# Ming Li

**Address**

David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, Ont. N2L 3G1, Canada
WWW homepage: http://cs.uwaterloo.ca/~mli/

**Contact Information**

Phone: 519-888-4659 (Office)
Cell Phone: 519-500-3026
E-mail: mli@uwaterloo.ca
Fax: 519-885-1208

**Education**

Harvard University, Postdoctoral fellow, supervisor: Les Valiant, 9/1986 – 8/1987
Cornell University, M.S. 1983, Ph.D. 1985 Computer Science. Advisor: Juris Hartmanis. 1981 – 1985
Wayne State University, M.S., Computer Science. Advisor: W. Grosky, 4/1980 – 12/1980
Chinese Academy of Sciences, Institute of Computing Technology, 1978 – 1980.

**Primary Positions**

University Professor, School of Computer Science, University of Waterloo, 2009 – present
Canada Research Chair in Bioinformatics, Tier I, School of Computer Science, University of Waterloo, 2002 – present
Professor (tenured), School of Computer Science, University of Waterloo, 7/1994 – present
Professor (tenured), Computer Science Department, University of California at Santa Barbara, 2000-2002 (On leave from UW)
Associate Professor, Computer Science Department, University of Waterloo, 7/1989 – 7/1994
Assistant Professor, Computer Science Department, York University, 1/1988 – 7/1989
Assistant Professor (non-tenure track), Aiken Computation Lab, Harvard University, 9/1987 – 1/1988
Assistant Professor, Dept. Computer Science and Information, Ohio State University, 4/1985 – 9/1986

**Research Interests**

Kolmogorov complexity and its applications, average-case analysis of algorithms, information distance, proteomics, peptidomics and cancer immunotherapy, machine learning, natural language processing, computational complexity.

**Prizes & Awards**

Overseas outstanding contributions award. Chinese Computing Federation, 2021.

Lifetime Achievement Award: CS-Can/Info-Can, 2020.

McGuffey Longevity Award: for book *An introduction to Kolmogorov complexity and its applications.* 2020.

Killam Prize, 2010. ($100,000). One award each year in Engineering, Natural Science, Health Science, Humanity, and Literature, in Canada.

Outstanding Contribution Award, IEEE Granular Computing, San Jose, August, 2010

Premier's Discovery Award (Innovation Leadership), 2009

IEEE GrC Conference, Pioneer Award, May, 2006

Award of Merit, 1997. The Federation of Chinese Canadian Professionals (FCCP)

**Honours & Fellowships**

Fellow, ISCB (Int'l Society of Computational Biology), 2021

Canada Research Chair in Bioinformatics, Tier I, Nov. 2002 (renewed 2009, 2016)

University Professor, University of Waterloo, June 2009.

Fellow, Royal Society of Canada, 2006

Fellow, ACM, 2006

Fellow, IEEE, 2006

Killam Research Fellowship, Canada Council for the Arts, 2001

E.W.R. Steacie Memorial Fellowship, NSERC, 1996

**Best Paper Awards & Highlights**

*COLING'2010*: F. Bu, X. Zhu and M. Li, Measuring the Non-compositionality of Multiword Expressions. Best paper award selected from over 800 submissions, August 21-26, 2010. pp. 116-124

*Computational System Biology'2007*: Z. Zhang, H. Lin, M. Li, MANGO: a new approach to multiple sequence alignment. San Diego, CA, Aug. 13-15, 2007.

*1999 Genome Informatics Workshop*: X. Chen, S. Kwong, and M. Li, A compression algorithm for DNA sequences and its application in genome comparison. Tokyo, Japan, Dec. 1999, pp. 51-61, and in *RECOMB'00*. Tokyo, Japan, April 2000, p. 107. Tokyo, Japan, Dec. 1999

*2007 Genome Informatics Workshop*: X. Gao, D. Bu, S.C. Li, J. Xu, M. Li, FragQA: predicting local fragment quality of a sequence-structure alignment. Singapore, Dec. 3-5, 2007.

*SIGKDD*: (Finalist for the best paper award.) X. Zhang, Y. Hao, X. Zhu, M. Li, Information distance from a question to an answer. pp. 874-883, Aug. 12-15, 2007.

*FOCS'85* (Disqualified for best student paper, due to graduation date.) M. Li, Simulating $2$ pushdown stores by $1$ tape in $O(n^{1.5})$ time. Portland, pp. 56-64, 1985.

NH Tran, R. Qiao, L. Xin, X. Chen, C. Lui, X. Zhang, B. Shan, A. Ghodsi, M. Li. Deep learning enables de novo peptide sequencing from data-independent acquisition mass spectrometry. *RECOMB'19*, Highlight track, 2019.

N.H, Tran, X. Zhang, L. Xin, B. Shan, M. Li. De novo peptide sequencing by deep learning. *RECOMB'18*, Highlight track, 2018.

**Selected Keynote/Distinguished Lectures**

1. Keynote speech, Deep learning and personalized immunotherapy. Future Science Prize Science Forum, Nov. 19, 2021, Beijing China. Online audience, 4 million.

2. Keynote Speech, Digital transformation in biotechnology. CNCC-21. Shen Zhen, China, Oct. 29, 2021.

3. Keynote Speech, Neoantigen discovery for personalized cancer immunotherapy. *RECOMB'2021*, Aug. 29 – Sept. 1, 2021. Virtual meeting.

4. Public Speech, "Self-supervised learning, 0-shot learning, and information distance" Tsinghua University, Jan. 7, 2021. 303242 people online attending.

5. Keynote Speech, *AI enabling cancer immunotherapy,* 12th Big Data Technonologies Conf. (BDTC), Dec. 5, 2019, Beijing, Great Wall Hotel.

6. Keynote Speech, AI, NLP, and Bioinformatics. *Beijing Academy of Artificial Intelligence Conf.*, Oct. 31, 2019. Beijing, China.

7. Keynote Speech, *3rd generation chatbot.* China Conference on Knowledge and Semantic Computing, Hangzhou, August 24-27, 2019

8. Keynote Speech, *3rd generation chatbot.* Tsinghua University, July 1st, 2019, at the Ceremony of the establishment of Tsinghua AI NLP Center.

9. Keynote Speech, *AI empowering cancer immunotherapy*, and Panelist (on NLP empowering industry), 2019 Global AI technology Conference (CAITC 2019). Nanjing, China, May 25-26, 2019.

10. Panelist: *Will AI surpass human intelligence*, ACM China Turing Meeting, Chengdu, May 18, 2019.

11. Keynote Speech: *Discovering Neoantigens for Cancer Immunotherapy. UWORCS*, April 10, 2019. University of Western Ontario.

12. Keynote Speech: *Chatbot technologies, 7th Guangzhou Electric Grid Conference.* March 26, 2019.

13. Keynote Speech: *Discovering Neoantigens for Cancer Immunotherapy. International 17th Asia Pacific Bioinformatics Conference (APBC 2019)*, Wuhan, China, Jan. 14-16, 2019.

14. Public Lecture: *Neoantigen, New Medicine and New Hope.* Sept 30, 2018, Beijing University of Technologies. (Also at Beijing 301 Hospital, Changsha Xiangya Hospital, Chengdu Huaxi Hospital, Peking Union Medical College Hospital.)

15. Keynote Speech: *SIGBIO China Symposium: Bioinformatics in AI Era.* Shanghai, China, May 19th, 2018.

16. Keynote Speech: *Challenges from Immunotherapy, CPM* 2018 (COCOON'2018), July 2-4, 2018, Qingdao, China.

17. Keynote Speech: 2017 International Forum on Innovation and Emerging Industrial Developments (IEID 2017), Shanghai, China, Nov 7-9, 2017.

18. Keynote Speech: *Deep context resolution.* Alibaba Yunxi Conference: Oct. 13, 2017, Hangzhou, China.

19. Keynote Speech: *Deep learning in bioinformatics.* The 8th IEEE Int'l Conf. on Big Knowledge, August 9-10, 2017.

20. Keynote Speech: *De novo peptide sequencing by deep learning.* The first ACSIC Symp. on Frontiers in Computing (SOFC), June 9-10, 2017, Chicago.

21. Keynote Speech: *Contextual sensitive chatting.* Artificial Intelligence Summit, Beijing Asia Grand Hotel, Beijing, China, March 28, 2017.

22. Keynote Speech: *De novo and antibody sequencing.* Fudan University Medical School, Proteomics Workshop, March 27, 2017.

23. Keynote Speech: *Chatting robots by deep learning.* Global Artificial Intelligence and Robotics Summit, Chinese Computer Federation, Shenzhen, China, August 12-13, 2016.

24. Keynote Speech: Chatting robots by deep learning. 7th Workshop on data mining and intelligent computing. Hefei, China, August 9-10, 2016.

25. Keynote Speech: *Chatting robots by deep learning.* The 3rd workshop on big data and computational intelligence. Beijing, July 29-July 31, 2016.

26. Keynote Speech: *Approximating Semantics.* 13th Chinese Computational Linguistics, Wuhan, China, Oct 18-19, 2014.

27. Keynote Speech: 10th Conf on Algorithmic Aspects of Information and Management. Vanckeeouver, July 8-11, 2014.

28. Invited tutorial: Unconventional Computation & Natural Computation, (Tutorial) London, Ontario. July 14-18, 2014

29. Keynote Speech: *Spaced Seeds.* ISBRA 2014, Zhang Jia Jie, China. June 28-30, 2014

30. Public Tianyi Lecture, Ningbo City, China, Oct. 21, 2013 *Latent Search.*

31. University of New Brunswick Annual Lecture in Computer Science: *Information Distance.* April 11-12, 2013.

32. Special Faculty Wide Colloquium: *Information distance from a question to an answer.* April 18th, 2012.

33. Keynote Speech: *Ideas in Bioinformatics.* Sharcnet Research Day, May 23, 2012. Guelph University, Ontario, Canada.

34. Keynote Lecture: *Spaced Seeds.* IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2012), Philadelphia, US, Oct. 4-7, 2012.

35. Keynote Speech: ICDM'2011 Workshop on Biological data mining and its applications in healthcare. Vancouver, Dec. 11, 2011.

36. Keynote Lecture: *Information Distance.* Solomonoff 85th Memorial Conference, Melbourne, Australia. Nov. 30–Dec. 2, 2011.

37. Invited Lecture: *Information distance.* The Fourteenth International Conference on Discovery Science (DS 2011), Oct. 5-7, 2011, Espoo, Finland.

38. Keynote Lecture: *Kolmogorov complexity and its applications in computer science.* Laurier Centennial Conference: Applied Mathematics, Modeling and Computational Science Conference. July 25-29, 2011.

39. Invited Lecture: *Spaced Seeds. International Bioinformatics Workshop'2011*, Xi'an, China, July 11-13, 2011.

40. Keynote Lecture: *Modern Homology Search.* The 4th Annual Meeting of the Asian Association for Algorithms and Computation (AAAC 2011), April 16-17, 2011, Hsinchu, Taiwan.

41. Keynote Lecture: *Information Distance and Chain Letters.* The 4th Annual Int'l Conf. on Combinatorial Optimization and Applications. Dec. 18-20, 2010, Hawaii.

42. Keynote Speech: ACM International Workshop on Data and Text Mining in Bioinformatics Oct. 29-30, 2010, Toronto.

43. Keynote Speech: *Kolmogorov Complexity and Information Distance.* IEEE Int'l Conf. on Granular Comput. August 14-16, 2010, San Jose.

44. Keynote Speech: 5th Canadian Student Conf. on Biomedical Comput. and Engineering, May 20-22, 2010, Waterloo, Ontario.

45. Keynote Lecture: 12th Descriptional Complexity of Formal Systems August 8-10, 2010. Saskatoon, Saskatchewan.

46. Keynote Lecture: *Optimized Spaced Seeds.* 1st International Conference on Bioinformatics and Computational Biology. New Orleans. April 8-10, 2009.

