

# Jim Shaw

---

jshaw@ds.dfc.harvard.edu — github — google scholar — <https://jim-shaw-blunote.github.io>

**ACADEMIC POSITIONS**      **Postdoctoral researcher**      2024 - Now  
Dana-Farber Cancer Institute (Department of Data Science)  
and Harvard Medical School  
Supervisor: Heng Li

**EDUCATION**      **PhD in Mathematics**      Received 2024  
University of Toronto  
Supervisor: Yun William Yu

**Master's in Mathematics**      Received 2020  
University of Toronto  
Supervisor: Yun William Yu

**BASc in Engineering Physics**      Received 2019  
*Minor in Honours Mathematics*  
University of British Columbia

**RESEARCH SUMMARY**      My research is on developing algorithms and tools for large-scale analysis of biological sequences. My current biological interest lies in microbiome informatics (especially metagenomics) and computational microbial genomics.

**SELECTED PUBLICATIONS**      **J. Shaw, Y.W. Yu.** Rapid species-level metagenome profiling and containment estimation with sylph (2024). *Nature Biotechnology*.

I developed a computational method called *sylph* for detecting microorganisms from metagenomics sequencing data that is more precise and 50 times faster than the previous state-of-the-art. An independent consortium called AllTheBacteria used sylph on *all publicly available* (1.9 million) bacterial sequencing samples.

**J. Shaw, Y.W. Yu.** Fast and robust metagenomic sequence comparison through sparse chaining with skani (2023). *Nature Methods*.

My new computational method, *skani*, can compare realistic microbial genomes more accurately and 20 times faster than the state-of-the-art. Skani is now used to define prokaryotic species by the Genome Taxonomy Database, a leading resource that sets the global standard for prokaryote classification.

**J. Shaw, Y.W. Yu.** Proving sequence aligners can guarantee accuracy in almost  $O(m \log n)$  time through an average-case analysis of the seed-chain-extend heuristic (2023). *Genome Research + RECOMB conference proceedings*

I mathematically proved that certain DNA search algorithms—used by tens of thousands of scientists—are fast and accurate. I showed that an asymptotically accurate algorithm can bypass the  $O(n^2)$  worst-case runtime to achieve  $O(n^{1+\epsilon} \log n)$  time, breaking a theoretical barrier in sequence alignment that stood for 40 years.

**PUBLICATIONS**, Proceedings in computational biology conferences can have associated journals—in **PROCEEDINGS**, this case, only the journal version is listed.

**AND**

**PREPRINTS**

\* indicates co-first authors.

1. X.B. Zhang, G. Oualline, **J. Shaw**, Y.W. Yu. skandiver: a divergence-based analysis tool for identifying intercellular mobile genetic elements (2024). Bioinformatics (ECCB conference associated).
2. **J. Shaw**, Y.W. Yu. Fairy: fast approximate coverage for multi-sample metagenomic binning (2024). Microbiome.
3. **J. Shaw\***, J.S. Gounot\*, H. Chen, N. Nagarajan, Y.W. Yu. Floria: Fast and accurate strain haplotyping in metagenomes (2024). Bioinformatics (ISMB conference associated).
4. **J. Shaw**, Y.W. Yu. Rapid species-level metagenome profiling and containment estimation with sylph (2024). Nature Biotechnology.
5. **J. Shaw**, Y.W. Yu. Fast and robust metagenomic sequence comparison through sparse chaining with skani (2023). Nature Methods.
6. A. Zheng, **J. Shaw**, Y.W. Yu. Mora: abundance aware metagenomic read re-assignment for disentangling similar strains (2024). BMC Bioinformatics.
7. **J. Shaw**, Y.W. Yu. Proving sequence aligners can guarantee accuracy in almost  $O(m \log n)$  time through an average-case analysis of the seed-chain-extend heuristic (2023). Genome Research. (RECOMB conference associated)
8. M.C. Frith, **J. Shaw**, J. Spouge. How to optimally sample a sequence for rapid analysis (2023). Bioinformatics.
9. **J. Shaw**, Y.W. Yu. Theory of local k-mer selection with applications to long-read alignment (2021). Bioinformatics.
10. **J. Shaw**, Y.W. Yu. flopp: Extremely Fast Long-Read Polyploid Haplotype Phasing by Uniform Tree Partitioning (2022). Journal of Computational Biology. (RECOMB conference associated)
11. R. Cotsakis\*, **J. Shaw\***, J. Tierny, J. Levine. Implementing Persistence-Based Clustering of Point Clouds in the Topology ToolKit (2020). Topological Methods in Visualization: Theory, Software and Applications.
12. S. Hu\*, O. Schnetz\*, **J. Shaw\***, K. Yeats\*. Further investigations into the graph theory of  $\phi^4$ -periods and the  $c_2$  invariant. (2020). Annales de l'Institut Henri Poincaré D.
13. D. Bertrand, **J. Shaw**, M. Narayan, H.Q.A. Ng, S. Kumar, C. Li, M. Dvornicic, J.P. Soldo, J.Y. Kho, O.T. Ng, T. Barkham, B. Young, K. Marimuthu, K.R. Chng, M. Sikic, N. Nagarajan. Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes (2019). Nature Biotechnology.

