Heng Li

Dana-Farber Cancer Institute & Harvard Medical School, Boston, MA 02215
Phone: 1-617-632-3032 E-Mail: lh3@me.com

Experience

Associate Professor, Dana-Farber Cancer Institute & Harvard Medical School, USA **May 2022 – present**
Assistant Professor **Oct 2018 – Apr 2022**

- Developed minimap2 for long-read mapping, miniprot for protein-to-genome alignment, hifiasm for accurate long read assembly and minigraph for pangenome representation. Developed hickit for single-cell Hi-C analysis.
- Key contributor to the NHGRI Human Pangenome Reference Project.

Affiliated Member, Broad Institute of MIT and Harvard, USA **Oct 2018 – present**
Senior Research Scientist **Oct 2009 – Sep 2018**

- Developed BWA-MEM for short- and long-read mapping, and fermi and miniasm for short- and long-read assembly. Established the foundation of multi-sample variant calling and inference models on low-coverage sequencing data. Developed software to analyze single-cell whole-genome sequencing data.
- Analyzed great ape, ancient DNA and large-scale population datasets and published the results.

Postdoctoral Fellow, Wellcome Trust Sanger Institute, UK **Sep 2006 – Sep 2009**

- Developed MAQ, BWA and samtools for short-read mapping and variant calling, which were the most widely used at the time. Led the design of the SAM/BAM format, the standard in sequencing data analysis.
- Formulated the PSMC model to infer population history; continued to develop novel algorithms to reconstruct gene trees with the prior knowledge on species tree.

Group Leader, Beijing Genomics Institute, China **Oct 2002 – Aug 2006**

- Developed software for gene finding, HLA typing from Sanger reads, sequence alignment and protein clustering. Constructed TreeFam, a database of gene trees, and devised novel algorithms for tree building.
- Analyzed the genomes or transcriptomes of rice, silkworm, chicken and mouse.

Research Assistant, Institute of Human Genetics, Aarhus University, Denmark **Mar 2003 – Mar 2004**

- Analyzed array-CGH data for breast cancer.

Education

PhD in Theoretical Physics, Chinese Academy of Science, China **Sep 2001 – Aug 2006**

Bachelor of Science in Physics, Nanjing University, China **Sep 1997 – Jul 2001**

Honors and Awards

- 2021 Sloan Research Fellow in computational and evolutionary biology
- Top 1% Highly Cited Researcher in Computer Science/Molecular Biology & Genetics in 2012–2021
- 2012 Benjamin Franklin Award for contributions in Open Source Bioinformatics (one per year)
- 2009-2010 AAAS Newcomb Cleveland Prize for the Most Outstanding Paper published in Science

Publications

Peer-reviewed Journal Articles

(Symbol * indicates co-first authorship; † indicates corresponding roles; double ** indicates mentees)

