

CURRICULUM VITAE

SCOTT J. EMRICH

CURRENT ADDRESS

Department of Computer Science and Engineering
University of Notre Dame
351 Fitzpatrick Hall, Notre Dame, IN 46556
Phone: (574) 631-0353; Email: semrich@nd.edu
Homepage: <http://www.nd.edu/~semrich>

EDUCATION

2007	Ph.D.	Bioinformatics and Computational Biology	Iowa State University
2002	B.S.	Biology and Computer Science	Loyola College in Maryland

ACADEMIC EXPERIENCE

Academic Appointments

University of Notre Dame

8/2007–present Assistant Professor, Dept. of Computer Science and Engineering

International

Indian Institute of Technology (IIT), Bombay

Fall 2006 Visiting Scholar, Kanwal Rekhi School of Information Technology

AFFILIATIONS

2008-present Eck Institute for Global Health, University of Notre Dame

HONORS AND AWARDS

2010	Invited and participated in 2010 Indo-American Frontiers of Engineering Symposium, National Academy of Engineering
2008	Iowa State University Zaffrano Prize for Graduate Research
2007	Iowa State University Research Excellence Award
2006	Electrical and Computer Engineering Research Excellence, Iowa State University
2006	<i>IEEE International Parallel and Distributed Processing Symposium (IPDPS)</i> best paper

PUBLICATIONS

Journal

- 2012 Moretti, C., Thrasher, A., Yu, L., Olson, M., **Emrich, S.**, and D. Thain. A framework for scalable genome assembly on clusters, clouds and grids. *IEEE Transactions on Parallel and Distributed Systems*, in press.
- 2012 O'Neil, S. T., and **S. J. Emrich** Haplotype and minimum-chimerism consensus determination using short sequence data, *BMC Genomics* (ICCABS 2010 special issue), in press.
- 2012 Megy, K.* , **Emrich, S. J.***, Lawson, D., Campbell, D., Dialynas, E., Hughes, D. S. T., Koscielny, G., Louis, C. , MacCallum, R. M., Redmond, S.N., Sheehan, A., Topalis, P., Wilson, D. and the VectorBase Consortium. VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics. *Nucleic Acids Research*, 40:D729-734.
- 2011 Rider, A. K., Siwo, G., Chawla, N. V., Ferdig, M. T., and **S. J. Emrich**. A Supervised Learning Approach to the Ensemble Clustering of Genes. *International Journal of Data Mining and Bioinformatics*,12:116.
- 2011 Carmichael, R., Braga-Henebry, P. Thain, D. and **S. J. Emrich**: Biocompute 2.0: an improved collaborative workspace for data intensive bio-science. *Concurrency and Computation: Practice and Experience* 23(17): 2305-2314.
- 2011 Samarakoon, U., Regier, A., Tan, A., Desany, B.A., Collins, B., Tan, J.C., **Emrich, S.J.** and M. T. Ferdig. High-throughput 454 sequencing for allele discovery and recombination mapping in *Plasmodium falciparum*. *BMC Genomics*, 12:116.
- 2010 Lawniczak, M.K.* , **Emrich, S.J.***, Holloway, A.K., Regier, A.P., Olson, M., White, B., Redmond, S., Fulton, L., Appelbaum, E., Godfrey, J., Farmer, C., Chinwalla, A., Yang, S.P., Minx ,P., Nelson, J., Kyung, K., Walenz, B.P., Garcia-Hernandez, E., Aguiar, M., Viswanathan, L.D., Rogers, Y.H., Strausberg, R.L., Sasaki, C.A., Lawson, D., Collins, F.H., Kafatos, F.C., Christophides, G.K., Clifton, S.W., Kirkness, E.F., and N. J. Besansky, Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science*, 330(6003): 512–514 (Google: 19 citations).
- 2010 Rider, A.K., Siwo, G., Chawla, N.V., Ferdig, M. and **S. J. Emrich**, A statistical approach to finding overlooked genetic associations. *BMC Bioinformatics*, 11:526.
- 2010 Lobo, N.F., Sangare, D. M., Regier, A. A., Reidenbach, K. R., Bretz, D. A., Sharakhova, M.V., **Emrich, S.J.**, Traore, S., F. Costantini, C., Besansky, N. J. and F. C. Collins, Breakpoint structure of the *Anopheles gambiae* 2Rb chromosomal inversion *Malaria Journal*, 9:293.
- 2010 Yu, L., Moretti, C., Thrasher,A. **Emrich, S.**, Judd, K., and D. Thain, Harnessing parallelism in multicore clusters with the All-Pairs, Wavefront, and Makeflow abstractions. *Cluster Computing*, 13:243.
- 2010 O'Neil, S.T., Dzurisin, J.D., Carmichael, R.D., Lobo, N.F., **Emrich, S.J.**, and J.J. Hellmann, Population-level transcriptome sequencing of nonmodel organisms *Erynnis propretius* and *Papilio zelicaon*. *BMC Genomics*, 11:310.
- 2009 Liu, S., Chen, H.D., Makarevitch, I., Shirmer, R., **Emrich, S.J.**, Dietrich, C.R., Barbazuk, W.B., Springer, N.M., and P.S. Schnable. High-throughput genetic mapping of mutants via Quantitative SNP-typing. *Genetics*, in press (Google: 17 citations).

