P.I.:** J.C. Venter.

2/1/03-1/31/06 National Science Foundation, "Bioinformatics Analysis of Regulatory Sites in Genomic DNA Sequences," DBI-0234704, \$310,191.

9/1/03-8/31/05 National Science Foundation, "A high-speed network connection for genomics research," ANI-0333537, \$183,035.

1/1/02-12/31/05 National Institutes of Health (NIAID), "Comprehensive Sequencing of Bacillus anthracis," N01-AI-15447, \$5,557,421. **PI:** C.M. Fraser.

8/15/01-8/14/05 National Science Foundation, "Pre-mRNA splicing signals in Arabidopsis," MCB-0114792, \$371,903. Subcontract from Univ. of Maryland (full award was \$1,010,298). **PI:** Stephen Mount.

8/15/03-7/31/05 National Science Foundation, "2003 and 2004 Computational Genomics Conferences," DBI-0314206, \$29,500.

9/15/99-2/28/03 National Science Foundation, "KDI: Intelligent Computational Genomic Analysis," KDI-9980088, \$795,000. Subcontract from Boston Univ (full award was \$1,700,000). **PI:** Simon Kasif.

9/1/99–9/30/02 National Institutes of Health, "Computational Modeling of Genes and Gene Structure," R01 LM06845-01, \$820,323.

9/1/01-6/30/02 National Science Foundation, "The Fifth Annual Computational Genomics Conference at TIGR," DBI-0120900, \$24,803.

8/25/99–7/31/02 National Science Foundation, "Interpolated Markov Models for DNA Sequence Analysis," IIS-9902923, \$330,000. (Co-PI: Arthur Delcher.)

8/1/01-7/31/04 National Science Foundation, "The Development of a Prokaryotic Annotation Engine" (PI: Owen White), DBI-0110270, \$900,000.

7/1/01-6/30/04 Department of Energy, "Continuation of the Comprehensive Microbial Resource" (PI: Owen White), DE-FG02-01ER63203, \$2,047,000.

3/15/00-2/28/05 NIH/NIAID, "Trypanosoma cruzi genome project at TIGR," (PI: N. El-Sayed), U01-AI45038, \$3,820,617.

7/1/96-6/30/00 National Human Genome Research Institute, National Institutes of Health, "Computational Techniques for Genomic Analysis," K01-HG00022-01, \$335,190. (Career Award)

7/1/99-6/30/04 Burroughs Wellcome Fund, "The Hopkins Program in Computational Biology," (PIs: G. Rose and M. Paulaitis), \$2,567,050. (This training grant established a joint Hopkins/TIGR training program.)

11/1/98-10/31/02 National Science Foundation, "Generation of a Tomato EST Database," (PI: E. Kirkness), \$3,246,725.

3/1/99-2/28/01 Department of Energy, "Development of a Comprehensive Microbial Resource," (PI: Owen White), DE-FC02-95ER61962 (Supplement), \$411,184.

8/15/96-7/31/00 National Science Foundation, "Finding Genes in DNA using Machine Learning Algorithms," IRI-9530462. \$214,497.

3/1/98-8/31/99 Merck Genome Research Institute, "Computational Analysis of Intergenic Regions in Microbial Genomes," \$120,000. (PI: H.O. Smith; Co-PIs: S. Salzberg, O. White.)

12/1/96-11/30/97 Pfizer Central Research (Department of Molecular Sciences), "Probabilistic Methods for Pattern Discovery in Biological Sequence Data." \$52,606.

3/1/93-2/28/97 National Science Foundation, "Knowledge Discovery for Molecular Biology: Algorithms and Experiments," IRI-9223591. \$197,698.

4/1/95-3/31/96 National Science Foundation SGER, "A Proposal to Study the Applicability of Astronomical Image Processing Methods to Low Contrast Digital Images," AST-9520463, \$49,986.

5/1/95-4/30/96 National Science Foundation, "Distributed Data Mining in Large Databases," CDA-9421531, \$50,000. (Co-PIs: S. Kasif, E. Brill, and K. Fasman.)

10/1/91-3/31/96 National Science Foundation, "A Geometric Framework for the Exploration and Analysis of Astrophysical Data," IRI-9116843, \$539,553. (PI: M. Goodrich; Co-PIs: S. Salzberg and H. Ford (Dept. of Physics and Astronomy).)

