

Personal Information

Name: Mingfu Shao **Phone:** +1 412 608 9150
Citizenship: China **Email:** mingfu.shao@cs.cmu.edu
Date of Birth: Jan. 1986 **Website:** www.cs.cmu.edu/~mingfus
Address: GHC 7401, 5000 Forbes Ave, Pittsburgh, PA 15213, USA

Professional Appointment

Nov. 2015—present **Lane Fellow**
Computational Biology Department, School of Computer Science,
Carnegie Mellon University, Pittsburgh, PA, USA
Supervisor: Prof. Carl Kingsford

Education

Sep. 2011—Jul. 2015 **Ph.D. in Computer Science**
School of Computer and Communication Sciences (IC),
École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
Advisor: Prof. Bernard M.E. Moret
Thesis: Models and Algorithms for Comparative Genomics

Sep. 2008—Jul. 2011 **M.S. in Computer Science**
Institute of Computing Technology (ICT),
Chinese Academy of Sciences (CAS), Beijing, China
Advisor: Prof. Dongbo Bu
Thesis: On the Protein Threading

Sep. 2004—Jul. 2008 **B.S. in Computer Science**
Department of Computer Science,
Beijing Institute of Technology, Beijing, China

Industrial Experience

Oct. 2007—Feb. 2008 **Internship**
Microsoft Research Asia, Beijing, China
Mentor: Yingnong Dang

Academic Services

Program Committee *WABI'17, ISMB'18, RECOMB-CCB'18*

Conference Reviewer *ISAAC'14, RECOMB'15, ISMB'15, ISBRA'15, RECOMB'16, ISMB'16, RECOMB-CG'16, WABI'16, RECOMB'17, RECOMB'18*

Journal Reviewer *Genome Biology and Evolution, PLOS ONE, Algorithms for Molecular Biology, Bioinformatics, IEEE/ACM Transactions on Computational Biology and Bioinformatics*

Teaching Activities

- Fall 2014 **Teaching Assistant:** *Advanced Algorithms*, IC, EPFL
Guest Lecture: Hungarian algorithm and Hopcroft-Karp algorithm
- Fall 2013 **Teaching Assistant:** *Advanced Algorithms*, IC, EPFL
Guest Lecture: Competitive analysis, Paging algorithms
- Fall 2012 **Teaching Assistant:** *Advanced Algorithms*, IC, EPFL
- Spring 2012 **Teaching Assistant:** *Computational Molecular Biology*, IC, EPFL
- Fall 2010 **Teaching Assistant:** *Algorithm Design and Analysis*, ICT, CAS
Guest Lecture: PTAS for closest string and closest substring problems
- Fall 2009 **Teaching Assistant:** *Algorithm Design and Analysis*, ICT, CAS

Awards

- 2015 Dimitris N. Chorafas Foundation Award (27 awardees worldwide and 2 at EPFL).
- 2015 Lane Fellowship, Carnegie Mellon University.
- 2015 Swiss NSF Early Postdoc.Mobility Fellowship (host by MIT, declined).
- 2014 Chinese Government Award for Outstanding Self-Financed Students Abroad
(500 awardees worldwide and 8 awardees in Switzerland in 2014).
- 2012 Outstanding Teaching Assistant Award, IC, EPFL.
- 2009 Merit Student of Advanced Computing Research Center, ICT, CAS.
- 2008 Beijing Outstanding Graduates.
- 2007 China Encouragement Scholarship.
- 2006 First Prize in China Collegiate Physics Contest.
- 2005 China Scholarship.