- 2009 Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., et al. The B73 maize genome: complexity, diversity, and dynamics. *Science*, 326:1112-1115 (cover of Nov 20th, 2009 issue; Google: 372 citations).
- 2007 Ohtsu, K., Smith, M., **Emrich, S.J.**, Borsuk, L.A., Zhou, R., Chen, T., Zhang, X., Timmermans, M., Beck, J., Buckner, B., Janick-Buckner, D., Nettleton, D., Scanlon, M.J., and P.S. Schnable. Expression of retrotransposons in the shoot apical meristem of maize (*Zea mays* L.). *Plant Journal*, **52**:391-404 (Google: 61 citations).
- 2007 Barbazuk, W.B.*, **Emrich, S.J.***, Chen, H.D., Li, L., and P.S. Schnable. SNP discovery in maize via 454 transcriptome sequencing. *Plant Journal*, **51**:910-918 (Google: 165 citations).
- 2007 Kalyanaraman, A.*, **Emrich, S.J.***, Schnable, P.S., and S. Aluru. Assembling genomes on large-scale parallel computers. *Journal of Parallel and Distributed Computing*, **67**:1240-1255 (Google: 13 citations).
- 2007 **Emrich, S.J.***, Barbazuk, W.B*., Li, L. and P.S. Schnable. Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Research*, **17**:69-73 (Google: 166 citations).
- 2007 **Emrich, S.J.***, Li, L.*, Wen, T.-J., Yandea-Nelson, M.D., Fu, Y., Guo, L., Chou, H.-H., Aluru, S., Ashlock, D.A., and P.S. Schnable. Nearly identical paralogs (NIPs): implications for maize (*Zea mays* L.) genome evolution. *Genetics*, **175**:429-439 (Google: 40 citations; featured in *Science*, Vol. 315, No. 5810, pp. 302 in Editor's Choice: Highlights of recent literature).
- 2005 Fu, Y.*, **Emrich, S.J.***, Guo, L., Wen, T.-J., Aluru, S., Ashlock, D.A., and P.S. Schnable. Quality assessment of Maize Assembled Genomic Islands (MAGIs) and experimental validation of predicted novel genes, *Proceedings of the National Academy of Science, USA*, **102**:12282-12287 (Google: 59 citations).
- 2005 Yao, H., Guo, L., Fu, Y., Borsuk, L.A., Wen, T.-J., Skibbe, D.S., Cui, X., Scheffler, B.E., Cao, J., **Emrich, S.J.**, Ashlock, D.A., and P.S. Schnable. Evaluation of seven *ab initio* gene prediction programs for the discovery of maize genes, *Plant Molecular Biology*, **3**:445-460 (Google: 21 citations).
- 2004 **Emrich, S.J.**, Aluru, S., Fu, Y., Wen, T., Narayanan, M., Guo, L., Ashlock, D.A., and P.S. Schnable. A strategy for assembling the maize (*Zea mays* L.) genome, *Bioinformatics*, **20**:140-147. (ISI: 27 citations; Google: 39 citations)
- 2003 **Emrich, S.J.**, Lowe, M., and A.L. Delcher. PROBEmer: A web-based software tool for selecting optimal DNA oligos. *Nucleic Acids Res.*, **31**:3746-3750. (ISI: 33 citations; Google: 47 citations)
- 2002 Lowe, M., Madsen, E. L., Schindler, K., Smith, C., **Emrich, S.**, Robb, F., and R. U. Halden. Geochemistry and microbial diversity of a Trichloroethene-contaminated superfund site undergoing intrinsic in situ reductive dechlorination. *FEMS Microbiol. Ecol.*, **40**:123-134. (Google: 56 citations)