47. Distinguished Lecture: *Can NMR Protein Structure Determination be Automated?* Computer Science Department, University of Alberta. March 16, 2009.

48. Keynote Lecture: *Thermodyanics of Computing. 14th Int'l Conf. DNA Computing (DNA14)*, Prague, June 2-6, 2008.

49. Keynote Lecture: *19th Int'l Conf on Genome Informatics (GIW2008)*. Brisbane, Australia, Dec. 1-3, 2008.

50. Distinguished lecture: *Optimized Spaced Seeds* Computer Science Department, University of Maryland at Baltimore County. Feb. 5, 2008.

51. Distinguished lecture: Computer Science Department, University of Texas, Dallas, Sept 14, 2007.

52. Keynote speaker: *Information Distance.* The 13th Annual International Computing and Combinatorics Conference (COCOON'07), July 16-19 2007, Banff, Alberta.

53. Invited Speaker: *Information Distance. 11th Int'l Conf. Develop. Language Theory (DLT)*, Turku, July, 3-6, 2007.

54. Keynote speaker: *Modern Homology Search.* Ohio Collaborative Conference on Bioinformatics, Miami University, Oxford Ohio, July 9-11, 2007.

55. Keynote speaker: *Modern Homology Search. ISBRA'2007.* Atlanta, May 7-10, 2007.

56. Distinguished Speaker: *Modern homology search*, Distinguished Lecture Series, Simon Fraser University, Oct. 26, 2006.

57. Keynote Speaker: *Fast homology search.* 31st Int'l Symp. on Math. Found. Comput. Sci. (MFCS'2006), Slovakia, Aug. 28–Sept. 1, 2006.

58. Distinguished Speaker: *Workshop on Bioalgorithmics*, July 12-15, 2006. Singapore.

59. Invited Speaker, *Complexity of motif ranking*, 2006 RECOMB Satellite Workshop on Regulatory Genomics: Singapore, July 17-18, 2006.

60. Keynote Speaker: *The information distance and applications.* The 11th International Conference on Implementation and Application of Automata (CIAA'06), Taipei, Taiwan, August 21-23, 2006.

61. Keynote Speaker: *Information distance.* IEEE Int'l Conf. on Granular Computing, Atlanta, May 10-12, 2006.

62. Invited Speaker: *Super seeds in bioinformatics and finance.* Symposium on Combinatorial Pattern Matching (CPM'05) Jeju island, Korea, June 19-22, 2005.

63. Invited Plenary Speaker: *Theory of spaced seeds.* International Symposium on Recent Trends in Theoretical Computer Science. Feb. 28 – Mar. 3, 2005, Kyoto, Japan.

64. Keynote Speaker: *Genome to annotation.* SIG meeting in structural bioinformatics, Jan. 17, 2005, Singapore.

65. Distinguished Speaker: *PatternHunter – fast and sensitive homology search.* Queens University, School of Computing Distinguished Seminar Series, http://www.cs.queensu.ca/seminars/dss/schedule.html, Nov. 12, 2004.

66. Invited Plenary Speaker: *Kolmogorov complexity and its applications.* The 3rd International Congress of Chinese Mathematicians (ICCM), http://www.ims.cuhk.edu.hk/conference/iccm2004/, Hong Kong, December 17-22, 2004.

67. Keynote Speaker: *Fast and sensitive homology search.* 4th IEEE International Conference on Data Mining (ICDM), http://icdm04.cs.uni-dortmund.de/, Brighton, UK, Nov. 2-4, 2004. Invited Talk: *A theory of parameter-free data mining.* Workshop on Foundations of Data Mining, Brighton, UK, Nov. 1, 2004.

68. Keynote speaker: *Fun and elegant ideas in bioinformatics.* The 3rd annual Toronto Undergraduate Biotechnology Symposium, March 26, 2004.

69. Keynote Speaker: *Fast and sensitive homology search.* 2003 Bioinformatics In Taiwan (BIT2003) Symposium, Taiwan, Sept 5-7, 2003. http://ymbc.ym.edu.tw/bit/3e.html.

70. Invited Public Lecture: *Chain letters and evolutionary histories.* Tsinghua University, Beijing, China, Oct. 2003.

71. Invited Plenary Lecture: *Bioinformatics Tools and Software.* Canada-Japan Science Symposium, Canadian Embassy in Japan. Jan. 2003. Japan.

72. Keynote Speaker: *Building tools to mine molecular sequence data.* 2002 Australian Joint Artificial Intelligence Conference, Canberra, Australia, Dec. 2-6, 2002. http://www.cs.adfa.edu.au/ abbass/AI02/

73. Invited Lecture: *Fast and sensitive homology search.* Theoretical Computer Science Symposium 100 Anniversary of Nihon University, Tokyo, Japan, 2002.

74. Invited Plenary Speaker: *PatternHunter: Any genome anywhere.* Second Annual "The Computational Challenges in the Post Genomic Age", Durham, NC. March 11-13, 2002.

75. Keynote Speaker: *Bioinformatics: new challenges to computer science.* Beijing Bioinformatics Conference. Nov. 27, 2000, Beijing, China.

76. Keynote Speaker: *Whole genome phylogeny,* Case Western Reserve-Netgenics-Athersys mini-symposium on computational genetics, Cleveland, Oct. 20, 2000.

77. Keynote Speaker/Panelist: *Bioinformatics challenges to computer science.* Bioinformatics Workshop, Wayne State University, Detroit, Oct. 26, 2000.

78. Keynote Speaker: *Bioinformatics: new challenges to computer science.* Bioinformatics Workshop, Dec. 2000. National Chung Zheng University, Taiwan.

79. Invited Plenary Speaker: *The incompressibility argument.* International Conference on Theoretical Computer Science (in honor of Manuel Blum's 60th Birthday). April 20-24, 1998, Hong Kong.

80. Keynote Speaker: *Average-case analysis using Kolmogorov complexity* CATS'98 – Computing: The Australasian Theory Symp. Feb. 2-3, 1998, Perth, Australia.

81. Keynote Speaker: *Average-case analysis via incompressibility method,* 11th Symp. Fundamentals of Computation Theory (FCT'97), Krakow, Poland, Sept. 1-3, 1997.

82. Distinguished Speaker: *Kolmogorov complexity and its applications,* Center for System Science and School of Computing Science Distinguished Speaker Series, Simon Fraser University, Nov. 21, 1996.

83. Public Lecture: *A tour through the world of Kolmogorov complexity,* City University of Hong Kong. Nov. 13, 1996.

84. Invited Plenary Lecture: *Inferring a DNA sequence from erroneous copies,* 6th Annual International Workshop on Algorithmic Learning Theory, Fukuoka, Japan, Oct. 18-20, 1995.

85. Invited Lecture: *Kolmogorov complexity and its applications,* ALCOM Summer School on Complexity, Barcelona, Sept. 5-9, 1994.

86. Invited Plenary Lecture: *DNA sequencing and learning,* DIMACS "Combinatorial methods for DNA mapping and sequencing" workshop, Oct. 6-9, 1994.

87. Keynote Lecture: *PAC-learning theory and its applications*, Canadian Workshop on Machine Learning, Banff, Calgary, May, 1994.

88. Keynote Lecture: *Theories of learning*, 3rd International Conference for Young Computer Scientists, Beijing, July, 1993.

89. Keynote Lecture: *Thermodynamics of computing and information distance*, 1993 Midwest theory conference, Notre Dame, Indiana, April 3, 1993.

90. Keynote Lecture: *Information distance*, 1993 Carleton Annual Theory Symposium, Ottawa, Oct. 16, 1993.

91. Keynote Lecture: *A tour through the world of Kolmogorov complexity*, 19th International Colloquium on Automata, Languages, and Programming (ICALP'92), Vienna, July 13-17, 1992.

**Patents**

M. Li, B. Ma and J. Tromp, A method for fast and sensitive homology search. US patent, issue by USPTO, March, 2017. Patent number US9652586B2.

N.H. Tran, M. Li, et al, Methods and systems for assembly of protein sequences. US patent pending. 15/599341, filed 2017-05-18.

N.H. Tran, M. Li, et al, Methods and systems for assembly of protein sequences. Canadsa patent pending. 2967752, filed 2017-05-18.

N.H. Tran, M. Li, et al, Methods and systems for de novo peptide sequencing using deep learning. US patent, application: 16/037949, filed 2018-07-19.

N.H. Tran, M. Li, et al, Methods and systems for DIA de novo peptide sequencing by deep learning. US and international patent. Filed, Dec. 2018.

M. Li, Y. Tang, D. Wang System, method, and computer program for correcting speech recognition information. US patent, provisional patent filed in Oct. 2011.

M. Li, Y. Tang, D. Wang System, method and computer program for correcting machine translation information. Issued under US Patent No 9256597 (Jan 27, 2016)

M. Li, Y. Yang, D. Wang System and method for universal translating from natural language questions to structured queries. Issued by USPTO on Nov. 10, 2015, Patent No. 9,183,511.

**Selected Trainees**.

- Fatema Zohora, Ph.D. 2017 –

- Zeping Mao, Ph.D. 2021 –

- Shuyang Zhang, MMath, 2021–

- Yonghan Yu, Ph.D., 2021 –

- Qianqui Zhang, MMath, 2022 –

- He Bai, Ph.D. 2019–

- Owain West, Ph.D. 2019–

- Anup Anand Deshmukh, graduate student, 2018-2020

- Sherry Xie, Ph.D. 2018–

- Natatie Zhang, Ph.D., 2019 –

- Arash Mollajafari Sohi, graduate student, 2018-2020

- Hieu Ngoc Tran. postdoc, 2016-2018.

- Nancy Zhang, postdoc, 2016-2017.

- Xiaopeng Yang, postdoc 2016-2018.

- J. Tromp, postdoc, Sponsored by NSERC International Postdoc Fellowship. 1993-1995. Renaissance Technologies.

- Karsten Verbeurgt. Ph.D. 1998

- B. Ma, Ph.D., 1999, and postdoc, 1999-2000. Professor at University of Waterloo.

- Kevin Lanctot, Ph.D. 2000. University of Waterloo

- Haoyong Zhang, M.Math. 1999. Microsoft.

- L. Zhang, postdoc, 1995. Professor at National University of Singapore.

- B. DasGupta, postdoc, 1995-1996. Professor at University of Illionois Chicago.

- Q. Gao, postdoc, 1987-1988. Professor at Academia Sinica, China.

- C.Z. Liang, M.Math. 2001. Professor at Insitute of Genetics, Academic Sinica, China.

- J. Badger, postdoc, 1998-2000. Research scientist at TIGR.

- G. Lin, postdoc, computational biology. Professor af University of Alberta.

- X. Chen, postdoc. 2000-2002. Faculty member at the Nanyang Institute of Technology.

- Dongbo Bu, postdoc, 2006-2009. Professor at Academia Sinica.

- Paul Kearney. Hired using my Steacie Fellowship replacement fund as a definite term assistant professor. Chief Scientific Officer, Integrated Diagnostics.

- J. Xu, Ph.D. 2003. Faculty member at Toyata Institute of Technology Chicago.

- T. Vinar, Ph.D. 2005. Faculty member Comenius University.

- B. Brejova, Ph.D. 2005. Faculty member at Comenius University.

- Jing Zhang. Ph.D. 2008. Google.

- Zefeng Zhang, Ph.D. 2008. Bioinformatics Solutions Inc.

- Hao Lin, PhD. 2008. Bioinformatics Solutions Inc.

- S.C. Li, Ph.D. 2010. Faculty member at City University of HK.

- X. Gao. Ph.D. 2011. Faculty member at KAUST.

- B. Alipanahi. Ph.D. 2012. Postdoc, University of Toronto. 23&Me

- Yang Tang, Research Associate, 2011-2013. Twitter.

- Sayd Bashir. Ph.D. 2011. Google.

- Christina Boucher. Ph.D. 2012. Faculty member at University of Colarado.

- Dan Holtby. Ph.D. 2013, Postdoc, University of Waterloo.

