CONTACT INFORMATION

Name:　Mingfu Shao　　Email:　mxs2589@psu.edu
Phone:　+1 412 608 9150　　Website:　https://sites.psu.edu/mxs2589
Address:　W310 Westgate Building, State College, PA 16802, USA

CURRENT APPOINTMENT

Assistant Professor, Jul. 2018—present
　Department of Computer Science and Engineering, College of Engineering,
　Huck Institutes of Life Sciences,
　The Pennsylvania State University, State College, PA, USA

EDUCATION & TRAINING

Lane Fellow, Nov. 2015—Jun. 2018
　Computational Biology Department, School of Computer Science,
　Carnegie Mellon University, Pittsburgh, PA, USA
　Supervisor: Prof. Carl Kingsford

Ph.D. in Computer Science, Sep. 2011—Jul. 2015
　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

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

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

HONORS & AWARDS

Charles K. Etner Career Development Professorship in Engineering,
　The Pennsylvania State University, 2018.

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), 2015.

Chinese Government Award for Outstanding Self-Financed Students Abroad
　(500 awardees worldwide and 8 awardees in Switzerland in 2014), 2014.

Outstanding Teaching Assistant Award, IC, EPFL, 2012.

Merit Student of Advanced Computing Research Center, ICT, CAS, 2009.

Beijing Outstanding Graduates, 2008.

China Encouragement Scholarship, 2007.

First Prize in China Collegiate Physics Contest, 2006.

China Scholarship, 2005.

PROFESSIONAL SERVICES

Grant Reviewer:
  NIH, Biodata Management and Analysis (BDMA) study section reviewer, 2020
  NSF, *ad hoc* reviewer, 2020

Program Committee:
  WABI'17, ISMB'18, RECOMB-CCB'18, APBC'19, HiCOMB'19,
  ISMB/ECCB'19, ACM/BCB'19, ISMB'20, RECOMB'21

Conference Reviewer:
  ISAAC'14, RECOMB'15, ISMB'15, ISBRA'15, RECOMB'16, ISMB'16,
  RECOMB-CG'16, WABI'16, RECOMB'17, RECOMB'18, WABI'19

Journal Reviewer:
  Genome Biology and Evolution, PLOS ONE, Algorithms for Molecular Biology,
  IEEE/ACM Transactions on Computational Biology and Bioinformatics,
  Bioinformatics, Nature Methods, Nature Communication

TEACHING

Fall 2020, Instructor: *Data Structures and Algorithms (CMPSC 465)*, Penn State

Fall 2019, Instructor: *Algorithm Design and Analysis (CSE 565)*, Penn State

Spring 2019, Instructor: *Data Structures and Algorithms (CMPSC 465)*, Penn State

Spring 2018, Instructor: *Advanced Topics in Computational Genomics*, CMU
  Topic: Single-cell RNA-seq Analysis

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

SUPERVISING

Graduate Students:
  Qian Shi, CSE PhD Student, Jan. 2019—current
  Qimin Zhang, CSE PhD Student Aug. 2019—current
  Xiaofei (Carl) Zang, BG PhD Student, Aug. 2020—current
  Xiang Li, CSE MS Student, Jan. 2020—current
  Raj Pandey, CSE MS Student, Sep. 2020—current

Undergraduate Students:
  Haoyuan Chen, May. 2020—present
  Amber Yuan, May. 2020—Sep. 2020
  Anbang Teng, May. 2019—Sep. 2019
  Thomas Isaac, Jan. 2019—Jul. 2019
  Harieasswar Lakshmidevi, Dec. 2018—Jul. 2019

RESEARCH GRANTS

NSF DBI-2019797, Jul. 2020—Jun. 2023, $400,000, PI (100%),
"BBSRC-NSF/BIO: IIBR Informatics: Collaborative Research: Inference of isoform-level regulatory infrastructures with studies in steroid-producing cells".

NIH R01HG011065, Feb. 2021—Jan. 2026, $1.85M, PI (100%),
"Computational Methods for Assembling Multiple RNA-seq Samples".