1. **Li H** (2023) Protein-to-genome alignment with miniprot. *Bioinformatics*, published online [PMID: 36648328]
2. Jarvis E. D, Formenti G, Rhie A, Guarracino A, Yang C, et al (2022) Semi-automated assembly of high-quality diploid human reference genomes. *Nature*, **611**:519-531. [PMID: 36261518]
3. Tan K.-T**, Slevin M. K, Meyerson M, **Li H**† (2022) Identifying and correcting repeat-calling errors in nanopore sequencing of telomeres. *Genome Biol.*, **23**:180. [PMID: 36028900]
4. Feng X**, Cheng H**, Portik D, **Li H**† (2022) Metagenome assembly of high-fidelity long reads with hifiasm-meta. *Nat Methods*, **19**:671-674. [PMID: 35534630]
5. Wang T, Antonacci-Fulton L, Howe K, Lawson H. A, Lucas J. K, et al (2022) The Human Pangenome Project: a global resource to map genomic diversity. *Nature*, **604**:437-446. [PMID: 35444317]
6. Cheng H**, Jarvis ED, Fedrigo O, Koepfli K-P, Urban L, Gemmell NJ, **Li H**† (2022) Haplotype-resolved assembly of diploid individuals without parental data. *Nat Biotechnol*, 40:1332-1335. [PMID: 35332338]
7. Kokot M, Gudyś A, **Li H**†, Deorowicz S† (2022) CoLoRd: compressing long reads. *Nat Methods.*, **19**:441-444. [PMID: 35347321]
8. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze A. V, et al (2022) The complete sequence of a human genome. *Science*, **376**:44-53. [PMID: 35357919]
9. Wagner J, Olson N. D, Harris L, McDaniel J, Cheng H**, et al (2022) Curated variation benchmarks for challenging medically relevant autosomal genes. *Nat Biotechnol.*, **40**:672-680. [PMID: 35132260]
10. Zhang HW**, Song L**, Wang X, Cheng H, Wang C, …, Liu XS†, **Li H**† (2021) Fast alignment and preprocessing of chromatin profiles with Chromap. *Nat Commun*, **12**:6566. [PMID: 34772935]
11. **Li H** (2021) New strategies to improve minimap2 alignment accuracy. *Bioinformatics*, **37**:4572-4. [PMID: 34623391]
12. Tan K.-T, Kim H, Carrot-Zhang J, Zhang Y, Kim W. J, et al (2021) Haplotype-resolved germline and somatic alterations in renal medullary carcinomas. *Genome Med.*, **13**:114. [PMID: 34261517]
13. Zhang HW**, Li H, Jain C, Cheng H, Au KF†, **Li H**†, Aluru S† (2021) Real-time mapping of nanopore raw signals. *Bioinformatics*, **37**:i477–i483 [PMID: 34252938]
14. Chu C, Borges-Monroy R, Viswanadham VV, Lee S, **Li H**, et al (2021) Comprehensive identification of transposable element insertions using multiple sequencing technologies. *Nat Commun*, **12**:3836 [PMID: 34158502]
15. **Li H**†, Rong J** (2021) Bedtk: finding interval overlap with implicit interval tree, *Bioinformatics*, **37**:1315–1316. [PMID: 32966548]
16. Feng X**, **Li H**† (2021) Higher rates of processed pseudogene acquisition in humans and three great apes revealed by long read assemblies. *Mol Biol Evol*, **38**:2958–2966 [PMID: 33681998].

17. Garg S**, Fungtammasan AA, Carroll A, Chou M, Schmitt A, …, Chin C-S†, Church G†, **Li H**† (2021) Chromosome-scale haplotype-resolved assembly of human genomes, *Nat Biotechnol*, **39**:309–312. [PMID: 33288905]

18. Xing D, Tan L, Chang CH, **Li H**†, Xie XS† (2021) Accurate SNV detection in single cells by transposon-based whole-genome amplification of complementary strands. *Proc Natl Acad Sci*, **118**:e2013106118 [PMID: 33593904].

19. Bonfield JK, Marshall J, Danecek P, **Li H**, Ohan V, *et al* (2021) HTSlib: C library for reading/writing high-throughput sequencing data. *Gigascience*, **10**:giab007 [PMID: 33594436].

20. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, …, **Li H** (2021) Twelve years of SAMtools and BCFtools. *Gigascience*, **10**:giab008 [PMID: 33590861].

21. Daher M, Basar R, Gokdemir E, Baran N, Uprety N, *et al* (2020) Targeting a cytokine checkpoint enhances the fitness of armored cord blood CAR-NK cells. *Blood*, **137**:624–636 [PMID: 32902645]

22. Cheng H**, Concepcion GT, Feng X**, Zhang H**, **Li H**† (2021) Haplotype-resolved de novo assembly with phased assembly graphs. *Nat Methods*, **18**:170–175 [PMID: 33526886].

23. **Li H**†, Feng X**, Chu C (2020) The design and construction of reference pangenome graphs with minigraph. *Genome Biology*, **21**:265. [PMID: 33066802]

24. Garg S, Aach J, **Li H**, Sebenius I, Durbin R, Church G (2020) A Haplotype-Aware De Novo Assembly of Related Individuals Using Pedigree Sequence Graph. *Bioinformatics*, **36**:2385–2392. [PMID: 31860070]

25. Gokhman D, Nissism-Rafinia M, Agranat-Tamir L, Housman G, García-Pérez, *et al* (2020) Differential DNA methylation of vocal and facial anatomy genes in modern humans. *Nat Commun*, **11**:1189. [PMID: 32132541]