Publications

1. Ma, C., **Shao, M.**, and Kingsford, C. SQUID: transcriptomic structural variation detection from RNA-seq. *Genome Biology*, 2018 (accepted, to appear)
2. **Shao, M.** and Kingsford, C. Accurate assembly of transcripts through phase-preserving graph decomposition. *Nat. Biotechnol.*, 35(12):1167–1169, 2017
3. **Shao, M.** and Kingsford, C. Efficient heuristic for decomposing a flow with minimum number of paths. *IEEE-ACM T. Comput. Bi.*, 2017 (accepted, to appear). doi: 10.1109/TCBB.2017.2779509
4. **Shao, M.**, Ma, J., and Wang, S. DeepBound: Accurate identification of transcript boundaries via deep convolutional neural fields. *Bioinformatics*, 33(14):i267–i273, 2017 (*ISMB'17*)
5. **Shao, M.** and Moret, B.M.E. On computing breakpoint distances for genomes with duplicate genes. *J. Comput. Biol.*, 26(6):571–580, 2017 (*journal version of RECOMB'16 proceedings*)

6. **Shao, M.** and Moret, B.M.E. On computing breakpoint distances for genomes with duplicate genes. In *Proc. 20th Int'l Conf. Comput. Mol. Biol. (RECOMB'16)*, volume 9649 of *Lecture Notes in Comp. Sci.*, pages 189–203. 2016
7. Nair, N.U., Hunter, L., **Shao, M.**, Grnarova, P., Lin, Y., Bucher, P., and Moret, B.M.E. A maximum-likelihood approach for building cell-type trees by lifting. *BMC Genomics*, 17(Suppl 1):14, 2016 (*APBC'16*)
8. **Shao, M.** and Moret, B.M.E. A fast and exact algorithm for the exemplar breakpoint distance. *J. Comput. Biol.*, 23(5):337–346, 2016 (*journal version of RECOMB'15 proceedings*)
9. **Shao, M.** and Moret, B.M.E. A fast and exact algorithm for the exemplar breakpoint distance. In *Proc. 19th Int'l Conf. Comput. Mol. Biol. (RECOMB'15)*, volume 9029 of *Lecture Notes in Comp. Sci.*, pages 309–322. 2015
10. **Shao, M.** and Moret, B.M.E. Comparing genomes with rearrangements and segmental duplications. *Bioinformatics*, 31(12):i329–i338, 2015 (*ISMB'15*)
11. **Shao, M.**, Lin, Y., and Moret, B.M.E. An exact algorithm to compute the double-cut-and-join distance for genomes with duplicate genes. *J. Comput. Biol.*, 22(5):425–435, 2015 (*journal version of RECOMB'14 proceedings*)
12. **Shao, M.**, Lin, Y., and Moret, B.M.E. An exact algorithm to compute the DCJ distance for genomes with duplicate genes. In *Proc. 18th Int'l Conf. Comput. Mol. Biol. (RECOMB'14)*, volume 8394 of *Lecture Notes in Comp. Sci.*, pages 280–292. 2014
13. **Shao, M.** and Moret, B.M.E. On the DCJ median problem. In *Proc. 25th Ann. Symp. Combin. Pattern Matching (CPM'14)*, volume 8486 of *Lecture Notes in Comp. Sci.*, pages 273–282. 2014
14. **Shao, M.**, Lin, Y., and Moret, B.M.E. Sorting genomes with rearrangements and segmental duplications through trajectory graphs. *BMC Bioinformatics*, 14(Suppl 15):S9, 2013 (*RECOMB-CG'13*)
15. **Shao, M.** and Lin, Y. Approximating the edit distance for genomes with duplicate genes under DCJ, insertion and deletion. *BMC Bioinformatics*, 13(Suppl 19):S13, 2012 (*RECOMB-CG'12*)
16. **Shao, M.**, Wang, S., Wang, C., Yuan, X., Li, S.-C., Zheng, W.-M., and Bu, D. Incorporating ab initio energy into threading approaches for protein structure prediction. *BMC Bioinformatics*, 12(Suppl 1):S54, 2011 (*APBC'11*)
17. Wei, Y., **Shao, M.**, Yang, J., Wang, C., Li, S.-C., and Bu, D. Approximating conserved regions of protein structures. In *Proc. 9th Ann. Int'l Comput. Systems Bioinf. Conf. (CSB'10)*, volume 9, pages 204–212. 2010