* denotes equal contribution

Peer-reviewed Conference

- 2012 Thrasher, A., Musgrave, Z., Thain, D. and **S. J. Emrich**. Shifting the Bioinformatics Computing Paradigm: A Case Study in Parallelizing Genome Annotation Using MAKER and Work Queue. *2nd IEEE International Conference on Computational Advances in Bio and medical Sciences*, Las Vegas, in press.

- 2012 Zhang, W., Zeng, E., Liu, D., Jones, S. and *S. J. Emrich* A Machine Learning Framework for Trait Based Genomics. *2nd IEEE International Conference on Computational Advances in Bio and medical Sciences*, Las Vegas, in press
- 2011 Theresa, C.B., A.N. Colaco, **S.J. Emrich**, S.T. O’Neil, and J.S. McLachlan. New genetic tools for estimating long-term changes in forest composition, 96th Annual ESA Conference.
- 2011 Yu, L., Carmichael, R., **S. J. Emrich**, and D. Thain. Managing worker pools for cloud workflows. *Proceedings of Cloud Computing and its Applications (CCA11)*; short abstract)
- 2011 Lanc, I., Bui, P., Thain, D. and **S. J. Emrich**. Adapting bioinformatics applications for heterogeneous systems: a case study. *Proceedings of the Second International Workshop on Emerging Computational Methods for the Life Sciences*.
- 2011 O’Neil, S.T. and **S. J. Emrich**. Robust haplotype reconstruction of eukaryotic sequence data with Hapler. *1st IEEE International Conference on Computational Advances in Bio and Medical Sciences*, Orlando. (Google: 3 citations)
- 2010 Rider, A., Siwo, G., **Emrich, S.J.**, Ferdig, M. and N. V. Chawla. A supervised learning approach to the unsupervised clustering of genes. *IEEE International Conference on Bioinformatics and Biomedicine*, Hong Kong.
- 2010 Zhang, W., Zeng, E., and **S.J. Emrich**. A two-stage machine learning approach for pathway analysis. *IEEE International Conference on Bioinformatics and Biomedicine*, Hong Kong.
- 2010 Thrasher, A., Carmichael, R., Bui, P., Thain, D., and **S.J. Emrich**. Taming complex bioinformatics workflows with Weaver, Makeflow, and Starch. *Workflows in Support of Large-Scale Science* held in conjunction with *Supercomputing 10*.
- 2010 Carmichael, R., Braga-Henebry, P., Thain, D. and **S.J. Emrich**. Biocompute: Towards a collaborative workspace for data intensive bio-science. *Emerging Computational Methods for the Life Sciences Workshop* held in conjunction with *19th International Symposium on High Performance Distributed Computing (HPDC)*.
- 2009 Regier, A., Cheng, C., Besansky, N. and **S.J. Emrich**. Improved inversion prediction in draft genomes. *Joint ISCB Africa ASBCB Conference on Bioinformatics of Infectious Diseases*, Bamako, Mali.
- 2009 Yu, L., Moretti, C., **Emrich, S.**, Judd, K. and D. Thain. Harnessing parallelism in multicore clusters with the All-Pairs and Wavefront abstractions. *18th International Symposium on High Performance Distributed Computing (HPDC)*.
- 2009 Regier, A., Olson, M. and **S.J. Emrich**. Alignment and analysis of closely related genomes, *International Conference on Bioinformatics and Computational Biology (Bi-CoB)*
- 2007 **Emrich, S.J.**, Kalyanaraman, A. and S. Aluru. Massively Parallel EST Clustering. *ISCA 20th International Conference on Parallel and Distributed Computing Systems (ISCA-PDCS’07)*.
- 2006 Kalyanaraman, A.*, **Emrich, S.J.***, Schnable, P.S., and S. Aluru. Assembling genomes on large-scale parallel computers. *IEEE International Parallel and Distributed Processing Symposium (IPDPS’06* ; Acceptance Ratio: 23%; best paper award).
- 2004 Ashlock, D.A., **Emrich, S.J.**, Bryden, K.M., Corns, S.A., T.-J. Wen, and P.S. Schnable. A comparison of evolved finite state classifiers and Interpolated Markov Models for improving PCR primer design, *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, 190–197 (Google: 4 citations).