6/1/93-12/1/93 National Science Foundation, Research Experience for Undergraduates award, supplement to Grant IRI-9116843, \$4,000.

3/1/91-2/28/96 National Science Foundation, CISE Institutional Infrastructure grant, "A Facility for Experimental Exploration and Validation," CDA-9015667, \$1,476,147. (PI: G. Masson.)

Software Systems (selected – all are open source)

1. Glimmer is a system for finding genes in microbial DNA. First released in 1997, Glimmer has become the world's leading bacterial gene finder, distributed to thousands of labs and institutions worldwide.

2. Bowtie is a system for very rapid alignment of short DNA sequence reads to a reference genome such as the human genome. Bowtie passed 200,000 downloads in late 2013.
3. Tophat and Cufflinks are systems for alignment of RNA-seq reads across splice junctions, and for assembly and quantitation of full-length transcripts from these alignments.
4. DIAMUND is a sensitive, efficient system for detecting mutations in exome data collected from families or from tumor samples. First released in 2013.
5. MUMmer, in its 3rd major release, is a system for alignment of the DNA sequences of entire genomes and chromosomes. Source code and executables have been freely available since 1999.
6. TopHat-Fusion is an enhanced version of TopHat with the ability to align reads across chromosomal fusion points, which results from the breakage and re-joining of different chromosomes, a common event in some tumors.
7. EDGE-pro aligns and quantitates transcript data from bacterial and archaeal RNA-seq experiments.
8. Hawkeye, a flexible graphical interface to genome assemblies from a variety of assemblers. Read the paper.
9. AMOScmp is a comparative genome assembler, which uses one genome as a reference on which to assemble another, closely related species.
10. Quake is a package to detect and correct substitution sequencing errors in whole-genome sequencing data sets with deep coverage, primarily for next-generation sequencing projects.
11. FLASH, Fast Length Adjustment of SHort reads, is a very fast program to merge paired-end reads that were sequenced from fragments that are shorter than twice the read length.
12. Minimus is a small, lightweight assembler for small jobs such as assembling a viral genome, assembling a set of reads from a single gene, or other tasks that don't require a large-genome assembler. Read the paper.
13. JIGSAW is a system that integrates the results of multiple gene finders and cDNA and protein sequence alignment programs to predict protein coding genes in eukaryotic genomes, including the human genome and the model organisms *C. elegans* and *D. melanogaster*.
14. GlimmerHMM is a system for finding genes in eukaryotic genomes, including the human genome, the malaria parasite *Plasmodium falciparum*, the model plant *Arabidopsis thaliana*, rice (*O. sativa*), *Theileria parva*, and others.
15. AMOS is an open-source genome assembly framework and software system comprising multiple genome assembly modules that are widely used for assembly of whole-genome shotgun data, including a scaffolder (Bambus), a comparative assembler (AMOS-Cmp), and an efficient small-scale assembler (Minimus).
16. AutoEditor is a system for automatically correcting the sequencing errors in large-scale genome sequencing projects. It is able to correct 85% of sequencing errors fully automatically.
17. VEIL and MORGAN were two early human gene finders, using Hidden Markov Models and decision trees respectively. These two systems were released in 1996 and 1997.
18. OC1 is a multivariate decision tree induction system. OC1 was applied to classification of data from molecular biology, astronomy, and numerous other application domains.

Professional Service

- 2014-present: Scientific Advisory Board, Lieber Institute for Brain Development
- 2014: Programme Committee, 1st Internatl. Conf. on Algorithms for Computational Biology
- 2011-2012: External Advisory Board, Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory
- 2010: Organizing committee, Beyond the Genome conference, October 2010.
- 2009: Co-chair, WABI-09, 8th Annual Workshop on Algorithms in Bioinformatics.
- 2007, 2009: Co-chair, 15th and 17th Annual International Conference on Microbial Genomics
- 2006-2011: Member, Maryland Stem Cell Research Commission
- 2008: Program Committee, Biology of Genomes Conference, Cold Spring Harbor Laboratory
- 2006-2011: Advisory council member, Computer Science Department, Princeton University.
- 1999-2005: Chair or Co-chair, 3rd through 8th Annual Conferences on Computational Genomics: 1999, 2000, 2001, 2003, 2004, 2005.