- Xuefeng Cui, Ph.D. 2014. Faculty member, Shangdong University

- Yahui Chen, M.Math. 2015, Google.

- Junnan Chen, M.Math. 2016, Amazon.

- Guangyu Feng, M.Math. 2017, Amazon.

- Kun Xiong, M.Math. 2014, CEO, RSVP Technologies Inc.

- Anqi Cui: PDF, 2015-2016, VP, RSVP Technologies Inc.

- Xin Chen: Visiting Assistant Professor 2017-2018

**Research Grants**

1. NSERC Discovery Grant, $74,000/year, 2011-2016; $64,000/year, 2016-2021.

2. CFI-CRC Chair bioinformatics lab equipment grant: $300,000, 2009. CRC Chair fund, $200,000/year, 2009-2016

3. NSERC Collaborative Grant. $53076 / year for 3 years, 2008

4. Premier's Discovery Award. total $250,000, 2009-2012.

5. NSERC IDRC Chair. $1,000,000 for 5 years, 2009-2014. $250,000 of this 1 million dollars will be spent at University of Waterloo over 5 years. Support source: NSREC.

6. NSERC Operating/Discovery Grant, $74,000/year, 2006-2011.

7. PI, CITO Champions of Innovation Program. "New generation bioinformatics software". $200,000, 2003-2005.

8. CFI-CRC Chair bioinformatics lab equipment grant: $300,000, 2003. CRC Chair fund, $200,000/year, 2002-2009

9. NSERC Operating/Discovery Grant, $68,000/year, 2001-2006.

10. Co-PI, Carbon Sequestration in Synechococcus: From Molecular Machines to Hierarchical Modeling. $220,000 USD, 2002-2005. The whole grant is 19.1 million over 3 years, led by Sandia National Lab, under DOE's Genome to Life (GTL) initiative.

11. PI, NSF, BDI, 0213903, Scalable homology search tools. $410,000 USD, 2002-2005.

12. PI, NSF, Theory of computation, CCR-0208595, Kolmogorov complexity and its applications. $201,430 USD, 2002-2005.

13. PI, LSI & Rigel Pharmaceuticals, PathwayFinder. $100,000 USD. (2002).

14. PI. NSF, ITR/ACS, 0124597: Supplement grant to ACI00-85801. $17,500 USD, 2001.

15. PI, NSF, ITR/ACS ACI00-85801: Computational techniques of applied bioinformatics. $312,500 USD, 2000-2004. (This is part of a larger project together with M. Clegg and T. Jiang, total $785k.)

16. Project leader, CITO grant: "Computational tools for molecular biology" $240,000, 1998-2000.

17. NSERC Operating Grant OGP0046506, $51,000/year, 1996-1999, $58,905/year, 1999-2001.

18. NSERC Steacie supplement grant ($92,000, 1997-1999).

19. CGAT Grant: "Efficient Algorithms for Multiple Sequence Alignment, Evolutionary Trees, and Restriction Mapping" ($300,000, 1994-1997). (with T. Jiang, *et al.*) CGAT—Canadian Genome Analysis and Technology project.

20. ITRC Grant (Ian Munro, Jeff Shallit, *et al*, $20,000/year, 1993 – 1998).

21. NSERC Operating Grant OGP0046506 ($37,000/year, 1993-1996).

22. NSERC International Scientific Exchange Award ($5400, 1992).

23. NSERC International Scientific Exchange Award ($5400, 1990).

24. NSERC Operating Grant OGP0046506, $35,428/year, 1990-1993.

25. NSERC Equipment Grant, $19,000, 1989.

26. NSERC Operating Grant OGP0036747, $22,890/year, 1988-1990.

27. PI, NSF Grant DCR-8606366 $80,000 USD, 1986-1988, with Yaacov Yesha.

28. PI: Ohio State University, Seed Grant $16,000 USD, 1986, with Yaacov Yesha.

PROFESSIONAL SERVICE AND OUTREACH

**Editorial Boards**

- (co-) Editor-in-Chief. *Journal of Bioinformatics and Computational Biology*, 2002 – 2016,

- Associate Editor-in-Chief, *Journal of Computer Science and Technology*, 2003 – present.

- *Journal of Computer and System Sciences*, 1992 – 2015.

- *SIAM Journal on Computing*, 2002 – 2008.

- *Information and Computation*, 1997 – 2006.

- *Journal of Combinatorial Optimization*, 1995 – 2014.

- *International Journal of Foundation of Computer Science*, 1999 – 2003.

- Editor: Science Press (China), Discrete Mathematics and Theoretical Computer Science Series.

- Editor: Higher Education Press (China): Frontiers of Modern Sciences Series.

- Special Issue co-editor (with Les Valiant): *Machine Learning*, 14:1(1994), special issue for *COLT'91*.

- Special Issue editor: *Journal of Computer and System Sciences*, 50:3(1995), for *COLT'92*.

- Special Issue co-editor (with Bill Gasarch): *Journal of Computer and System Sciences*, for *COLT'94*.

- Special Issue co-editor (with Paul Vitányi): *Journal of Computer and System Sciences*, special issue for *2nd European Computational Learning Theory Conference* (EuroCOLT'95).

- Special Issue co-editor (with Dingzhu Du): *Theoretical Computer Science*, special issue for *COCOON'95*, Vol. 181 No. 2 (30 July, 1997)

- Special Issue co-editor (with Tao Jiang): *International J. Foundation of Computer Science*, special issue on "Computational Biology", 1995.

- Special Issue editor: *Theoretical Computer Science* for ALT'97.

- Special Issue co-editor (with R. Karp, P. Pevzner, R. Shamir): *Journal of Computer and System Sciences*, Bioinformatics Special Issues, 2002, 2003, 2004.

**Other Professional Services**

- Member, the Future Science Prize committee. Since 2021.

- Conference Chair, Asia Pacific Bioinformatics Conference. Singapore, Jan. 17-21, 2005. CCL'2015 (Chinese Computational Linguistics Conference), Guangzhou. Nov. 2015.

- Recent service on conference program committees: *RECOMB'2017, RECOMB'2018, WABI'2016, RE-COMB'2014, CiE'2012, IDCM'2009, APBC'2009, FOCS'2009, GIW'2009, WADS'2009, WABI'2009, RE-COMB'2008, GIW'2007, WABI'2007, CPM'2007, KDD'2007, SODA'2007, CSB'2006, ISMB'2006, CPM'2006,* Spain. *CPM'2004,* Istanbul, Turkey. *2nd IEEE Computer Society Bioinformatics Conference* CSB2003, Aug. 11-14, 2003. *FCT'03,* Malmo, Sweden, August 12-15, 2003. *ISAAC* 2002, Vancouver, Nov. 21-23, 2002. *ACM Symp. Theory of Computation (STOC'01),* Greece, 2001; *40th IEEE Sypm. Foundation of Computer Science, (FOCS'99),* NY, 1999; *FST&TCS'99,* India, Dec. 13-15, 1999; *LATIN'2000,* Punta del Este, Uruguay, April 10-14, 2000; *ICYCS'99.* Nanjing, China, Oct. 20-23, 1999; *CATS'99.* Auckland, New Zealand, Jan., 1999; *ISAAC'98.* South Korea, December 14-16, 1998;

- Conference program committee chair for: *8th Algorithmic Learning Theory Conference* (ALT'97), Sendai, Japan, Oct. 6-8, 1997; *International Computing and Combinatorics Conference,* (COCOON'95), Xi'an, China, Aug. 24-26 1995.

- Refereeing: I have refereed hundreds of papers and proposals for many journals including *SIAM J. Comput., Theor. Comput. Sci., IEEE Trans. Inform. Theory, Inform. Computation, Journal of Computer and System Sciences, Inform. Process. Lett., Math. Syst. Theory, Journal of the ACM, Bioinformatics, Journal of Bioinformatics and Computational Biology, AI Journal, Machine Learning,* and agencies including *NSERC, CGAT (Canadian genome project), NIH, NSF, Hong Kong Research council, City University of Hong Kong, Hong Kong University, Australian National University*

- External examiner of Bioinformatics program, Nanyang Technical University of Singapore. 2002–present.

- Created the Waterloo's undergraduate (co-op) Bioinformatics program joint between Computer Science Department and Biology Department, accepting 50 students each year, since 2001.

- Advisory committee: Computer Science Department, Wayne State University, 2003-2006.

- Scientific Advisory Committee: Designing Oil Seeds for Tomorrow's Market (Genome Canada / Genome Alberta, 14 million project): 2008–2010.

- Advisory Board: Book series: Theory and Applications of Computability, Springer. 2009-present.

**Media Coverage**

- ACM Tech News: Using Machine Learning to Develop Personalized Cancer Vaccines University of Waterloo Cheriton School of Computer Science (Canada). At: https://technews.acm.org/archives.cfm?fo=2021-03-mar/mar-12-2021.html, March 11, 2021

- University of Waterloo Daily Bulletin: Using machine learning to develop personalized cancer treatment. At https://uwaterloo.ca/daily-bulletin/2021-04-21#using-machine-learning-to-develop-personalized-cancer-treatment, April 21, 2021

- NSERC impact story https://www.nserc-crsng.gc.ca/Media-Media/ImpactStories-ArticlesPercutants_eng.asp: Personalizing cancer treatment through machine learning June 17, 2021

- Scientific Chinese: Cover page, No. 11, 2016.

- CCTV, 10 minute interview on Chatting Robots. Broadcast Feb. 27, 2017.

- CCTV 4. Together with Legend CEO Yang Yuanqing 2005

- Radio Canada International Interview for IDRC, at: http://www.rciviva.ca/rci/ch/dossiers/80504.shtml

- Waterloo Region Record, Friday, April 16, 2010, Front page story

- *Globe and Mail*, April 13, 2010 (coverage of Killam prize), 2006 (coverage of Royal Society of Canada), 2013 (University of Waterloo)

- News coverages (for Steacie, Killam awards) in the Chinese news papers including Tsingdao Daily, World Daily, People's Daily, Suzhou Daily

- K-W FM 98.3 Radio Chinese 1/2 hour interview.

PUBLICATIONS

**Books**

1. M. Li and P. Vitányi, *An introduction to Kolmogorov complexity and its applications*, Springer-Verlag, 1st Edition 1993 (xx+546 pp). 2nd Edition 1997 (xx+637 pp). 3rd Edition 2008 (xx+792 pp). 4th Edition, 2019. This resesarch monograph is considered the standard reference in the field. It has been used to teach graduate research courses in many top universities. At http://citeseer.nj.nec.com/source.html, this book was ranked as one of the most cited publications in computer science (31st most cited computer science source document). 5 star rating by all readers at amazon.com.

2. M. Li and H.A. Li, *A course on Java programming*. Science Press, 1997 (In Chinese, 252 pp).

3. M. Li and P. Vitányi, *Descriptional complexity*. Science Press, 1999 (In Chinese, 220 pp, Translator: Qi Cheng).

4. M. Li. *Lower Bounds in Computational Complexity*. Ph.D. thesis. Cornell Univeristy, 1985.

**Papers in Popular Magazines**

5. C.H. Bennett, M. Li and B. Ma, Chain letters and evolutionary histories. *Scientific American*, 288:6(June 2003) (feature article), 76-81.

6. Dana MacKenzie, On a roll, *New Scientist*, Nov. 6, 1999, 44-47. (Article devoted to our work on Kolmogorov complexity and Heilbronn triangles.) Also translated in to French appeared in *Courrier International* (12/23/99).

7. Wolfgang Blum, Die ZEIT, April 13, 2000 (#16), p. 40: "Geometrisch Eingekreist". (Article devoted to our work on Kolmogorov complexity)

8. Interview article on our PatternHunter system. Steve Meloan: Exploring the new frontier: Java technology powers the "Post-Genomic" era. http://java.sun.com/features/2001/09/genome.html. It appeared as the cover story of java.sun.com website which has 2 million viewers per day.