Invited papers

- 2009 Moretti, C., Olson, M., **Emrich, S.**, and D. Thain. Highly scalable genome assembly on campus grids. 2nd Workshop on Many-Task Computing on Grids and Supercomputers, Portland, OR.
-

BOOK CHAPTERS

- 2011 Thain, D., Albrect, M., Bui, H., Bui, P., Carmichael, R., **Emrich, S.** and P. Flynn. Data Intensive Computing with Clustered Chirp Servers. *Data Intensive Distributed Computing: Challenges and Solutions for Large-scale Information Management*, IGI Global, Chapter 7.
- 2007 Barbazuk, W.B., **Emrich, S.J.** and P.S. Schnable. SNP Mining from Maize 454 EST Sequences. *Cold Spring Harbor Protocols*
- 2005 **Emrich, S.J.**, Kalyanaraman, A and S. Aluru. Algorithms for Large-Scale Sequence Clustering and Assembly of Biological Sequence Data. *Handbook of Computational Molecular Biology*, CRC Press, Chapter 13.
-

INVITED PRESENTATIONS

- 2/15/12 Challenges and opportunities of non-model informatics
CANGS bioinformatics conference, Las Vegas, NV
- 3/17/11 Ecoinformatics: Algorithms and challenges
Invited seminar, Michigan State University
- 1/27/11 Scalable computing for ecoinformatics
Invited bioinformatics seminar, Indiana University Bloomington
- 6/21/10 Challenges of modern bioinformatics
Invited Panelist, Emerging Computational Methods for the Life Sciences workshop, HPDC'10
- 6/11/10 Bioinformatics for arthropod gene discovery. 4th Annual Arthropod Genomics Symposium, Kansas City, MO.
- 10/14/09 Enabling, cost-effective bioinformatics analysis for arthropod genomics, 1st International Workshop on Information Systems for Insect Pests, Rennes, France, November 2009 (joint INIRA/INRA workshop).
- 3/2/09 Sequencing your favorite genome for \$1000
Invited Seminar, Biomedical Research Workshop
University of Notre Dame
- 2/20/09 Recent advances in *An. gambiae* genome analysis
Invited talk, African Virtual Conference on Bioinformatics (Afbix '09)
- 4/10/08 Lean on me: Comparative Anopheles genomics
Invited Seminar, Workshop on Interdisciplinary Biomedical Research
University of Notre Dame and Indiana University School of Medicine
- 3/26/07 Maize genome assembly on IBM BlueGene/L
Invited Seminar, Animal Breeding and Genetics
Iowa State University

- 10/18/06 Massively parallel plant genomics
Invited Seminar, Interdepartmental Plant Physiology
Iowa State University
- 1/12/04 A strategy for assembling the maize genome
Plant & Animal Genome XII Bioinformatics Workshop, San Diego CA
-

EXTERNAL SUPPORT

Current

F. Collins (PI): VectorBase: An NIAID Bioinformatics Resource Center for Invertebrate Vectors of Human Pathogens. National Institutes of Health. \$14,671,665 (5 years; Emrich scientific manager and co-PI, 50% effort)

P. Clark (PI): Synonymous Codon Usage as a Predictive Tool for Disease Susceptibility, \$9600. Indiana Clinical and Translational Sciences Institute.

COURSE EXPERIENCE

- CSE40532/60532 Bioinformatics Computing, Fall 2008 + 2010
– developed from scratch
- CSE20212 Fundamentals of Computing, Spring 2008–2012
– Outcomes based on Prof. Flynn’s previous version but lectures redone
- CSE40833/60833 Introduction to Parallel Algorithms and Programming, Fall 2007 + 2009
– Text based on comprehensive lecture notes by Prof. Aluru at ISU, otherwise redeveloped
-

CURRENT GRADUATE STUDENTS

- Shawn O’Neil PhD candidate; joined 2009 (**Schmidt Fellow, Global Health Fellow**)
- Andrew Rider PhD candidate; joined 2008 (**Global Health Fellow**; coadvised with Nitesh Chawla)
- Wei Zhang PhD in progress; joined 2010
- Irena Lanc PhD in progress; joined in 2009
- Andrew Thrasher PhD in progress; joined in 2009
- Lauren Assour PhD in progress; joined in 2010
(**Naughton Fellow**, coadvised by Ken Wolfe at Trinity College Dublin)
- Rory Carmichael MS in progress; joined 2011
-