- 2003-2005 Program committee, IEEE Computer Society Computational Systems Bioinformatics Conference. 2003 and 2004 Steering Committee member.
- 2005: Area chair, Intelligent Systems for Molecular Biology (ISMB), the International Conference on Computational Biology.
- 2005: Scientific organizing committee, ESF-EMBO Symposium on Comparative Genomics.
- 1996–2002: Program committee member (every year), Intelligent Systems for Molecular Biology (ISMB) conference.
- 2002-2004: External Advisory Board member, Computer Science Department, Virginia Tech.
- 2000-2003: External Advisory Board member, Center for Bioinformatics and Computational Biology, University of Maryland, College Park.

National Institutes of Health, National Science Foundation, and other scientific review panels:

- Jan. 2014. Review panel, NIH Genetics and Genomics special study section.
- Nov. 2013. Chair, R13/U13 Review panel, NIH/NIGMS.
- August 2013. Special Emphasis Review panel, NIH/NICHHD.
- April 2013. Review panel, NIH Early Independence Awards.
- June 2012. Reviewer, NIH/NIAID Vaccines against Microbial Diseases Study Section
- March 2012. Review panel, NIH/NHGRI Encode program.
- March 2011. Reviewer, NIH/NIGMS EUREKA program.
- Nov. 2011. Chair of review panel, NIH Genomic Resources Grants for Community Resource Projects.
- July 2007-2011. Member, BDMA (Biodata Management and Analysis) study section, NIH
- June 2010. Review panel, NIH National Centers for Biomedical Computing program.
- May 2007. Review panel, NIH GCAT (Genomics, Computational Biology, and Technology) study section.
- July 2006. Review panel, NIH/NHGRI Large-scale sequencing centers program.
- May 2006. Site visit panel, Center for Cancer Research Nanobiology program, National Cancer Institute
- April 2006. Review Panel, NIH/NHGRI Genomic Database Resource program.
- May 2005. Review Panel, DOE Genomes to Life program.
- March 2005. Review Panel, NIH/NIDDK Clinical Nutrition Research Units, RFA DK-04-016.
- May 2004. Review panel, NIH National Centers for Biomedical Computing initiative.
- July 2003. Review panel, NIH/NHGRI Encyclopedia of DNA Elements (ENCODE) program.
- Mar. 2003. Review panel, NIH BISTI initiative, pre-NPEBC centers.
- Oct. 2002. NIAID Blue Ribbon Panel on Bioterrorism and Its Implications for Biomedical Research.
- Aug. 2002. NHGRI review panel, RFA HG-02-002, Model Organism Databases.
- Apr. 2002. Site visit panel, Laboratory of Experimental and Computational Biology, National Cancer Institute.
- Oct. 2001-Sept. 2002. Advisory Panel member, NSF Biological Databases and Informatics Program.
- Nov. 2001. Review Panel, NIH Biomedical Information Science and Technology Initiative (BISTI).
- April 2001. Review Panel, NSF, Information Technology Research Program.
- April 2001. Review panel NIH Genetic Sciences study section.
- Feb. 2001. Review Panel, NSF, CISE Research Infrastructure Program.
- June 2000. Genome Study Section member, NHGRI, NIH.
- March 2000. Special Study Section member, Genetic Sciences, NIH.
- July 1999. Site visit committee, National Center for Research Resources, NIH.
- Mar. 1999. SBIR/STTR review panel, Genetic Sciences Initial Review Group, NIH.
- Dec. 1998. Review Panel, NSF Digital Libraries Initiative part II.
- Aug. 1998. Review Panel, National Library of Medicine, NIH.
- Sep. 1998. Review Panel, Natl. Heart, Lung, and Blood Institute, NIH.

- October 1997. Review Panel, NSF, CISE/IRIS Research Instrumentation Program.
- November 1996. Review Panel, NSF, Knowledge Models and Cognitive Systems Program.
- June 1994. Review Panel, NSF, Research Initiation Awards, Database and Expert Systems Program.
- Apr. 1994. Review Panel, NSF, Research on Digital Libraries (a joint initiative with ARPA and NASA).
- Oct. 1993. Review Panel, NSF, CISE/IRIS Research Instrumentation Program.
- Mar. 1993. Review Panel, NSF, Research on Scientific Databases Initiative.