9. W. Kirchherr, M. Li, and P. Vitányi, The miraculous universal distribution. *The Mathematical Intelligencer*, 19:4(fall 1997), 7-15.

**Papers in Refereed Journals**

10. A. Ahmadzadegan, P. Simidzija, M. Li, A. Kempf, Learning to utilize correlated auxiliary noise: a posible quantum advantage. *Nature, Scientific Report*, 11:21624, 2021.

11. F. Zohora, Z. Rahman, NH Tran, L. Xin, B. Shan, M. Li. Deep neural network for detecting arbitrary-precision peptide features through attention based segmentation. *Nature, Scientific Report* Sept 14, 2021.

12. R. Qiao, N.H. Tran, L. Xin, P. Shan, M. Li, A. Ghodsi, Computationally instrument-resolution-independent de novo peptide sequencing for high-resolution devices. *Nature Machine Intelligence*. March, 2021.

13. N.H. Tran, R. Qiao, L. Xin, X. Chen, P. Shan, M. Li, Personalized deep learning of individual immunopeptidomes to identify neoantigens for cancer vaccines. *Nature Machine Intelligence* 2, 764-771(2020).

14. WJ Zhang, X. Chen, L. Xin, P. Shan, ZG. Luo, M. Li. ChimST: An efficient spectral library search tool for peptide identification for chimeric spectra in data-dependent acquisition. *ACM/IEEE Trans. on Comput. Biol. and Bioinform.*. 2020

15. M. Li, Deep learnings deciphers protein-RNA interaction, *Genomics, Proteomics, & Bioinformatics.*, 17(5), 2019, 475-477.

16. F. Zohora, Z. Rahman, NH Tran, L. Xin, B. Shan, M. Li. Deeplso: a deep learning model for peptide feature detection. *Nature, Scientific Report* 9, 17168, Nov. 20, 2019.

17. NH Tran, R. Qiao, L. Xin, X. Chen, C. Lui, X. Zhang, B. Shan, A. Ghodsi, M. Li. Deep learning enables de novo peptide sequencing from data-independent acquisition mass spectrometry. *Nature Methods*, Vol 16, January 2019, 63-66.

18. N.H. Tran, X. Zhang, M. Li, Deep Omics. *Proteomics*. 18:2(2018). 1700319. (Cover page story)

19. N.H, Tran, X. Zhang, L. Xin, B. Shan, M. Li. De novo peptide sequencing by deep learning. *Proc. National Academy of Sciences of the USA*, 114(31), 8247-8252, August 1, 2017.

20. C.Y. Yue, H. Cao, K. Xiong, A. Cui, H. Qin, M. Li, Enhanced question understanding with dynamic memory networks for textual question answering, 2017. *Expert Systems with Applications*, 80:1(2017), 39-45.

21. N.H. Tran, M.Z. Rahman, L. He, L. Xin, B. Shan, M. Li. Complete de novo assembly of monoclonal antibody sequences. *Nature, Scientific Reports 6* (Published on line: August 26, 2016), doi:10.1038/srep31730.

22. J. Wooley, M. Li, L. Wong, JBCB, the first decade. *Journal of Bioinformatics and Computational Biology* 12:03(2014).

23. X. Zhang, Z. Luo, M. Li Merge-weighted dynamic timing warping for speech recognition. *J Comput. Sci. Tech.* 29:6 (2014), 1072-1082.

24. B. DasGupta, X. He, M. Li, J. Tromp, L. Zhang. Nearest neighbor interchange and related distances. *Encyclopedia of Algorithms*, 2014. Springer.

25. X. Cui, S.C. Li, L. He, M. Li, Fingerprinting protein structures effectively and efficieintly. *Bioinformatics*, 30:7(2014) 949-955.

26. X.F. Cui, S.C. Li, D.B. Bu, M. Li, Towards reliable automatic protein structure alignment. *Nucleic Acids Research*, 37:12(2013), 1-7.

27. Y. Tang, D. Wang, J. Bai, X. Zhu, M. Li, Information distance between what I said and what it heard. *Communications of the ACM*. 56:7 (July, 2013), 70-77.

28. D. Holtby, S.C. Li and M. Li. LoopWeaver - Loop Modeling by the Weighted Scaling of Verified Proteins *J. Computat. Biol.*, 20:3(2013), 212-223.

29. X.F. Cui, S.C. Li, D.B. Bu, B. Alipanahi, M. Li. Protein structure idealization: How accurately is it possible to model protein structures with dihedral angles? *Alg. of Molecular Biology*, 8:5(2013)

30. B. Alipanahi, N. Krislock, H. Wolkowicz, A. Ghodsi, L. Donaldson and M. Li. Determining protein structures from NOESY distance constraints by semidefinite programming. *J. Computat. Biol.*, 20:4(2013) 296-310.

31. D. Bu, S.C. Li, M. Li, Clustering 100,000 potein structure decoys in minutes. *IEEE/ACM Trans. Comput. Biol. Bioinf.* 9:3(2012), 765-773.

32. F. Bu, X. Zhu, M. Li. A New Multiword Expression Metric and Its Applications *J. Comput. Sci. Tech.*, 26:1(2011), 3-13.

33. S.C. Li, D. Bu, M. Li Residues with similar hexagon neighborhoods share similar side-chain conformations. *ACM/IEEE Trans. Comput. Biology Bioinfo.* 9:1(2011), 240-248

34. Richard Jang, X. Gao, M. Li, Towards fully automated structure-based NMR assignment of 15N-labeled proteins from automatically picked peaks. *J. Comput. Biol.*, 18:3(2011), 374-363. Special Issue for *RE-COMB 2010*.

35. S.C. Li, D. Bu, J. Xu, and M. Li. Finding nearly optimal GDT scores. *J. Comput. Biol.*, 18:5(2011), 693-704.

36. M. Li, Can we determine a protein structure quickly? *J. Comput. Sci. Tech.* 25:1(2010) 95-106.

37. J.W. Zou, X. Deng, M. Li, Detecting market trends by ignoring it, some days. *J. Universal Computer Science*. 2010.

38. Y. Zhao, B. Alipanahi, S.C. Li, M. Li, Protein secondary structure prediction using NMR chemical shift data. *JBCB* 8:5(2010).

39. B. Alipanahi, X. Gao, E. Karakoc, F. Balbach, S.C. Li, G. Feng, L. Donaldson, M. Li. Error tolerant NMR backbone resonance assignment for automated structure generation. *J. Bioinform & Computat. Biol.*, 1(2011), 1V26.

40. X. Gao, J. Xu, S.C. Li, and M. Li. Predicting local quality of a sequence-structure alignment, *J. Bioinform & Computat. Biol.*, 7:5(2009) 789-810.

41. X. Gao, D. Bu, J. Xu, M. Li, Improving consensus contact prediction via server correlation reduction. *BMC Structural Biology*. 9:28(2009), 1-14.

42. P. Vitnyi, F.J. Balbach, R.L. Cilibrasi, M. Li. Normalized information distance. In *Information theory and statistical learning*, pp. 45-82. Springer US, 2009.

43. B. Alipanahi, X. Gao, E. Karakoc, L. Donaldson, and M. Li. PICKY: a novel SVD-based NMR spectra peak picking method. *Bioinformatics*. ISMB'09 Special issue. 25(2009): i268-i275.

44. D. Bu, M. Li, S.C. Li, J. Qian, J. Xu, Finding compact structural motifs. *Theoretical Computer Science*, 410:30-32(2009), 2834-2839.

45. S.C. Li and M. Li. On two open problems of 2-interval patterns. *Theoretical Computer Science*, 410:24-25(2009), 2410-2423.

46. Chen *et al*. The genome of Schistosoma japonicum reveals unique features of host-parasite interplay. *Nature*, Vol 460, July 16, 2009.

47. B. Brejova, T. Vinar, Y. Cheng, S. Wang, G. Zhao, D. Brown, M. Li, Y. Zhou, Finding genes in *Schistosoma Japonicum*: annotating novel genomes with help of extrinsic evidence. *Nucleic Acids Research*. 2009; doi: 10.1093/nar/gkp052.

48. X. Zhang, Y. Hao, X. Zhu, M. Li, New information measure and its application in question answering system, *J. Comput. Sci. Tech.*, 23:4(2008), 557-572.

49. H. Lin, Z. Zhang, M.Q. Zhang, B. Ma, M. Li. ZOOM! Zillions of oligos mapped. *Bioinformatics*. 24:21(2008), 2431-2437.

50. S.C. Li, D. Bu, J. Xu, and M. Li, Fragment-HMM: a new approach to protein structure prediction. *Protein Science*, 17(2008) 1925-1934.

51. Z. Zhang, H. Lin, M. Li, MANGO: multiple alignment with n gapped oligos. *Journal of Bioinformatics and Computational Biology*, 6:3(2008), 521-541.

52. S.C. Li, D. Bu, J. Xu, and M. Li, Designing succinct structural alphabet. *Bioinformatics*, *ISMB* Special Issue, 2008.

53. X. Cui, T. Vinar, B. Brejova, D. Shasha, and M. Li. Homology search for genes. *Bioinformatics* 23(2007), i97-i103.

54. B. Ma, L. Wang, M. Li, Near optimal alignment within a band in polynomial time. *J. Comput. Syst. Sci.*, 73(2007), 997-1011.

55. J. Zhang, B. Jiang, M. Li, J. Tromp, X. Zhang, M.Q. Zhang, Computing exact p-values for DNA motifs. *Bioinformaics*, 23(March 2007), 531 - 537.

56. B. Ma and M. Li. On the Complexity of Spaced Seeds. *J. Comput. Syst. Sci.*, 73(2007) 1024-1034. (Bioinformatics Special Issue)

57. B. Lucier, T. Jiang, M. Li, Average-case analysis of Quicksort and binary insertion tree height using incompressibility. *Inform. Process. Lett.* 103(2007) 45-51.

58. M. Li, Information distance and its applications. *Int'l J. Found. Comput. Sci.*, 18:4(2007), 669-681.

59. B. Brejova, D. Brown, M. Li, T. Vinar, ExonHunter: a comprehensive approach to gene finding. *Bioinformatics* 21:S1(2005), i57-i65.

60. Y. Hao, X.Y. Zhu, M.L. Huang, M. Li. Discovering patterns to extract protein-protein interactions from the literature: part II. *Bioinformatics*, 21:15(2005), 3294-3300.

61. J. Liu, B. Ma, M. Li. PRIMA: Peptide robust identification from MS/MS spectra. *J. Bioinform. Comput. Biol.*, 4:1(2006) 125-138.

62. M.L. Huang, X.Y. Zhu, M. Li, A hybrid method for relation extraction from biomedical literature. *International Journal of Medical Informatics* Special issue: Natural language processing on biomedicine and its applications. 75:6(2006), 443-455.

63. J. Liu and M. Li, Finding cancer biomarkers from mass spectrometry data by decision lists. Journal of Computational Biology. 12:7(2005), 969-977.

64. J. Xu, D. Brown, M. Li and B. Ma, Optimizing multiple spaced seeds for homology search. *Journal of Computational Biology*, 13:7(2006) 1355-1368

65. D. Kisman, M. Li, B. Ma, L. Wang, tPatternHunter: gapped, fast and sensitive translated homology search. *Bioinformatics*, 21:4(2005), 542-544.

66. M.L. Huang, X.Y. Zhu, Y. Hao, D.G. Payan, K. Qu, M. Li, Discovering patterns to extract protein-protein interactions from full texts. *Bioinformatics*. 20(2004) 3604-3612.

67. J. Xu, M. Li, Y. Xu, Protein Threading by Linear Programming: Theoretical Analysis and Computational Results. *Journal of Combinatorial Optimization*. 8:4(Dec. 2004), 403–418.

68. M. Li, X. Chen, X. Li, B. Ma, P.M.B. Vitanyi, The similarity metric, *IEEE Trans. Information Theory*, 50:12(2004), 3250-3264.