26. Ruan J† and **Li H**† (2020) Fast and accurate long-read assembly with wtdbg2. *Nat Methods*, **17**:155–158. [PMID: 31819265]

27. **Li H** (2019) Identifying centromeric satellites with dna-brnn. *Bioinformatics*, **35**:4408–4410. [PMID: 30989183]

28. Wenger AM, Peluso P, Rowell WJ, Chang PC, Hall RJ, *et al* (2019) Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol*, **37**:1152-1162. [PMID: 31406327]

29. Vasimuddin M, Misra S, **Li H**, Aluru S. (2019) Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. in 2019 IEEE Interna- tional Parallel and Distributed Processing Symposium, IPDPS 2019, Rio de Janeiro, Brazil, May 20-24, 2019:314–324

30. Luo S, Yu JA, **Li H**, Song YS (2019) Worldwide genetic variation of the IGHV and TRBV immune receptor gene families in humans. *Life Sci Alliance*, **26**:2(2). [PMID: 30808649]

31. Regier AA, Farjoun Y, Larson DE, Krasheninina O, Kang HM, *et al* (2018) Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. *Nat Commun*, **9**:4038. [PMID: 30279509]

32. Tan L, Xing D, Chang CH, **Li H**, Xie XS (2018) Three-dimensional genome structures of single diploid human cells, *Science*, **361**:924–928. [PMID:30166492]

33. **Li H**†, Bloom JM, Farjoun Y, Fleharty M, Gauthier L, Neale B† and MacArthur D† (2018) A synthetic-diploid benchmark for accurate variant calling evaluation, *Nat Methods*, **15**:595–597. [PMID: 30013044]

34. **Li H** (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**:3094–3100. [PMID: 29750242]

35. Schneider V. A, Graves-Lindsay T, Howe K, Bouk N, Chen H.-C, et al. (2017) Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. Genome Res., 27:849-864 [PMID: 28396521]

36. Chen C*, Xing D*, Tan L*, **Li H***, Zhou G, Huang L, Xie XS (2017) Single-cell whole genome analyses by Linear Amplification via Transposon Insertion (LIANTI), *Science*, **356**:189-94. [PMID: 28408603]

37. Mallick S*, **Li H***, Lipson M*, Mathieson I*, Gymrek M, Racimo F, Zhao M, Chennagiri N, Nordenfelt S, Tandon A, et al. (2016) The Simons Genome Diversity Project: 300 genomes from 142 diverse populations, *Nature*, **538**:201–6. [PMID: 27654912]

38. **Li H** (2016) Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences, *Bioinformatics*, **32**:2103-10. [PMID: 27153593]

39. Mancuso N, Rohland N, Rand KA, Tandon A, Allen A, Quinque D, Mallick S, **Li H**, Stram A, Sheng X, *et al.* (2016) The contribution of rare variation to prostate cancer heritability, *Nat Genet.*, **48**:30-5 [PMID: 26569126]

40. **Li H** (2016) BGT: efficient and flexible genotype query across many samples, *Bioinformatics*, **32**:590-2. [PMID: 26500154].

41. The 1000 Genomes Project Consortium (2015) A global reference for human genetic variation, *Nature*, **526**:68-74. [PMID: 26432245]

42. **Li H** (2015) FermiKit: assembly-based variant calling for Illumina resequencing data, *Bioinformatics*, **31**:3694-6. [PMID: 26220959]

43. Palkopoulou E, Mallick S, Skoglund P, Enk J, Rohland N, **Li H**, Omrak A, Vartanyan S, Poinar H, Götherström A, *et al.* (2015) Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth, *Curr Biol.*, **25**:1395-400. [PMID: 25913407]

44. **Li H** (2015) BFC: correcting Illumina sequencing errors, *Bioinformatics*, **31**:2885-7 [PMID: 25953801]

45. Do R, Balick D, **Li H**, Adzhubei I, Sunyaev S, Reich D (2014) No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans, *Nat Genet.*, **47**:126-31. [PMID: 25581429]

46. Fu Q, **Li H**, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, Johnson PL, Aximu-Petri A, Prüfer K, de Filippo C, *et al.* (2014) Genome sequence of a 45,000-year-old modern human from western Siberia, *Nature*, **514**:445-9. [PMID: 25341783]

47. Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M, *et al.* (2014) Ancient human genomes suggest three ancestral populations for present-day Europeans, *Nature*, **513**:409-13. [PMID: 25230663]

48. **Li H** (2014) Fast construction of FM-index for long sequence reads, *Bioinformatics*, **30**:3274-5. [PMID: 25107872]

49. **Li H** (2014) Towards Better Understanding of Artifacts in Variant Calling from High-Coverage Samples, *Bioinformatics*, **30**:2843-2851 [PMID: 24974202]

50. Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C, *et al.* (2014) The complete genome sequence of a Neanderthal from the Altai Mountains, *Nature*, **505**:43-9. [PMID: 24352235]

51. Ruan J, Jiang L, Chong Z, Gong Q, **Li H**, Li C, Tao Y, Zheng C, Zhai W, Turissini D, *et al.* (2013) Pseudo-Sanger sequencing: massively parallel production of long and near error-free reads using NGS technology, *BMC Genomics*, **14**:711. [PMID: 24134808]

52. Genovese G, Handsaker RE, **Li H**, Kenny EE and McCarroll SA (2013) Mapping the human reference genome's missing sequence by three-way admixture in Latino genomes, *Am J Hum Genet.*, **93**:411-21. [PMID: 23932108]

53. Prado-Martinez J, Sudmant PH, Kidd JM, **Li H**, Kelley JL, Lorente-Galdos B, Veeramah B, Woerner A, O'Connor TD, Santpere G, *et al.* (2013) Great ape genetic diversity and population history, *Nature*, **499**:471-5 [PMID: 23823723].

54. Genovese G, Handsaker RE, **Li H**, Altemose N, Lindgren AM, Chambert K, Pasaniuc B, Price AL, Reich D, Morton CC *et al.* (2013) Using population admixture to help complete maps of the human genome, *Nat Genet*, **45**:406-14. [PMID: 23435088]

55. Sankararaman S, Patterson N, **Li H**, Pääbo S, Reich D (2013) The date of interbreeding between Neandertals and modern humans, *PLoS Genet*, **8**:e1002947. [PMID: 23055938]

56. Li S, Li R, **Li H**, Lu J, Li Y, Bolund L, Schierup MH, Wang J (2013) SOAPindel: efficient identification of indels from short paired reads, *Genome Res*, **23**:195-200 [PMID: 22972939]

57. 1000 Genomes Project Consortium (2012) An integrated map of genetic variation from 1,092 human genomes, *Nature*, **491**:56-65. [PMID: 23128226]

58. Meyer M, Kircher M, Gansauge MT, **Li H**, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K, de Filippo C (2012) A high-coverage genome sequence from an archaic Denisovan individual, *Science*, **338**:222-226. [PMID: 22936568]

59. Sun JX, Helgason A, Masson G, Ebenesersdóttir SS, **Li H**, Mallick S, Gnerre S, Patterson N, Kong A, Reich D, Stefansson K (2012) A direct characterization of human mutation based on microsatellites, *Nat Genet*, **44**:1161-5. [PMID: 22922873]

60. Pasaniuc B, Rohland N, McLaren PJ, Garimella K, Zaitlen N, **Li H**, Gupta N, Neale BM, Daly MJ, Sklar P, *et al.* (2012) Extremely low-coverage sequencing and imputation increases power for genome-wide association studies, *Nat Genet*, **44**:631-5. [PMID: 22610117]

61. **Li H.** (2012) Exploring single-sample SNP and INDEL calling with whole-genome de novo assembly, *Bioinformatics*, **28**:1838-44. [PMID: 22569178]

62. Hu X, Yuan J, Shi Y, Lu J, Liu B, Li Z, Chen Y, Mu D, Zhang H, Li N, *et al.* (2012) pIRS: Profile-based Illumina pair-end reads simulator, *Bioinformatics*, **28**:1533-5. [PMID: 22508794]

63. **Li H.** (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data, *Bioinformatics*, **27**:2987-93 [PMID: 21903627]

64. **Li H.** and Durbin R. (2011) Inference of Human Population History From Whole Genome Sequence of A Single Individual, *Nature*, **475**:493-6. [PMID: 21753753]