Other program committees and external review committees

Oct. 1998. External review committee, Functional Genomics and Bioinformatics Conf., Virginia Tech Institute of Genomics.

Program committee member, The 14th International Conference on Machine Learning (ICML-97).

Program committee member, Thirteenth National Conference on Artificial Intelligence (AAAI-96).

Program committee member, 8th IEEE International Conference on Tools with Artificial Intelligence, 1996.

Tutorial presentation, "Machine Learning Applications in Molecular Biology," ISMB-96 Conference.

Aug. 1991. Organizing committee, IJCAI Workshop on Representing Knowledge in Medical Decision Support Systems.

Referee for more than 25 journals including *Science* and *Nature*.

Supervision of Graduate Students and Postdoctoral Fellows

Ph.D. students at Johns Hopkins University (2011-present)

Samarjeet Prasad. Thesis advisor, 2013-. BCMB Ph.D. program.

Foram Ashar. Thesis committee member, 2012-. Human genetics Ph.D. program.

Cory Smith. Thesis committee member, 2012-. Human genetics Ph.D. program.

Taeyoung Hwang. Thesis committee member, 2013-. Biomedical Engineering Ph.D. program.

Ph.D. students at University of Maryland (with their current positions)

Derrick Wood, Ph.D. expected June 2014.

Daehwan Kim, Ph.D. 2013. Thesis advisor. Postdoctoral fellow, Salzberg lab.

Ben Langmead, Ph.D. 2012. Thesis advisor. Assistant Professor, Johns Hopkins University.

Samuel Angiuoli, Ph.D. 2011. Thesis advisor. Director, Bioinformatics Software Engineering, Institute for Genome Sciences, University of Maryland.

David Kelley, Ph.D. 2011. Thesis advisor. Postdoctoral fellow, the Broad Institute and Harvard University.

Michael Schatz, Ph.D. 2010. Thesis advisor. Assistant Professor, Cold Spring Harbor Laboratory.

Adam Phillippy, Ph.D. 2010. Thesis advisor. Principal Investigator, National Biodefense Analysis and Countermeasures Center (NBACC).

Bruce (Cole) Trapnell, Ph.D. 2010. Thesis advisor. Postdoctoral fellow, the Broad Institute and Harvard University.

James White, Ph.D. 2010, Applied Mathematics. Dissertation committee member.

Jessica Miller, Ph.D. 2009, Molec. and Cell Biol. Dissertation committee member.

Xue Wu, Ph.D. 2008. Dissertation committee member.

Eugene Melamud, Ph.D. 2006, Molec. and Cell Biol. Dissertation committee member.

Ph.D. students at Johns Hopkins University (pre-2006)

Jonathan Allen, Ph.D. September 2006. Thesis advisor. Research scientist, Lawrence Livermore National Laboratory.

Mihaela Pertea, Ph.D. August 2001. Thesis advisor. Assistant Professor, Johns Hopkins University.

John Sheppard, Ph.D. January 1997. Thesis advisor. Chaired Professor, Montana State University.

Sreerama K. Murthy, Ph.D. October 1995. Thesis advisor. Chairman, Teqnum Consultancy Services.

Lewis Stiller, Ph.D. June 1995. Dissertation committee member.

Kai Ming Ting, Ph.D. December 1995, University of Sydney, Australia. External examiner.

David Heath, Ph.D. October 1992. Dissertation committee member.

Kathleen Romanik, Ph.D. June 1992, University of Maryland. Dissertation committee member.

Walter Harley, Ph.D. in Psychology, June 1990. Dissertation committee member.