69. D. Brown, M. Li and B. Ma, A tutorial of recent developments in the seeding of local alignments. *J. Bioinformatics and Computational Biology*, 2:4(2004), 819-842.

70. X. Chen, B. Francia, M. Li, B. Mckinnon, A. Seker, Shared information and program plagiarism detection. *IEEE Trans. Information Theory*, 50:7(July 2004), 1545-1550.

71. M. Li, B. Ma, D. Kisman and J. Tromp, PatternHunter II: highly sensitive and fast homology search. *J. Bioinformatics and Computational Biology*, 2:3(2004), 417–440.

72. J. Badger, P. Kearney, M. Li, J. Tsang, T. Jiang, Selecting the branches for an evolutionary tree: a polynomial time approximation scheme. *Journal of Algorithms*, 51(2004), 1-14.

73. H. Lin, Z.F. Zhang, Q.F. Zhang, D.B. Bu and M. Li, A note on the single genotype resolution problem. *J. Computer Science and Technology*, 19:2(2004), 254-257.

74. U. Keich, M. Li, B. Ma and J. Tromp, On spaced seeds for similarity search. *Discrete Applied Math.* 138(2004), 253-263.

75. B. Ma, K. Zhang, A. Doherty-Kirby, C. Hendrie, M. Li, C. Liang, G. Lajoie. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Communication in Mass Spectrometry*. 17:20(2003), 2337-2342.

76. J. Xu, M. Li, D. Kim and Y. Xu, RAPTOR: optimal protein threading by linear programming. *Journal of Bioinformatics and Computational Biology*, 1:1(2003), 95-118.

77. C. Vogl, J. Badger, P. Kearney, M. Li, M. Clegg, T. Jiang, Probabilistic analysis indicates discordant gene trees in chloroplast evolution. *J. Mol. Evol.*, 56:3(2003), 330-340.

78. J. Xu and M. Li, Assessing RAPTOR's new linear programming approach for fold recognition in CAFASP3. *PROTEINS: Structure, Function, and Genetics*, 53(S6):579-584. October 2003 Invited paper by CASP5 organizers (after peer voted as the most interesting new method for 2002's CASP5 competition at www.forcasp.org). We have introduced a new method to do protein threading using linear programming. It ranked number 1 among the individual automatic protein 3D structure prediction programs in CASP5/CAFASP3 competition for fold recognition. RAPTOR was licensed to major pharmaceutical companies including Merck, Boehringer-Ingelheim, and Genentech.

79. M. Li, J. Tromp, P. Vitanyi, Sharpening Occam's razor. *Information Processing Letters* 85:5(2003), 267–274.

80. H. Sun, M. Li, Y. Xu, MOLVIE: an interactive visualization environment for molecular structures, *Computer Methods and Programs in Biomedicine*, 71:1(2003), 85-90.

81. K. Lanctot, M. Li, B. Ma, S. Wang, and L. Zhang, Distinguishing string selection problems. *Information and Computation*, 185(2003), 41-55.

82. Mouse Genome Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature*, 409:6822(2002), 520-522. (M. Li is a member of MGSC.)

83. G.S. Heffelfinger *et al*, Carbon Sequestration in Synechococcus Sp.: From Molecular Machines to Hierarchical Modeling. *OMICS*, 6:4(2002), 305-330. (M. Li is a coauthor.)

84. M. Li, B. Ma and L. Wang, Finding similar regions in many sequences. *Journal of Computer and System Sciences* (Special Issue for STOC'99). 65(2002), 73-96.

85. X. Chen, M. Li, B. Ma, J. Tromp, DNACompress: fast and effective DNA sequence compression. *Bioinformatics*. 18(2002), 1696-1698.

86. M. Li, B. Ma, L. Wang, On the Closest String and Substring Problems. *J. ACM*, 49:2(2002), 157-171.

87. B. Ma, J. Tromp, M. Li, PatternHunter: Faster and more sensitive homology search. *Bioinformatics*, 18:3(2002), 440-445. The optimized spaced seeds, invented here, have changed the way we do homology search. This new method is widely recognized as a major innovation in bioinformatics and is adopted by BLAST (http://www.ncbi.nlm.nih.gov/Web/Newsltr/FallWinter02/blastlab.html, news release). Today, our spaced seeds are serving thousands of homology search queries daily by the scientists all over the world via NCBI BLAST and PatternHunter. PatternHunter has been used by the Mouse Genome Sequence Consortium to compare the mouse genome and the human genome, taking 20 CPU-days instead of 20 CPU-years otherwise by BLAST at the time [80].

88. T. Jiang, M. Li, P. Vitányi, The average-case area of Heilbronn-type triangles. *Random Structures and Algorithms*, 20:2(2002), 206-219.

89. T. Jiang, P. Kearney, and M. Li, A polynomial time approximation scheme for inferring evolutionary trees from quartet topologies and its application. *SIAM J. Computing*, 30:6(2001), 1942-1961.

90. X. Chen, S. Kwong, and M. Li, A compression algorithm for DNA Sequences, *IEEE Engineering in Medicine and Biology Magazine*, (Special Issue on Bioinformatics) 20:4(2001), 61-66.

91. M. Li, J. Badger, X. Chen, S. Kwong, P. Kearney, H. Zhang, An information-based sequence distance and its application to whole mitochondrial genome phylogeny, *Bioinformatics*, 17:2(2001), 149-154.

92. H.C. Wang, J. Badger, P. Kearney, M. Li, Analysis of codon usage patterns of microbial genomes with self-organizing map: methods and applications to genome exploration. *Mol. Biol. Evol.*, 18:5(2001), 792–800.

93. Q. Gao, M. Li, P. Vitányi, Applying MDL to learning best model granularity. *Artificial Intelligence*, 121:1-2(2000), 1–29.

94. T. Jiang, M. Li, P. Vitányi, Average-case analysis of algorithms using Kolmogorov complexity. *Journal of Computer Science and Technology*, 15:5(2000), 402-408.

95. B. DasGupta, X. He, T. Jiang, M. Li, J. Tromp, and L. Zhang, On computing the nearest neighbor interchange distance. in D.Z. Du, P.M. Pardalos and J. Wang (eds.), Proceedings of the DIMACS Workshop on Discrete Problems with Medical Applications, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, American Mathematical Society, Vol. 55, pp. 125-143, 2000.

96. B. Ma, L. Wang, M. Li, Fixed topology alignment with recombination. *Discrete Applied Math.* (Special Issue on Computational Molecular Biology), 104(2000), 281-300.

97. B. Ma, M. Li, and L. Zhang, From gene trees to species trees. *SIAM J. Comput.*,30:3(2000), 729-752.

98. T. Jiang, M. Li, P. Vitányi, A lower bound on the average-case complexity of Shellsort. *J. Assoc. Comp. Mach.*, 47:5(2000), 905–911. The average-case analysis of Shellsort has been a major open question of 40 years since the beginning of computer science, 1959. Using our "incompressibility method", we gave a simple analysis of an $\Omega(pn^{1+\frac{1}{p}})$ average-case lower bound for $p$-pass Shellsort, for every $p$.

99. P. Vitányi and M. Li, Minimum decription length induction, Bayesanism, and Kolmogorov complexity. *IEEE Trans. Inform. Theory*, IT-46:2(2000), 446–464.

100. H. Buhrman, T. Jiang, M. Li, and P. Vitányi, New applications of the incompressibility method: Part II. *Theoretical Computer Science*, 235:1(2000), 59-70.

101. T. Jiang, P. Kearney, and M. Li, Some open questions in computational molecular biology, *Journal of Algorithms*, 34:1(2000), 194-201; and *SIGACT News*, 30:3(1999), 43-49.

102. M. Charikar, C. Chekuri, T.Y. Cheung, Z. Dai, A. Goel, S. Guha, and M. Li, Approximation algorithms for directed Steiner problems. *Journal of Algorithms*, 33(1999), 73-91.

103. B. DasGupta, X. He, T. Jiang, M. Li, and J. Tromp, On the linear-cost subtree-transfer distance between phylogenetic trees. *Algorithmica*, 25(1999), 176-195.

104. M. Li and L. Zhang, Twist-rotation transformations of binary trees and arithmetic expressions. *Journal of Algorithms*, 32(1999), 155-166.

105. T. Jiang, M. Li, and P. Vitányi, New applications of the incompressibility method. *The Computer Journal*, 42:4(1999).

106. H. Buhrman, M. Li, J. Tromp, and P. Vitányi, Kolmogorov random graphs and the incompressibility method. *SIAM J. Comput.*, 29:2(1999), 590-599.

107. B. DasGupta, T. Jiang, S. Kannan, M. Li, and Z. Sweedyk, On the complexity and approximation of syntenic distance. *Discrete Applied Mathematics* (special issue in computational biology), 88:1-3(1998), 59-82.

108. M. Li, J. Tromp, and P. Vitányi, Reversible simulation of irreversible computation. *Physica D*, 120(1998) 168-176.

109. R. Beigel, W. Gasarch, M. Li, and L. Zhang, Addition in $\log_2 n$ steps on average: a simple analysis, *Theoretical Computer Science*, 1-2(191), 1998, 245-248.

110. C.H. Bennett, P. Gács, M. Li, P. Vitányi, and W. Zurek, Information Distance. *IEEE Trans. Inform. Theory*, 44:4(July 1998), 1407-1423. (Also in STOC'93.) The information distance derived from the first physical principles here is now widely accepted as the standard concept to measure "ultimate distance" between any two information carrying entities. [89,5] have opened the door of using such tools in real applications from measuring distance between two genomes to measuring distance between a query and an answer on the internet. Hundreds of applications have followed. In a SIGKDD04 paper (pp. 206-215), Keogh, Lonardi and Ratanamahatana have demonstrated that our methodology was better than all 51 methods for time series clustering they have found in SIGKDD, SIGMOD, ICDM, ICDE, SSDB, VLDB, PKDD, PAKDD conferences.

111. J. Kececioglu, M. Li, and J. Tromp, Reconstucting a DNA sequence from erroneous copies. *Theoretical Computer Science*, 185:1(1997), 3-13.

112. M. Li, J. Tromp, and L. Zhang, On the nearest neighbour interchange distance between evolutionary trees. *Journal of Theoretical Biology*, 182(1996), 463-467.

113. M. Li, J. Tromp, and P. Vitányi, How to share concurrent asynchronous wait-free variables. *Journal of the ACM*, 43:4(1996), 723-746.

114. T. Jiang and M. Li, $k$ one-way heads cannot do string-matching. *Journal of Computer and System Sciences*, 53:3(Dec. 1996), 513-524.

115. J.H. You, R. Cartwright, and M. Li, Iterative belief revision in extended logic programming. *Theoretical Computer Science*, 170(1996), 383-406.

116. T. Hancock, T. Jiang, M. Li, and J. Tromp, On approximating decision lists and trees. *Information and Computation*, 126:2(1996), 114-122.

117. T. Jiang and M. Li, DNA sequencing and string learning. *Math. Syst. Theory*, 29(1996), 387-405.

118. M. Li and P. Vitányi, Reversibility and adiabatic computation: trading time and space for energy. *Proc. Royal Society of London, Series A*, (452)1996, 769-789.

119. T. Jiang and M. Li, On the approximation of shortest common supersequences and longest common subsequences. *SIAM J. Comput.*, 24:5(1995), 1122-1139.

120. M. Li and P. Vitányi, A new approach to formal language theory by Kolmogorov complexity. *SIAM J. Comput.*, 24:2(1995), 398-410.

121. M. Li and P. Vitányi, Statistical properties of finite sequences with high Kolmogorov complexity. *Math. Syst. Theory,* 27(1994), 365-376.

122. M. Kearns, M. Li, and L. Valiant, Learning Boolean formulas. *Journal of the ACM*, 41:6(1994), 1298-1328. (Also in STOC'87) This paper and [127] are pioneering papers in computational learning theory and have made major impact to the field. Many later papers have investigated and improved problems and results in these papers.