65. **Li H.** (2011) Improving SNP discovery by base alignment quality. *Bioinformatics*, **27**:1157-8. [PMID: 21320865]

66. **Li H.** (2011) Tabix: Fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics*, **27**:718-9. [PMID: 21208982]

67. Reich D., Green R.E., Kircher M., Krause J., Patterson N., Durand E.Y., Viola B., Briggs A.W., Stenzel U., Johnson P.L., *et al.* (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, **468**:1053-60. [PMID: 21179161]

68. 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**:1061-73. [PMID: 20981092]

69. **Li H.** and Homer N. (2010) A survey of sequence alignment algorithms for next-generation sequencing. *Brief Bioinform*, **11**:473-83. [PMID: 20460430]

70. Green R.E., Krause J., Briggs A.W., Maricic T., Stenzel U., Kircher M., Patterson N., **Li H.**, Zhai W., Fritz M.H. *et al.* (2010) A draft sequence of the Neandertal genome. *Science*, **328**:680-4. [PMID: 20448178]

71. **Li H.**, and Durbin R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler Transform. *Bioinformatics*, **26**:589-95. [PMID: 20080505]

72. Li R., Fan W., Tian G., Zhu H., He L., Cai J., Huang Q., Cai Q., Li B., Bai Y., Zhang Z. *et al.* (2009) The sequence and de novo assembly of the giant panda genome. *Nature*, **463**:311-7. [PMID: 20010809]

73. Schierup MH, Mailund T, **Li H**, Wang J, Tjønneland A, Vogel U, Bolund L, Nexø BA (2009) Haplotype frequencies in a sub-region of chromosome 19q13.3, related to risk and prognosis of cancer, differ dramatically between ethnic groups. *BMC Med Genet.*, **10**:20. [PMID: 19257887]

74. Sudbery I., Stalker J., Simpson J.T., Keane T., Rust A.G., Hurles M.E., Walter K., Lynch D., Teboul L., Brown S.D. *et al.* (2009) Deep short-read sequencing of chromosome 17 from the mouse strains A/J and CAST/Ei identifies significant germline variation and candidate genes that regulate liver triglyceride levels. *Genome Biol*, **10**:R112. [PMID: 19825173]

75. **Li H.***, Handsaker B.*, Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, **25**:2078-9. [PMID: 19505943]

76. Holt K.E., Teo Y.Y., **Li H.**, Nair S., Dougan G., Wain J. and Parkhill J. (2009) Detecting SNPs and estimating allele frequency in clonal bacterial populations by sequencing pooled DNA. *Bioinformatics*, **25**:2074-5. [PMID: 19497932]

77. **Li H.** and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, **25**:1754-60. [PMID: 19451168]

78. Vilella A.J., Severin J., Ureta-Vidal A., **Li H.**, Durbin R. and Birney E. (2009) EnsemblCompara GeneTrees: complete, duplication aware phylogenetic trees in vertebrates. *Genome Res*. **19**:327-35. [PMID: 19029536]

79. Li J., Gromov P., Gromova I., Moreira J.M., Timmermans-Wielenga V., Rank F., Wang K., Li S., **Li H.**, Wiuf C., Yang H., *et al.* (2008) Omics-based profiling of carcinoma of the breast and matched regional lymph node metastasis. *Proteomics*, **8**:5038-52. [PMID: 19003862]

80. Wang J.*, Wang W.*, Li R.*, Li Y.*, Tian G., Goodman L., Fan W., Zhang J., Li J., Zhang J., *et al*. (2008) The diploid genome sequence of an Asian individual. *Nature*, **456**:60-5. [PMID: 18987735]

81. Bentley D.R., Balasubramanian S., Swerdlow H.P., Smith G.P., Milton J., Brown C.G., Hall K.P., Evers D.J., Barnes C.L., *et al*. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, **456**:53-9. [PMID: 18987734]

82. **Li H.**, Ruan J. and Durbin R. (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.*, **18**:1851-8. [PMID: 18714091]

83. Down T.A.*, Rakyan V.K.*, Turner D.J., Flicek P., **Li H.**, Thorne N.P., Kulesha E., Gräf S., Tomazou E.M., Bäckdahl L., *et al.* (2008) A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylation analysis. *Nat. Biotech.*, **26**:779-85. [PMID: 18612301]