Postdoctoral Fellows and trainees

- Mihai Pop, Ph.D. Johns Hopkins University. Research Scientist in my group, 2000-2005. Currently an Associate Professor (tenured) at the University of Maryland, College Park
- Natalia Volfovsky, Ph.D. Hebrew University. Postdoc 1999-2002. Currently at the National Cancer Institute, Frederick, MD.
- Maria Ermolaeva, Ph.D. Moscow State University. Research Scientist in my group, 1999-2004.
- Pawel Gajer, Ph.D. Research scientist in my group, 2000-2004. Currently at Institute for Genome Sciences, University of Maryland.
- Jonathan Allen, Ph.D., Johns Hopkins University. Currently a research scientist at Lawrence Livermore National Laboratory..
- Carl Kingsford, Ph.D. Princeton University. Postdoc 2006-2007. Currently an Associate Professor at Carnegie Mellon University.
- Arthur Brady, Ph.D. Tufts University. Postdoc 2008-2011. Currently a Research Scientist at the Institute for Genome Sciences, University of Maryland School of Medicine.
- Todd Treangen, Ph.D. Polytechnic Univ. of Catalonia, Barcelona. Postdoc 2010-2012. Currently a Research Scientist at the National Biodefense Analysis and Countermeasures Center, Battelle National Biodefense Institute.
- Henry Lin, Ph.D. University of California, Berkeley. Postdoc 2010-2011, currently an associate research scientist at Columbia University.
- Tanja Magoc, Ph.D. University of Texas-El Paso. Postdoc 2010-2013.
- Stefan Canzar, Ph.D. Saarland University (Saarbrücken, Germany). Postdoc 2012-present.
- Stephan Pabinger, Ph.D. University of Technology, Graz (Austria). Postdoc 2012-2013.
- Daehwan Kim, Ph.D. 2013. Thesis advisor. Postdoc 2013-present.

Master's students, University of Maryland:

Derrick Wood, M.S. 2011.

David Kelley, M.S. 2010.

Benjamin Langmead, M.S. 2009 (co-advisor).

Megan Smedinghoff, M.S. 2009.

Michael Schatz, M.S. 2008.

Adam Phillippy, M.S. 2008.

Master's students, Johns Hopkins University:

Tugba Suzek, M.S.E. 2001.

Baris Suzek, M.S.E. 2001.

Mihaela Pertea, M.S.E. 1998.

Radu State, M.S.E. 1998.

Truxton Fulton, M.S.E. 1996.

Xin Chen, M.S.E. 1996.

John Rachlin, M.S.E. 1994.

Suraj Surendrakumar, M.S.E. 1992.

Scott Cost, M.S.E. 1991.

University Service and Teaching

Courses developed and taught:

- CMSC 828G Bioinformatics for Metagenomics (Univ. of Maryland)
- CMSC 828N/H Computational Gene Finding and Genome Assembly (Univ. of Maryland)
- CS 600.439 Computational Biology (Johns Hopkins University)
- CS 600.435 Artificial Intelligence (Johns Hopkins University)
- CS 600.350 Expert Systems (Johns Hopkins University)
- CS 600.661 Machine Learning (Johns Hopkins University)
- CS 600.665 Natural Language Processing (Johns Hopkins University)

John Hopkins University, 2011-present:

Founding Director, Center for Computational Biology

Associate Director, Institute for Data Intensive Engineering and Science (IDIES)

Leadership committee, High Performance Research Computing Facility (HPRCF)

Chair, Scientific Management Committee, HPRCF

Search committee member, Director of HPRCF

Served on multiple ad hoc appointment and tenure committees

Faculty search committee, Bloomberg Chair in Statistical Genomics, 2013-2014

University of Maryland, 2005-2011:

Scientific Advisory Committee, Maryland Pathogen Research Institute (MPRI)

Information Technology Council member, 2007-2010.

College Appointments, Promotion, and Tenure (APT) committee member, 2007-2008.

John Hopkins University, 1989-97:

Chair, Computer Science Dept. Graduate Admissions Committee, 1993–1995. Committee member, 1992-1996.

Member of the Steering Committee of the Engineering Faculty Assembly, 1993–1996.

Member of Computer Science Graduate Student-Faculty Forum, 1995–1997.

Member of Dean's Review Committee on Academic Computing, 1993–95.

Member of Computer Science Dept. Equipment Committee, 1990–1995.

Member of Whiting School Dean's Committee on International Education, 1991–92.

Manager of Computer Science Undergraduate Computing Lab, 1989–91.

Member of Faculty Invention Committee 1989–90.