OXFORD

2020 ISCB accomplishments by a Senior Scientist Award: Steven Salzberg

Christiana N. Fogg[1], Diane E. Kovats[2,*] and Ron Shamir[3]

[1]Kennsington, MD 20895, USA, [2]International Society for Computational Biology, VA, 20176, USA and [3]Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel

*To whom correspondence should be addressed.

Contact: dkovats@iscb.org

The Outstanding Contributions to IISCB honors a leader in the fields of computational biology and bioinformatics each year with the Accomplishments by a Senior Scientist Award. This award is the highest honor conferred by ISCB to a scientist who is recognized for significant research, education, and service contributions. Steven L. Salzberg, Bloomberg Distinguished Professor of Biomedical Engineering, Computer Science and Biostatistics at Johns Hopkins University and Director of the Center for Computational Biology, is being honored as the 2020 winner of the ISCB Accomplishments by a Senior Scientist Award. He will be recognized and give a keynote address at ISMB 2020 virtual conference being held on July 13–16, 2020.

Steven Salzberg: a journey between industry and academia

Steven Salzberg grew up in Columbia, SC. Throughout his childhood and young adulthood, he was always interested in science and deeply enjoyed reading science fiction. Salzberg was also fascinated by astronomy and considered studying physics. As an undergraduate at Yale University in the 1970's, he explored several majors and thought he had settled on English Literature but added Computer Science as a second major upon taking an introductory computer programming class. He recalled, 'This is the kind of math I thought I would really like to study', and he was soon captivated by artificial intelligence (AI) and natural language processing. At the advice of his undergraduate advisor, Salzberg spent a year after graduation gaining more programming experience by working at a local power company in South Carolina, where he worked on an IBM mainframe and used self-training courses to learn COBOL and IBM assembler. Salzberg said, 'It was a very boring sort of application, but I was still interested in programming. I liked the idea I could work on something technical and within a short period of time, I would have results that would do what I intended'.

Salzberg returned to Yale and completed his M.S. in computer science. He then joined a startup in Boston during the first blush of AI, although this and many other AI startups failed in the late 1980's due to lack of computing power and other technical limitations. One of Salzberg's advisors at the startup was AI pioneer Bill Woods, who held an adjunct appointment at Harvard University and later became Salzberg's graduate advisor in the Department of Computer Science. Salzberg had managed to avoid taking any biology classes as an undergraduate, but he heard about the Human Genome Project (HGP) while he was in graduate school in the late 1980's. He said, 'The Human Genome Project sounded like the most exciting thing in all of science at the time, and I wanted to be a part of that'. While completing his PhD project in machine learning, he started sitting in on biology classes, including a course by the late Stephen Jay Gould and reading on his own to learn about genomics and genetics. He was determined to figure out a way to using his computing knowledge to get involved in the HGP.

Salzburg continued doing research in machine learning as he started in his first academic position at Johns Hopkins University. He was still curious about genomics and recalled going to a talk in the early 1990's by Temple Smith about sequence differences between exons and introns. It dawned on Salzberg that he could use machine learning to distinguish exons from introns, which could be used as a strategy for gene finding. This became Salzberg's entrance into genomics. During this time, Salzberg was also introduced to Nobel Laureate Hamilton Smith, a notable microbiologist who discovered type II restriction enzymes. Salzberg recalled, '[Smith] had a secret passion for computer programming. He wanted to talk to computer scientists who were interested in genomics—that was me. And I was interested in learning more about genomics'. Salzberg and Smith began working together to understand how computer programs could be made for tasks like gene finding. Smith had also started collaborating with J. Craig Venter, and in 1997, both Smith and Salzberg began working at Venter's non-profit research institute, The Institute for Genomic Research (TIGR). Salzberg became the Director of Bioinformatics at TIGR and developed with his colleague Art Delcher the GLIMMER gene finder, a software system still used today to identify coding regions in bacteria, archea and viruses. In the early 2000's, the first *Mycobacterium tuberculosis* genomes were being sequenced by both TIGR and The Sanger Center. This led Salzberg and his colleagues to develop MUMmer, a system that could be used to compare large genomes. He also got involved in the HGP through the development of a gene finder that could analyze the human genome and, with his colleague Mihaela Pertea, also built other eukaryotic gene finders for plant, fungus and parasite genomes. Salzberg and his colleagues were called upon by the FBI after the 2001 anthrax attacks to analyze the genome of the

anthrax bacteria, and that work identified genetic mutations that eventually pinpointed the source of the bacteria to a biodefense lab in Fort Detrick, Maryland. In 2003, Salzberg co-founded the Influenza Genome Sequencing project with David Lipman, which involved the sequencing and analysis of thousands of influenza isolates.

Salzberg then moved to the University of Maryland, College Park in 2005, where he was the Horvitz Professor of Computer Science. He returned to JHU in 2011, where he is currently the Bloomberg Distinguished Professor of Biomedical Engineering, Computer Science and Biostatistics and the Director of the Center for Computational Biology in the Whiting School of Engineering. As next-generation sequencing technology developed, Salzberg's research interests shifted toward developing algorithms for large-scale genome assembly and sequence alignment, including the development of the open-source Tuxedo suite of programs (Bowtie, Tophat and Cufflinks).

Salzberg's current interests include the development of an improved human gene catalog and assembly and annotation of an Ashkenazi human reference genome. Recent technical advances have made this undertaking feasible, and the research community has desperately needed other reference genomes beyond the only publicly available genome, GRCh38. Salzberg is also working with colleagues on developing methods for using shotgun sequencing as a diagnostic tool for infectious diseases. They have tested their techniques on biopsy materials from patients with difficult-to-diagnose brain infections and on samples collected from eye infections, and the technology has the potential to work on a much broad range of infections.

Salzberg has trained numerous students and post-doctoral fellows throughout his time in academia and at TIGR, and he has focused on matching highly motivated individuals with projects that get them excited. Like many computational biologists, Salzberg is continually in search of interesting data associated with problems that matter, whether they involve the nature of the human genome, human health and disease, or any of a much broader range of microbial, plant and animal genomes.

Salzberg's body of work includes more than 300 publications, including many highly cited manuscripts. His contributions have been recognized through his election as a member of the American Academy of Arts and Sciences, a Fellow of the American Association for the Advancement of Science, a Fellow of the International Society for Computational Biology (ISCB) and a member of the Board of Scientific Counselors of the National Library of Medicine at NIH. All of Salzberg's bioinformatics systems have been released as free, open-source software and he won the 2013 Benjamin Franklin Award for Open Science for his advocacy of open-source software and of open sharing of genome sequence data. Salzberg is also a contributor to *Forbes* magazine and writes a widely read column that debunks pseudoscience and explains scientific and medical findings with honesty and clarity.

Salzberg is greatly honored to be the 2020 recipient of ISCB's Accomplishments by a Senior Scientist award. He has always felt at home at ISMB meetings since their inception and is touched by this award since it is bestowed upon him by his computational biology colleagues.