PRE-PRINTS

1. Shi, Q. and **Shao, M.** Coral accurately bridges paired-end RNA-seq reads alignment. *bioRxiv*, 2020. doi: 10.1101/2020.03.03.975821

PUBLICATIONS

1. Tung, L.H., **Shao, M.**, and Kingsford, C. Quantifying the benefit offered by transcript assembly with scallop-lr on single-molecule long reads. *Genome Biology*, 20(1):1–18, 2019

2. **Shao, M.** and Kingsford, C. Efficient heuristic for decomposing a flow with minimum number of paths. *IEEE-ACM Transactions on Comput. Biol. and Bioinf.*, 16:658–670, 2019. doi: 10.1109/TCBB.2017.2779509

3. Xin, H., **Shao, M.**, and Kingsford, C. Context-aware seeds for read mapping. In *Proc. 19th Int'l Workshop on Algorithms in Bioinf. (WABI'19)*, volume 143 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 15:1–15:13. 2019. doi: 10.1101/643072

4. Ma, C., **Shao, M.**, and Kingsford, C. SQUID: transcriptomic structural variation detection from RNA-seq. *Genome Biology*, 19(1):52, 2018

5. **Shao, M.** and Kingsford, C. Accurate assembly of transcripts through phase-preserving graph decomposition. *Nature Biotechnology*, 35(12):1167–1169, 2017

6. **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*)

7. **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*)

8. **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

9. 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*)

10. **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*)

11. **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

PUBLICATIONS (CONTINUED)

12. **Shao, M.** and Moret, B.M.E. Comparing genomes with rearrangements and segmental duplications. *Bioinformatics*, 31(12):i329–i338, 2015 (*ISMB'15*)

13. **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*)

14. **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

15. **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

16. **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*)

17. **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*)

18. **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*)

19. 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 PRESENTATIONS

"Accurate assembly of transcripts through phase-preserving graph decomposition", *26th Int'l Conf. Intelligent Systems Mol. Biol. (***ISMB'18***)*, Chicago, IL, USA, Jul. 2018.

"DeepBound: Accurate identification of transcript boundaries via deep convolutional neural fields", *25th Int'l Conf. Intelligent Systems Mol. Biol. (***ISMB'17***)*, Prague, Czech Republic, Jul. 2017.

"Theory and algorithm for the minimum path flow decomposition problem", *7th RECOMB Massively Parallel Sequencing (***RECOMB-seq'17***)*, Hong Kong, China, May 2017.

"On computing breakpoint distances for genomes with duplicate genes", *20th Int'l Conf. Comput. Mol. Biol. (***RECOMB'16***)*, Santa Monica, CA, USA, Apr. 2016.

"Comparing genomes with rearrangements and segmental duplications", *23th Int'l Conf. Intelligent Systems Mol. Biol. (***ISMB'15***)*, Dublin, Ireland, Jul. 2015.

"A fast and exact algorithm for the exemplar breakpoint distance", *19th Int'l Conf. Comput. Mol. Biol. (***RECOMB'15***)*, Warsaw, Poland, Apr. 2015.

"On the DCJ median problem", *25th Ann. Symp. Combin. Pattern Matching (***CPM'14***)*, Moscow, Russia, Jun. 2014.

CONFERENCE PRESENTATIONS (CONTINUED)	"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, Apr. 2014.

"Accurate assembly of transcripts through phase-preserving graph decomposition", *26th Int'l Conf. Intelligent Systems Mol. Biol. (***ISMB'18***)*, Chicago, IL, USA, Jul. 2018.

"Sorting genomes with rearrangements and segmental duplications through trajectory graphs", *11th RECOMB Comp. Genomics (***RECOMB-CG'13***)*, Lyon-Villeurbanne, France, Oct. 2013.

"Approximating the edit distance for genomes with duplicate genes under DCJ, insertion and deletion", *10th RECOMB Comp. Genomics (***RECOMB-CG'12***)*, Niteroi, Brazil, Oct. 2012.

"Incorporating ab initio energy into threading for protein structure prediction", *9th Asia Pacific Bioinf. Conf. (***APBC'11***)*, Inchon, Korea, Jan. 2011.

INVITED TALKS

"Efficient algorithms for large-scale transcriptomics and genomics", School of Medicine, The Pennsylvania State University, Feb. 2019.

"Efficient algorithms for large-scale transcriptomics and genomics", Computer Science Department and Center for Computational Molecular Biology, Brown University, Apr. 2018.

"Efficient algorithms for large-scale transcriptomics and genomics", Department of Computer Science and Engineering, Penn State University, Mar. 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, Feb. 2018.

"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, Nov. 2016.

"Fast algorithms for problems in genome rearrangements", Shandong University, Jinan, China, May 2016.

"On computing breakpoint distances for genomes with duplicate genes", University of Pittsburgh, Pittsburgh, PA, USA, Apr. 2016.

"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.

INVITED TALKS (CONTINUED)

"Models and algorithms for comparative genomics", Kunming Animal Institute, CAS, Kunming, China, Aug. 2015.

"On the rearrangement distances for genomes with duplicate genes", University of South Carolina, Columbia, SC, USA, Apr. 2014.

"Small phylogeny construction and genes correspondence assignment", University of Lausanne, Lausanne, Switzerland, Aug. 2013.