**ACCEPTED  
CONFERENCE  
TALKS**

These talks were reviewed by a conference committee and accepted for oral presentation. Some are proceedings associated while others are not.

1. ISMB, Montreal, Canada. 2024. (Proceedings)
2. Great Lakes Bioinformatics Conference, Pittsburgh, USA. 2024. (Talk only)
3. Genome Informatics, Cold Spring Harbor, USA. 2023. (Talk only)
4. ISMB, Lyon, France. 2023. (Talk only)

5. RECOMB, Istanbul, Turkey. 2023. (Proceedings)
6. RECOMB, Padova, Italy (virtual). 2021. (Proceedings)

**INVITED SEMINARS AND TALKS**

1. Indian Institute of Science. Department of Computational and Data Sciences Seminar. Virtual, 2024.
2. University of Waterloo. Computer Science Seminar. Waterloo, Canada, 2024.
3. University of Tokyo. Computational Biology Seminar. Tokyo, Japan. 2023.
4. University of Toronto. Combinatorics student seminar. Toronto, Canada. 2022.

**PROFESSIONAL EXPERIENCE AND INTERNSHIPS**

**Visiting PhD student** Jan 2024 - June 2024  
*Carnegie Mellon University*, Pittsburgh, USA  
 Advised by Yun William Yu.

**Visiting Computational Biology Researcher** Sept 2023 - Dec 2023  
*University of Tokyo*, Tokyo, Japan  
 Advised by Martin C. Frith.

**Data Science Internship** May 2019 - August 2019  
*DeepND*, Vancouver, BC, Canada

**Mathematics Research Intern** May 2018 - September 2018  
*University of Waterloo*, Waterloo, ON, Canada  
 Advised by Karen Yeats.

**Computational Genomics Intern** May 2017 - September 2017  
*Genome Institute of Singapore*, Singapore  
 Advised by Niranjana Nagarajan.

**Software Engineer Intern** January 2016 - May 2016  
*Zaber Technologies*, Vancouver, BC, Canada

**POSTERS**

1. RECOMB. Boston, USA. 2024. (**Best poster award: 2/284 selected**).
2. ISMB. Lyon, France. 2023.
3. RECOMB. San Diego, USA. 2022.

**PROFESSIONAL SERVICE**

1. Reviewer for Genome Biology, Bioinformatics Advances, BMC Bioinformatics, Briefings in Bioinformatics.
2. Secondary reviewer for ISMB 2023 and RECOMB 2022, 2023, and 2024 conferences.

**HONOURS AND AWARDS**

This section is roughly sorted by competitiveness/prestige.

1. NSERC-CGS D Scholarship (2022) - Awarded \$105,000 over 3 years through a national PhD scholarship competition.
2. RECOMB 2024 Best Poster Award (**2/284 selected**)
3. NSERC Michael Smith Foreign Supplement (2023) - Awarded \$6000 to travel to Japan for research.
4. PhD Entrance Scholarship (2020) - Awarded \$5000 for performance in the MSc of Mathematics program.

5. Trek Excellence Award (2017) - Awarded \$1500 for being ranked in the top 5% of students in the Faculty of Applied Science at the University Of British Columbia for the 2016-2017 year.
6. Singapore International Pre-Graduate Award (2017) - \$6000 awarded to pursue a research internship in Singapore for four months.
7. NSERC Undergraduate Summer Research Award (2018) - \$4500 awarded to pursue summer research at the University of Waterloo.
8. Donald J. Evans Scholarship in Engineering (2017) - \$500 awarded on the recommendation of the Faculty of Applied Science at the University of British Columbia.
9. NSERC Industrial Undergraduate Student Research Award (2016) - \$4500 awarded to pursue an industrial internship.

## **TEACHING**

All teaching assistant positions involved a combination of leading tutorials and recitations, marking, and office hours.

### **University of Toronto**

1. MAT 135 Calculus 1 TA: Fall 2019, Winter 2022.
2. MAT 136 Calculus 2 TA: Winter 2020
3. MAT 223 Linear Algebra 1 TA: Fall 2020, Fall 2021, Winter 2022
4. MAT 224 Linear Algebra 2 TA: Winter 2021
5. MAT learning centre TA: Fall 2019, Fall 2022, Winter 2023.