123. A. Blum, T. Jiang, M. Li, J. Tromp, and M. Yannakakis, Linear approximation of shortest superstrings. *Journal of the ACM*, 41:4(1994), 630-647. (Also in STOC'91).

124. T. Jiang and M. Li, Approximating shortest superstrings with constraints. *Theoret. Comput. Sci.*, 134(1994), 473-491.

125. M. Li and P. Vitányi, Kolmogorov complexity arguments in combinatorics. *J. Combinatorial Theory, Ser A*, 66:2(1994), 226-236. Erratum, 69(1995), p.183.

126. B. Chor, A. Israeli, and M. Li, Wait-free consensus using asynchronous hardware. *SIAM J. Comput.*, 23:4(1994), 701-712.

127. M. Geréb-Graus and M. Li, Three one-way heads cannot do string matching. *Journal of Computer and System Sciences*, 48(1994), 1-8.

128. T. Jiang and M. Li, On the complexity of learning strings and sequences. *Theoretical Computer Science*, 119(1993), 363-371.

129. M. Kearns and M. Li, Learning in the presence of malicious errors. *SIAM J. Comput.*, 22:4(1993), 807-837. (Also in STOC'88)

130. A. Israeli and M. Li, Bounded time-stamps. *Distributed Computing*, 6(1993), 205-209.

131. D. Foulser, M. Li, and Q. Yang, Theory and algorithms for plan merging. *Artificial Intelligence*, 57(1992), 143-181.

132. T. Jiang, M. Li, and D. Du, A note on shortest common superstrings with flipping. *Inform. Process. Lett.*, 44:4(1992), 195-199.

133. M. Li and P. Vitányi, Optimality of wait-free atomic multiwriter variables. *Inform. Process. Lett.*, 42(1992), 107-112.

134. M. Li and P. Vitányi, Inductive reasoning and Kolmogorov complexity. *Journal of Computer and System Sciences*, (special issue for Structure'89) 44:2(1992), 343-384.

135. M. Li and P. Vitányi, Average case complexity equals worst-case complexity under the Universal Distribution. *Inform. Process. Lett.*, 42(1992), 145-149.

136. M. Li, L. Longpré, and P. Vitányi, The power of the queue. *SIAM J. Comput.*, 21:4(1992), 697-712.

137. M. Li and P. Vitányi, Learning simple concepts under simple distributions. *SIAM J. Comput.* 20:5(1991), 911-935.

138. M. Li and Y. Yesha, Resource bounds for parallel computation of threshold and symmetric functions. *Journal of Computer and System Sciences*, 42:1(1991), 119-137.

139. M. Li and Y. Yesha, New lower bounds for parallel computation. *Journal of the ACM*, 36:3(1989), 671-680.

140. F. Fich, M. Li, P. Ragde, and Y. Yesha, On the power of concurrent-write PRAMs with read-only memory. *Information and Computation*, 83:2(1989), 234-244.

141. C. Bajaj and M. Li, Geometric optimization and Dp-completeness. *Discrete and Computational Geometry*, 4:1(1989), 3-13.

142. M. Chrobak and M. Li, $k + 1$ heads are better than $k$ for PDA's. *Journal of Computer and System Sciences* (Special issue for FOCS'86), 37:2(1988), 144-155.

143. M. Li and P. Vitányi, Tape versus queue and stacks: The lower bounds. *Information and Computation*, 78:1(1988), 56-85.

144. M. Li, Simulating $2$ pushdown stores by $1$ tape in $O(n^{1.5})$ time. *Journal of Computer and System Sciences* (Special Issue for FOCS'85), 37:1 (1988), 101-116. This paper solved several open questions in complexity theory and formal language theory, open for over 10 years. It showed, surprisingly, that one non-deterministic tape can simulate 2 pushdown stores in less than $n^2$ time, using a planar graph separator theorem.

145. M. Li, A separator theorem for one-dimensional graphs under linear mapping. *Inform. Process. Lett.*, 27(1988), 9-11.

146. M. Li and P. Vitányi, Kolmogorovskaya slozhnost' dvadsat' let spustia. *Uspekhi Math. Nauk* (*Russian Mathematical Surveys*), 43:6(1988), 129-166 (In Russian).

147. M. Li and Y. Yesha, Separation results for ROM and nondeterministic models of parallel computation. *Information and Control*, 73:2(1987), 102-128.

148. M. Li and Y. Yesha, String-matching cannot be performed by a 2-head 1-way DFA. *Inform. Process. Lett.*, 22(1986), 231-235.

149. M. Li, W. Grosky, and R. Jain, Normalized quadtrees with respect to translations. *Journal of Computer Graphics and Image Processing*, 20(1982), 72-81.

**Edited Books / Book Chapters**

150. D. Brown, M. Li and B. Ma, Homology search methods. In *The Practical Bioinformatician.* L. Wong Ed. World Scientific, 2004, 217-244.

151. M. Li, Compressing DNA sequences. Chapter 7 in *Current Topics in Computational Molecular Biology*, T. Jiang, Y. Xu, M.Q. Zhang (eds.), the MIT Press. 2002.

152. M. Li and P. Vitányi, Algorithmic complexity. In *International Encyclopedia of Social and Behavioral Sciences*. Elsevier Science, 2000.

153. B. DasGupta, X. He, T. Jiang, M. Li, J. Tromp, L. Wang, and L. Zhang, Computing distances between evolutionary trees. In *Handbook in Combinatorial Optimization*, D.Z. Du and P. Pardalos (eds), Jan. 1999.

154. T. Jiang, M. Li, P. Vitányi, Some examples of average-case analysis by the incompressibility method. In *Jewels are Forever: Contributions on Theoretical Computer Science in Honor of Arto Salomaa*, J. Karhumaki, H. Maurer, G. Paun, and G. Rozenberg (eds.), Springer, 1999, pp. 250-261.

155. T. Jiang and M. Li, Shortest common superstring. Chapter 8 in *Pattern matching algorithms*, A. Apostolico and Z. Galil (eds), pp. 237-266, 1997.

156. M. Li and A. Makuoka, Eds., *Algorithmic Learning Theory, Proceedings of the Eighth International Workshop*. Lecture Notes in Artificial Intelligence, Springer-Verlag, Heidelberg, 1997.

157. M. Li and P. Vitányi, Average-case analysis using Kolmogorov complexity. In *Advances in Complexity and Algorithms*, edited by D. Du and K. Ko, Kluwer, 1997.

158. T. Jiang, M. Li, and B. Ravikumar, Formal models and computability. Chapter 3 of *Handbook for Computer Science and Engineering*, Allen Tucker, Jr.(ed.), CRC Press, pp. 31-63, 1997.

159. T. Jiang, M. Li, and B. Ravikumar, Basic notions in computational complexity. Chapter 30 of *Algorithms and Theory of Computation Handbook*. Mike Atallah(ed.), CRC Press, 1998.

160. T. Jiang, M. Li, B. Ravikumar and K. Regan, Formal grammars and languages. Chapter 31 of *Algorithms and Theory of Computation Handbook*. Mike Atallah(ed.), CRC Press, 1998.

161. T. Jiang, M. Li, B. Ravikumar and K. Regan, Computability. Chapter 32 of *Algorithms and Theory of Computation Handbook*. Mike Atallah(ed.), CRC Press, 1998.

162. M. Li, Where do we go for dinner, Paul? In *A dynamic and quick intellect*, pp. 89-96, J. Tromp (ed.), 1996.

163. D. Du and M. Li, Eds., *Combinatorics and Computing, Proceedings of the First International Conference*, Lecture Notes in Computer Science, Springer-Verlag, Heidelberg, 1995.

164. M. Li and P. Vitányi, Computational machine learning in theory and praxis. In *Computer Science Today, Recent Trends and Developments*, Springer-Verlag, LNCS Memorial Volume 1000, (Editor: J. van Leeuwen) pp. 518-535, 1995.

165. T. Jiang and M. Li, Optimization problems in molecular biology. In *Advances in Optimization and Approximation*, edited by D.Z. Du and J. Sun, Kluwer Academic Publishers, 1994, pp. 195-216,

166. M. Li and P. Vitányi, A brief introduction to Kolmogorov complexity and its applications. In book *Chinese Mathematics into the 21st Century*, invited article, (Edited by Wu Wen-tsun and Cheng Min-de), Peking University Press, 1991, pp. 85-140.

167. M. Li and P. Vitányi, Kolmogorov complexity and its applications. In *Handbook of Theoretical Computer Science*, (Managing Editor: J. van Leeuwen), 1990, pp. 189-254, Elsevier/MIT Press. Translated into Russian and Japanese.

168. M. Li and P. Vitányi, Applications of Kolmogorov complexity in theory of computing. In *Complexity Theory Retrospective*. Springer-Verlag Alan Selman (ed.), 1989, pp. 147-203.

**Papers in Peer-Reviewed Conference Proceedings**

169. MH Li, M. Li, K. Xiong, J. Lin Multi-task dense retrival via model uncertainty fusion for open-domain question answering. *EMNLP'2021*. Nov. 7-11, 2021. (Findings of EMNLP)

170. A. Deshmukh, Q.Q. Zhang, M. Li, J. Lin, L. Mou, Unsupervised chunking as syntactic Structure induction with a knowledge-transfer approach. *EMNLP'2021*. Nov. 7-11, 2021. (Findings of EMNLP)

171. H. Bai, P. Shi, J. Lin, Luchen Tan, K. Xiong, W. Gao, J. Liu, M. Li, Semantics of the Unwritten: the effect of end of paragraph and sequence tokens on text generation with GPT2, ACL Workshop 2021.

172. H. Bai, P. Shi, J. Lin, K. Xiong, W. Gao, M. Li, SegaTron: Segment-aware transformer for language modeling and understanding, *AAAI'2020*.

173. Z. Tu, W. Yang, Z. Fu, Y. Xie, L. Tan, K. Xiong, M. Li, J. Lin, Content analysis, recommendation and classification, search and ranking. *ICTIR'2020*. 2020.

174. H. Shadid, M. Li, J. Lin, Two birds, one stone: a simple, unified model for text generation from structured and unstructured data. *ACL'2020*.

175. Y. Xie, W. Yang, L. Tang, K, Xiong N, J. Yuan B. Huai, M. Li, J. Lin. Distant supervision for multi-stage fine-turing in retrieval-based question answering. *WWW 2020*.

176. NH Tran, R. Qiao, L. Xin, X. Chen, C. Lui, X. Zhang, B. Shan, A. Ghodsi, M. Li. Deep learning enables de novo peptide sequencing from data-independent acquisition mass spectrometry. *RECOMB'19*, Highlight track, 2019.

177. W. Yang, R. Qiao, H. Qin, A. Sun, L. Tan, K. Xiong, M. Li, End-to-end context reconstruction in Chinese dialogue. *ACL workshop*. 2019.

178. W. Yang, L. Tan, C. Lu, A. Cui, H. Li, X. Chen, K. Xiong, M. Wang, M. Li, J. Pei, and J. Lin. Detecting customer complaint escalation with recurrent neural networks and manually-engineered features. 2019 Annual Conf North American Chapter of the Association for Computational Linguistics. June 2-7, 2019, Minneapolis, MN.

179. W. Yang, Y. Xie, A. Lin, X. Li, L. Tan, K. Xiong, M. Li, J. Lin. End-to-end open domain question answering with BERTserini. 2019 Annual Conf North American Chapter of the Association for Computational Linguistics. June 2-7, 2019, Minneapolis, MN.

180. W. Yang, Y. Xie, L. Tan, K. Xiong, M. Li, J. Lin. Data augmentation for BERT fine-tuning in open-domain question answering. arXiv:1904.06652, 2019.

181. N.H, Tran, X. Zhang, L. Xin, B. Shan, M. Li. De novo peptide sequencing by deep learning. *RECOMB'18*, Highlight track, 2019.