Conference Oral Presentations

- Jul. 2017 “DeepBound: Accurate identification of transcript boundaries via deep convolutional neural fields”,
25th Int'l Conf. Intelligent Systems Mol. Biol. (ISMB'17), Prague, Czech Republic
- May 2017 “Theory and algorithm for the minimum path flow decomposition problem”,
7th RECOMB Massively Parallel Sequencing (RECOMB-seq'17), Hong Kong, China
- Apr. 2016 “On computing breakpoint distances for genomes with duplicate genes”,
20th Int'l Conf. Comput. Mol. Biol. (RECOMB'16), Santa Monica, CA, USA

- Jul. 2015 “Comparing genomes with rearrangements and segmental duplications”,
23th Int’l Conf. Intelligent Systems Mol. Biol. (ISMB’15), Dublin, Ireland
- Apr. 2015 “A fast and exact algorithm for the exemplar breakpoint distance”,
19th Int’l Conf. Comput. Mol. Biol. (RECOMB’15), Warsaw, Poland
- Jun. 2014 “On the DCJ median problem”,
25th Ann. Symp. Combin. Pattern Matching (CPM’14), Moscow, Russia
- Apr. 2014 “An exact algorithm to compute the DCJ distance for genomes with duplicate genes”,
18th Int’l Conf. Comput. Mol. Biol. (RECOMB’14), Pittsburgh, PA, USA
- Oct. 2013 “Sorting genomes with rearrangements and segmental duplications through trajectory graphs”,
11th RECOMB Comp. Genomics (RECOMB-CG’13), Lyon-Villeurbanne, France
- Oct. 2012 “Approximating the edit distance for genomes with duplicate genes under DCJ, insertion and deletion”,
10th RECOMB Comp. Genomics (RECOMB-CG’12), Niteroi, Brazil
- Jan. 2011 “Incorporating ab initio energy into threading for protein structure prediction”,
9th Asia Pacific Bioinf. Conf. (APBC’11), Incheon, Korea

Invited Seminar Talks

- Apr. 2018 “Efficient algorithms for large-scale transcriptomics and genomics”,
Computer Science Department and Center for Computational Molecular Biology,
Brown University
- Mar. 2018 “Efficient algorithms for large-scale transcriptomics and genomics”,
Department of Computer Science and Engineering, Penn State University
- Feb. 2018 “Efficient algorithms for large-scale transcriptomics and genomics”,
Department of Electrical Engineering and Computer Sciences and Center for
Computational Biology, University of California, Berkeley
- Nov. 2016 “New methods for transcript assembly using flow decomposition”,
École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
- Nov. 2016 “New methods for transcript assembly using flow decomposition”,
Eidgenössische Technische Hochschule Zürich (ETH), Switzerland
- May 2016 “Fast algorithms for problems in genome rearrangements”,
Shandong University, Jinan, China
- Apr. 2016 “On computing breakpoint distances for genomes with duplicate genes”,
University of Pittsburgh, Pittsburgh, PA, USA
- Aug. 2015 “Models and algorithms for comparative genomics”,
Institute of Computing Technology, CAS, Beijing, China
- Aug. 2015 “Models and algorithms for comparative genomics”,
CAS-MPG Partner Institute for Computational Biology,
Shanghai Institutes for Biological Sciences, CAS, Shanghai, China
- Aug. 2015 “Models and algorithms for comparative genomics”,
Beijing Institute of Genomics, CAS, Beijing, China

- Aug. 2015 “Models and algorithms for comparative genomics”,
Kunming Animal Institute, CAS, Kunming, China
- Apr. 2014 “On the rearrangement distances for genomes with duplicate genes”,
University of South Carolina, Columbia, SC, USA
- Aug. 2013 “Small phylogeny construction and genes correspondence assignment”,
University of Lausanne, Lausanne, Switzerland