84. Campbell P.J.*, Stephens P.*, Pleasance E.D.*, O'Meara S., **Li H.**, Santarius T., Stebbings L.A., Leroy C., Edkins S., Hardy C., *et al.* (2008) Identification of somatically acquired rearrangement in cancer using genome-wide massively parallel paired-end sequencing. *Nat. Genet.*, **40**:722-9. [PMID: 18438408]

85. Ruan J.*, **Li H.***, Chen Z., Coghlan A., Coin L.J., Guo Y., Hériché J.K., Hu Y., Kristiansen K., Li R., *et al.* (2008) TreeFam: 2008 Update. *Nucleic Acids Res.*, **36**:D735-740. [PMID: 18056084]

86. **Li H.***, Guan L.*, Liu T.*, Guo Y.*, Zheng W., Wong G. and Wang J. (2007) A cross-species alignment tool (CAT). *BMC Bioinformatics*, **8**:439. [PMID: 17880681]

87. Gorodkin, J., Cirera, S., Hedegaard, J., Gilchrist, M. J., Panitz, F., Jorgensen, C., Scheibye-Knudsen, K., Arvin, T., Lumholdt, S., Sawera, M., Green, T., *et al.* (2007) Porcine transcriptome analysis based on 97 non-normalized cDNA libraries and assembly of 1,021,891 expressed sequence tags. *Genome Biol.*, **8**:R45. [PMID: 17407547]

88. Li S.*, Ma L.*, **Li H.***, Vang S., Hu Y., Bolund L. and Wang J. (2007) Snap: an integrated SNP annotation platform. *Nucleic Acids Res.*, **35** (Database issue):D707-10. [PMID: 17135198]

89. Ruan J.*, Guo Y.*, **Li H.***, Hu Y., Song F., Huang X., Kristiensen K., Bolund L. and Wang J. (2007) PigGIS: Pig Genomic Informatics System. *Nucleic Acids Res.*, **35** (Database issue):D654-7. [PMID: 17090590]

90. **Li H.***, Coghlan A.*, Ruan J.*, Coin L.J., Hériché J.K., Osmotherly L., Li R., Liu T., Zhang Z., Bolund L., *et al.* (2006) Treefam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res.*, **34** (Database issue):D572-80. [PMID: 16381935]

91. **Li H.***, Liu J.*, Xu Z*., Jin J., Fang L., Gao L., Li Y., Xing Z., Gao S., Liu T., *et al.* (2005) Test data sets and evaluation of gene prediction programs on the rice genome. *J Comput Sci & Technol.*, **20**:446-53.

92. Yu J.*, Wang J.*, Lin W.*, Li S.*, **Li, H.***, Zhou J.*, Ni P.*, Dong W., Hu S., Zeng C., *et al.* (2005) The Genomes of Oryza sativa: a history of duplications. *PLoS Biol.*, **3**. [PMID: 15685292]

93. Wang J., Zhang J., Zheng H., Li J., Liu D., **Li H.**, Samudrala R., Yu J. and Wong G. (2004) Mouse transcriptome: neutral evolution of 'non-coding' complementary DNAs. *Nature*, **431**. [PMID: 15495343]

94. Xia Q., Zhou Z., Lu C., Cheng D., Dai F., Li B., Zhao P., Zha X., Cheng T., Chai C., *et al.* (2004) A draft sequence for the genome of the domesticated silkworm (Bombyx mori). *Science*, **306**:1937-40. [PMID: 15591204]

95. Wong G., Liu B., Wang J., Zhang Y., Yang X., Zhang Z., Meng Q., Zhou J., Li D., Zhang J., *et al.* (2004) A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature*, **432**:717-22. [PMID: 15592405]

96. Li C., Ni P., Francki M., Hunter A., Zhang Y., Schibeci D., **Li H.**, Tarr A., Wang J., Cakir M., *et al.* (2004) Genes controlling seed dormancy and pre-harvest sprouting in a rice-wheat-barley comparison. *Funct Integr Genomics*, **4**:84-93. [PMID: 14770301]

PhD Thesis

Li H. (2006) Constructing the TreeFam database. PhD thesis, Institute of Theoretical Physics, Chinese Academy of Science.