182. X. Yang, X. Lin, S. Suo, M. Li Generating thematic Chinese poetry using conditional variational autoencoders with hybrid decoders. *IJCAI'2018*.

183. B. Ye, G. Feng, A. Cui, M. Li, Learning question similarity with recurrent neural networks. IEEE ICBK, 2017, regular paper. Hefei, China, August 9-10, 2017.

184. G. Feng, K. Xiong, Y. Tang, A. Cui, J. Bai, H. Li, Q. Yang, M. Li. Question classification by approximating semantics. *Proc. WWW'2015* pp. 407-417.

185. X.F. Cui, S.C. Li, D.B. Bu, M. Li, Towards reliable automatic protein structure alignment. *WABI 2013*, Sept. 2-4, 2013, Sophia Antipolis, France, pp. 18-32

186. X.F. Cui, S.C. Li, D.B. Bu, B. Alipanahi, M. Li, How accurately can we model protein structures with dihedral angles? *WABI 2012*, Sept. 10-12, 2012, Ljubljana, Slovenia, pp. 274-287.

187. X. Zhang, H. Wang, T.J. Collins, Z. Luo, and M. Li, Classifying Stem Cell Differentiation Images by Information Distance. *PKDD'2012*, Sept. 24-28, Bristol, UK, pp. 269-282.

188. B. Alipanahi, N. Krislock, H. Wolkowicz, A. Ghodsi, L. Donaldson and M. Li, Protein Structure by Semidefinite Facial Reduction. RECOMB 2012. pp. 1-11, Barcelona, Spain.

189. D. Holtby, S.C. Li and M. Li, LoopWeaver - Loop Modeling by the Weighted Scaling of Verified Proteins RECOMB 2012. pp. 113-126. Barcelona, Spain, April 21-24

190. M. Li, Information distance and its extensions. *Discovery Science 2011*. Epsoo, Finland, Oct 5-7, 2011. pp 18-28 1 page in *ALT'2011*.

191. F. Bu, X. Zhu and M. Li, Measuring the Non-compositionality of Multiword Expressions. *COLING 2010*. Best paper award selected from over 800 submissions, Beijing, August 21-26, 2010. pp. 116-124

192. Richard Jang, X. Gao, M. Li, Towards automated structure-based assignment. *RECOMB 2010*, pp. 189-207. Lisbon, Portugal, April 25-28.

193. C. Long, M. Huang, X. Zhu, M. Li, Multi-document summarization by information distance. *IEEE Int'l Conf. Data Mining, 2009 (ICDM 2009)*. Miami, Florida, Dec. 6-9, 2009. Short paper. pp 866-871.

194. C. Long, J. Zhang, M. Huang, X. Zhu, M. Li, B. Ma, Specialized review selection for feature rating estimation. *IEEE/WIC/ACM Int'l Conf. Web Intelligence (WI-IAT 2009)*. Regular paper. pp 214-221. Milan, Italy, Sept 15-18.

195. C. Long, X.Y. Zhu, M. Li, B. Ma, Information shared by many objects. *ACM 17th Conf. Info. and Knowledge Management, (CIKM'2008)*. Oct. 26-30, 2008, Napa Valley, California. Regular paper.

196. B. Alipanahi, X. Gao, E. Karakoc, L. Donaldson, and M. Li, PICKY: a novel SVM-based NMR spectra peak picking method. *ISMB'09*.

197. S.C. Li, D. Bu, J. Xu, M. Li, Designing succinct structural alphabet. *ISMB'08*, Toronto, July 2008.

198. S.C. Li, D. Bu, J. Xu and M. Li, Finding largest well-predicted subset of protein structure models *CPM'08*, Italy June, 2008.

199. J. Zhang, X. Gao, J. Xu and M. Li, Rapid and accurate side chain prediction with local backbone information. *RECOMB'08*, 2008.

200. X. Zhang, Y. Hao, X. Zhu, M. Li, Information distance from a question to an answer. In *Proc. 13th ACM SIGKDD*. pp. 874-883, Aug. 12-15, 2007. (Finalist for the best paper award.)

201. X. Gao, D. Bu, S.C. Li, J. Xu, M. Li, FragQA: predicting local fragment quality of a sequence-structure alignment. In *GIW'2007*. Best paper award. Singapore, Dec. 3-5, 2007

202. Z. Zhang, H. Lin, M. Li, MANGO: a new approach to multiple sequence alignment. *CSB'2007*. Best paper award. San Diego, CA, Aug. 13-15, 2007.

203. X. Gao, D. Bu, S. Li, M. Li, J. Xu, Consensus contact prediction by linear programming. *CSB'2007*. San Diego, CA, Aug. 13-15, 2007.

204. X. Cui, T. Vinar, B. Brejova, D. Shasha, M. Li, Homology search for genes. *ISMB'2007*.

205. M. Li, Information distance and its applications. *11th. Int'l Conf. Impl. Appl. Automata*, O.H. Ibarra and H.C. Yen Eds, LNCS 4094, 2006, pp. 1-9.

206. J. Qian, S.C. Li, D. Bu, M. Li, J. Xu, Finding compact structural motifs in many proteins. *CPM'2007*

207. J. Zhang, X. Chen, M. Li, Calculating exact p-value for structured motif. *CPM'2007*

208. M. Li, B. Ma, L. Zhang, Superiority and complexity of the spaced seeds. *SODA*, 2006, pp. 444-453.

209. B. Brejova, D. Brown, M. Li, T. Vinar, ExonHunter: a comprehensive approach to gene finding. *ISMB'2005*, June 25-29, 2005, Michigan.

210. T. Tang, J. Xu, M. Li, Discovering sequence-structure motifs from protein segments and two applications. *PSB'05*. Hawaii, Jan. 2005.

211. J. Liu, B. Ma, M. Li, PRIMA: Peptide robust identification from MS/MS spectra. *APBC'05*, Singapore, Jan. 17-21, 2005.

212. J. Xu, L. Yu, M. Li, Consensus fold recognition by predicted model quality. *APBC'05*, Singapore, Jan. 17-21, 2005.

213. Y.S. Jiao, M. Li, J.Y. Xu, On the k-Closest Substring and k-Consensus Pattern problems. *CPM'04*, Istanbul, Turkey, 2004.

214. J. Xu, D. Brown, M. Li, B. Ma, Optimizing multiple spaced seeds for homology search. *CPM'04*, Istanbul, Turkey, 2004.

215. D. Yao, J. Wang, Y. Lu, N. Noble, H. Sun, X. Zhu, N. Lin, D. Payan, M. Li, K. Qu, PathwayFinder: paving the way towards automatic pathway extraction. In *The Second Asia-Pacific Bioinformatics Conference* (APBC2004), Jan. 18-22, 2004, Dunedin, New Zealand.

216. M. Li, B. Ma, D. Kisman, J. Tromp, PatternHunter II: highly sensitive and fast homology search. *GIW'03*.

217. M. Li, X. Chen, X. Li, B. Ma, P. Vitanyi, Finding similarity by compression. *IEEE Int'l Symp. on Information Theory*, Yokohama, Japan, in June, 2003.

218. J. Xu, M. Li, G. Lin, D. Kim, Y. Xu, Protein threading by linear programming. *PSB'03*, Hawaii. Jan. 2003, pp 264-275.

219. M. Li, X. Chen, X. Li, B. Ma, P. Vitanyi, Similartiy Metric. ACM-SIAM *SODA'03*. Jan. 12-14, 2003, Baltimore, MD, pp. 863-872.

220. C. Liang, J. Smith, C. Hendrie, K.W.M. Siu, and M. Li, A comparative study of mass spec de novo sequencing software. *51st ASMS Conference*, Poster, Montreal, Canada, June 8-12, 2003.

221. B. Ma, K. Zhang, G. Lajoie, A. Doherty-Kirby, C. Liang, M. Li, PEAKS: A powerful software tool for *de novo* sequencing of peptides from MS/MS data, *50th ASMS Conference*. June. 2002.

222. M. Li, J. Tromp, P. Vitanyi, Sharpening Occam's razor. *Proc. COCOON'2002*, Singapore, 2002, pp. 411-419.

223. M. Li, B. Ma, L. Wang, Near optimal alignment within a band in polynomial time. *Proc. 32nd ACM Symp. Theory of Computing* (STOC'00), Portland, Oregon, 2000, pp. 425-434.

224. X. Chen, S. Kwong, and M. Li, A compression algorithm for DNA sequences and its application in genome comparison. In GIW'99 (best paper award), Tokyo, Japan, Dec. 1999, pp. 51-61, and in *RECOMB'00*. Tokyo, Japan, April 2000, p. 107.

225. D. Bryant, V. Berry, P. Kearney, T. Jiang, M. Li, T. Wareham, and H. Zhang, A practical algorithm for recovering the best supported edges of an evolutionary tree. *Proc. 11th Ann. ACM-SIAM Symp. on Discrete Algorithms* (SODA'2000), Jan. 2000, pp. 287–296.

226. K. Lanctot, M. Li, E.H. Yang, Estimating DNA sequence entropy. *Proc. 11th Ann. ACM-SIAM Symp. on Discrete Algorithms* (SODA'2000), Jan. 2000, pp. 409–418.

227. D. Bryant, J. Tsang, P. Kearney, and M. Li, Computing the quartet difference between two trees. *Proc. 11th Ann. ACM-SIAM Symp. on Discrete Algorithms* (SODA'2000), Jan. 2000, pp. 285–286.

228. M. Li, B. Ma, and L. Wang, Finding similar regions in many sequences. *Proc. 31st ACM Symp. Theory of Computing* (STOC'99), May 1-4, 1999, Atlanta. This paper solves several open questions on concensus problems. For example, it improves a $\frac{4}{3}$ approximation algorithm Hamming center problem to a polynomial approximation scheme (PTAS): find string $s$ which is Hamming distance $d$ away from a set of given strings. It also gives a PTAS for the well-known star-alignment problem when only constant number of gaps on each sequence is allowed.

229. T. Jiang, M. Li, P. Vitányi, Average-case complexity of Shellsort. *ICALP'99*, July 11-15, 1999, Prague.

230. H. Buhrman, T. Jiang, M. Li, and P. Vitányi, New applications of the incompressibility method. *ICALP'99*, July 11-15, 1999, Prague.

231. V. Berry, T. Jiang, P. Kearney, M. Li, T. Wareham, Quartet cleaning: improved algorithms and simulations. *ESA'99*, 1999, Prague.

232. T. Jiang, M. Li, P. Vitányi, The expected size of Heilbronn's triangles. *14th Annual IEEE Conference on Computational Complexity*, Atlanta, 1999.

233. K. Lanctot, M. Li, B. Ma, S. Wang, and L. Zhang, Distinguishing string search problems. *Proc. 10th Ann. ACM-SIAM Symp. on Discrete Algorithms* (SODA'99), Jan. 1999, pp. 633-642.

234. P. Kearney, M. Li, J. Tsang, and T. Jiang, Recovering branches on the tree of life: an approximation algorithm. *Proc. 10th Ann. ACM-SIAM Symp. on Discrete Algorithms* (SODA'99), Jan. 1999, pp. 537-546.

235. T. Jiang, P. Kearney, and M. Li, Orchestrating quartets: approximation and data correction. *Proc. 39th IEEE Symp. Foundation of Computer Science* (FOCS'98), Palo Alto, California, Nov 8-11, 1998, pp. 416-425.

236. P. Vitányi and M. Li, Minimum description length induction, Bayesianism and Kolmogorov complexity. *Proc. 1998 IEEE International Symp. Inform. Theory*, MIT, Cambridage, Aug. 16-21, 1998.

237. M. Li and L. Zhang, Better approximation of diagonal-flip transformation and rotation transformation. *Proc. 4th Int'l Computing and Combinatorics Conference* (COCOON'98), Aug. 12-14, 1998, Taipei, Taiwan, pp. 85-94.

238. B. Ma, L. Wang, M. Li, Fixed topology alignment with recombination. *9th Ann. Conf. on Combinatorial Pattern Matching*, (CPM'98). July 20-22, 1998.

239. B. Ma, M. Li and L. Zhang, On reconstructing species trees from gene trees in term of duplications and losses. *2nd Int'l Conf. Compututational Molecular Biology* (RECOMB'98), New York, pp. 182-191, 1998

240. M. Charikar, C. Chekuri, T.Y. Cheung, Z. Dai, A. Goel, S. Guha, and M. Li, Approximation algorithms for directed Steiner tree problems. *Proc. 9th Ann. ACM-SIAM Symp. on Discrete Algorithms* (SODA'98), 1998.

241. P. Vitányi and M. Li, On prediction by data compression. *Proc. 9th European Conference on Machine Learning*. Lecture Notes in AI, Vol. 1224, Springer-Verlag, 1997, pp. 14-30.

242. M. Li and P. Vitányi, Average-case analysis and Kolmogorov complexity. In *11th Symp. Fundamentals of Computation Theory* (FCT'97), Krakow, Poland, Sept. 1-3 1997.

243. H. Buhrman, M. Li, and P. Vitányi, Kolmogorov random graphs and the incompressibility method. *IEEE Proc. Conference on Compression and Complexity of Sequences*. 1997.

244. B. DasGupta, T. Jiang, S. Kannan, M. Li, and Z. Sweedyk, On the complexity and approximation of syntenic distance. In *Proc. 1st Symp. on Computational Molecular Biology* (RECOMB'97) Santa Fe, Jan. 20-23, 1997.

245. B. DasGupta, X. He, T. Jiang, M. Li, J. Tromp, and L. Zhang, On distances between phylogenetic trees. *Proc. 8th Ann. ACM-SIAM Symp. on Discrete Algorithms* (SODA'97), Jan. 5-7, 1997, New Orleans.

246. M. Li and P. Vitányi, On reversible computation. *Proc. 4th IEEE Physics of Computation Workshop*, Boston, Nov. 22-24, 1996.

247. M. Li and P. Vitányi, Ideal MDL and its relation to Bayesianism. ISIS'96, Australia, August 1996.

248. M. Li, J. Tromp, and L. Zhang, On the nearest neighbour interchange distance measure. *Proc. 2nd International Conf. on Computing and Combinatorics*, (COCOON'96), Hong Kong, June 17-19, 1996.

249. M. Li and P. Vitányi, Reversible simulation of irreversible computation. *Proc. 11th IEEE Conf. on Computational Complexity*, Philadelphia, May 24-27, 1996, pp. 301-306.

250. P. Vitányi and M. Li, Algorithmic arguments in physics of computation. *Proc. 4th Workshop on Algorithms and Data Structures* (WAD'95), Ottawa, August, 1995, Kingston, Canada. (Invited lecture of P. Vitányi)

251. J. Kececioglu, M. Li, and J. Tromp, Reconstructing a DNA sequence from erroneous copies. *Proc. 6th International Workshop on Algorithmic Learning Theory* (ALT'95), Fukuoka, Japan, Oct. 18-20, 1995, pp. 151-152. (Invited lecture of M. Li.)

252. T. Hancock, T. Jiang, M. Li, and J. Tromp, Lower bounds for learning decision lists and trees. *Proc. 12th Annual Symp. Theoret. Aspects of Comput. Sci.* (STACS'95). Munich, Germany, 1995, pp. 527-538.

253. T. Jiang and M. Li, On the approximation of shortest common supersequences and longest common subsequences. *Proc. 21st International Colloquium on Automata, Languages, and Programming* (ICALP'94), 1994, pp. 191-202.

254. M. Li and A. Viola, Learning secondary structures of proteins. *Proc. 6th International Conference on Computing and Information*, Peterborough, Ontario, May 26-28, 1994.

255. C.H. Bennett, P. Gács, M. Li, P. Vitányi, and W. Zurek, Thermodynamics of computation and information distance. *Proc. 25th ACM Symp. Theory of Computing* (STOC'93), 1993, pp. 21-30.

256. T. Jiang and M. Li, $k$ one-way heads cannot do string-matching. *Proc. 25th ACM Symp. Theory of Computing* (STOC'93), 1993, pp. 62-70. This paper settles a 13 year old well-known conjecture of Galil and Seiferas who showed that 6 two ways heads can do string-matching in linear time, and wondered if $k$ 1-ways heads can do the job.

257. T. Jiang and M. Li, Approximating shortest superstrings with constraints. *Proc. 3rd Workshop on Algorithms and Data Structures* (WAD'93), Montréal, August, 1993, pp. 385-396.

258. M. Li and P. Vitányi, Philosophical issues in Kolmogorov complexity. *Proc. 19th International Colloquium on Automata, Languages, and Programming* (ICALP'92), invited lecture of M. Li, 1992, pp. 1-15.

259. M. Li and P. Vitányi, Theories of learning. *Proc. 3rd International Conference for Young Computer Scientists* (ICYCS'93), invited lecture of M. Li, Beijing, July, 1993, pp. 10-20.

260. M. Li, Q. Gao and P. Vitányi, Recognizing on-line handwritten characters using MDL. *Proc. 1993 IEEE Information Theory Workshop*, Japan, 1993, pp. 24-25.

261. M. Li and P. Vitányi, Theory of thermodynamics of computation. *Proc. 2nd IEEE Physics of Computation Workshop*, Dallas, Texas, Oct. 4-6, 1992, pp. 42-46.

262. M. Li and P. Vitányi. Inductive reasoning. *Proc. DIMACS Workshop on Human Language*, E.S. Ristad, Editor, March 20-22, 1992, pp. 127-148. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Volume 17, American Math. Society, 1994.

263. A. Blum, T. Jiang, M. Li, J. Tromp, and M. Yannakakis, Linear approximation of shortest superstrings. *Proc. 23rd ACM Symp. on Theory of Computing* (STOC'91), 1991, pp. 328-336.

264. M. Li and P. Vitányi, Kolmogorov complexity and combinatorics. *Proc. 6th IEEE Structure in Complexity Theory Conference*, 1991, pp. 154-163.

265. T. Jiang and M. Li, On the complexity of learning strings and sequences. *Proc. 4th ACM Conf. Computational Learning Theory* (COLT'91), Santa Cruz. 1991, pp. 367-274.

266. D. Foulser, M. Li, and Q. Yang, A quantitative theory of plan merging. *Proc. 9th National Conference on Artificial Intelligence* (AAAI'91), Anaheim, CA, 1991, pp. 673-679.

267. M. Li, Towards a DNA sequencing theory (learning a string). *Proc. 31st IEEE Symp. Found. Computer Science* (FOCS'90), 1990, pp. 125-134. This paper and [121] have solved a major question in computer science open for 10 years on linear approximation of shortest common superstrings. This problem also arises in shotgun DNA sequence assembly and our solution has provided theoretical foundations for shotgun DNA sequencing. Our algorithm and proof have been described in all pioneering compuational biology textbooks including [Michael Waterman. *Introduction to Computational Biology: maps, sequences and genomes*, Chapman & Hall, 1995].

268. M. Li, Towards a DNA sequencing theory: learning a string. *Proc. 3rd ACM Workshop on Computational Learning Theory* (COLT'90), Rochester, NY, Aug. 6-8, 1990, page 394.

269. M. Li and P. Vitányi, Two applications of the universal distribution. *Proc. 1990 AAAI Spring Symposium Series: The Theory and Application of Minimum Length Coding*, Stanford, CA, 1990.

270. M. Li and P. Vitányi, A theory of learning simple concepts under simple distributions and average case complexity. *Proc. 30th IEEE Symp. Found. Computer Science* (FOCS'89), 1989, pp. 34-39.

271. M. Li and P. Vitányi, A theory of learning simple concepts under simple distributions. *Proc. 2nd ACM Workshop on Computational Learning Theory* (COLT'89), Santa Cruz, July 31 – August 2, 1989, p. 386.

272. M. Li and P. Vitányi, A new approach to formal language theory by Kolmogorov complexity. *Proc. 16th Int'l Colloquium on Automata, Languages and Programming* (ICALP'89), LNCS 372, Italy, 1989, pp. 506-520.

273. M. Li and P. Vitányi, How to share concurrent asynchronous wait-free variables. *Proc. 16th International Colloquium on Automata, Languages and Programming* (ICALP'89), LNCS 372, Italy, 1989, pp. 488-505.

274. Q. Gao and M. Li, An application of minimum description length principle to online recognition of hand-printed alphanumerals. *Proc. 11th International Joint Conference on Artificial Intelligence* (IJCAI'89), Vol. 1, Detroit, MI, 1989, pp. 843-848.

275. M. Li and P. Vitányi, Inductive reasoning and Kolmogorov complexity. *Proc. 4th IEEE Structure in Complexity Theory Conference*, 1989. pp. 165-185.

276. M. Kearns and M. Li, Learning in the presence of malicious errors. *Proc. 18th ACM Symp. on Theory of Computing* (STOC'88), 1988, pp. 267-280.

277. M. Li and P. Vitányi, Two decades of applied Kolmogorov complexity. *Proc. IEEE 3rd Structure in Complexity Theory Conference*, 1988, pp. 80-101.

278. M. Li and U. Vazirani, On the learnability of finite automata. *Proc. 1st ACM Workshop on Computational Learning Theory* (COLT'88), MIT, 1988, pp. 325-336.

279. M. Kearns, M. Li, L. Pitt, and L. Valiant, On the learnability of Boolean formulae. *Proc. 19th ACM Symp. Theory of Computing* (STOC'87), New York City, 1987, pp. 285-295.

280. A. Israeli and M. Li. Bounded time-stamps. *Proc. 28th IEEE Symposium of Foundation of Computer Science* (FOCS'87), 1987, pp. 371-382.

281. M. Kearns, M. Li, L. Pitt, and L. Valiant, Recent results on Boolean concept learning. *Proc. 4th Workshop on Machine Learning* (ML'87) (Ed. Pat Langley), Irvine, 1987, pp. 337-352.

282. B. Chor, A. Israeli, and M. Li, On processor coordination using asynchronous hardware. *Proc. 6th ACM Symp. on Principles of Distributed Computing* (PODC'87), 1987, pp. 86-97.

283. M. Li and Y. Yesha, Probabilistic and deterministic parallel complexity of symmetric functions. *Proc. 14th International Colloquium on Automata, Languages, and Programming* (ICALP'87), Lecture Notes in Computer Science, 267, 1987, pp. 326-335.

284. M. Li and Y. Yesha, New lower bounds for parallel computation. *Proc. 18th ACM Symp. Theory of Computing* (STOC'86), Berkeley, May, 1986, pp. 177-187.

285. M. Chrobak and M. Li, $k + 1$ heads are better than $k$ for PDA's. *Proc. 27th IEEE Symp. Found. Computer Science* (FOCS'86), Toronto, 1986, pp. 361-367. This paper settles a 20 year open question first raised by Harrison and Ibarra: there is a language accepted by a $k + 1$ head PDA but not acceptable by a $k$ head PDA.

286. J. Hartmanis, M. Li, and Y. Yesha, Containment, separation, complete sets, and immunity of complexity classes. *Proc. 13th International Colloquium on Automata, Languages, and Programming* (ICALP'86), LNCS 226, Rennes, France, 1986, pp. 136-145.

287. M. Li, L. Longpre, and P. Vitányi, The power of the queue. *Proc. IEEE 1st Structure in Complexity Theory Conference*, Berkeley, CA. June, 1986. In LNCS Vol 223, pp. 219-233.

288. M. Li, Simulating $2$ pushdown stores by $1$ tape in $O(n^{1.5})$ time. *Proc. 26th IEEE Symp. Found. Computer Science* (FOCS'85). Portland, pp. 56-64, 1985.

289. M. Li, Lower bounds by Kolmogorov complexity. *Proc. 12th International Colloquium on Automata, Languages, and Programming* (ICALP'85), LNCS 194, Nafplion, Greece, 1985, pp. 383-393.