UNDERGRADUATE RESEARCH ASSISTANTS

- Brian Kachmarck Summer 2010–present
- Paul Baranay Summer 2010–present

THESES SUPERVISED

- 2011 Allison A.P. Regier
"A flexible comparative genomics framework for integrating heterogeneous sequence data"
Doctoral Thesis, Computer Science and Engineering, University of Notre Dame
- 2010 Benjamin Drda
"Hive: an alignment tool for distributed systems"
Honors Thesis. Undergraduate Engineering Honors, University of Notre Dame
- 2009 Michael R. Olson
"New methods for assembly and validation of large genomes"
Master's Thesis, Computer Science and Engineering, University of Notre Dame
- 2008 Allison A.P. Regier
"Challenges in working with draft genomes,"
Master's Thesis, Computer Science and Engineering, University of Notre Dame

GRADUATE STUDENT ALUMNI

- Alison Regier 2007–2011 (**Presidential Fellow**); currently a research scientist at WashU (St. Louis)
- Michael Olson 2007–2009 ; currently software architect for CenterX

UNDERGRADUATE RESEARCH ALUMNI

- Pengqui Cheng (Tsinghua) 2011
- Sicong Zhang (Tsinghua) 2011
- Benjamin Drda 2007–2010
- Steven Bohlsen 2009 (Indiana University - South Bend)

UNIVERSITY SERVICE

University

- 2011–present Co-director, Genomics and Bioinformatics Core Facility (GBCF)
- 2008–2011 Director, Bioinformatics Core Facility (BCF)

Departmental Committees

- 2007–2010 Graduate Studies Committee
- 2008–present Curriculum Committee

Student Committees

- CSE: Scott Christley (2008 PhD), Chris Moretti (2010 PhD), Ryan Kennedy (2010 PhD), Darcy Davis (2011 PhD), Ryan Connaughton (2011 PhD), Li Yu, James Gentile

Biology: Jenica Abrudan, Changde Cheng, Masha Unger, Derrick Parker, Jihyun Won

Outside chair: 1 former student (AME), 4 collaborative (Biology), 4 university request

Service at Iowa State University

2006 BCB Graduate Curriculum Committee, which redesigned the PhD core curriculum in bioinformatics and computational biology.

PROFESSIONAL MEMBERSHIPS

International Society for Computational Biology (ISCB)

Association for Computing Machinery (ACM)

Institute of Electrical and Electronics Engineers (IEEE)

EXTERNAL SERVICE ACTIVITIES

Steering Committee Member

2010–2012 Arthropod Genomics Symposium

Program Chair

2012 Emerging Computational Methods for the Life Sciences workshop
ACM Symposium on High-Performance Parallel and Distributed Computing (HPDC)

2011 Parallel Bio-Computing Workshop
International Conference on Parallel Processing and Applied Mathematics

Program Committee Member

2011–2012 IEEE International Conference on Computational Advances in Bio and medical Sciences (ICCABS)

2010–2011 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)

2009–2012 International Conference on Bioinformatics and Computational Biology (BiCoB)

2009–2012 Using Emerging Parallel Architectures for Computation Science Workshop
International Conference on Computational Science

2011 ACM Conference on Bioinformatics, Computational Biology and Biomedicine

2011 ACM Symposium on High-Performance Parallel and Distributed Computing (HPDC)

2010 Workshop on Parallel Programming and Applications on Accelerator Clusters, IEEE Cluster

2007–2011 Parallel Bio-Computing Workshop
International Conference on Parallel Processing and Applied Mathematics

2009, 2011 International Conference on Contemporary Computing, Applications

2008–2009 IEEE/ACM Supercomputing (SC), Applications

2008 HICOMB Workshop, IEEE International Parallel and Distributed Processing Symposium (IPDPS)

2008 ACS/IEEE International Conference on Computer Systems and Applications (AICCSA)

Tutorials

2007 Computational Problems in Maize Genomics
International Conference on Computational Systems Bioinformatics (CSB)

2006 Exploring Computational Biology with a Massively Parallel
High Performance Computing Environment
International Conference on Intelligent Systems for Molecular Biology (ISMB)

Professional Committees

Education Committee
International Society for Computational Biology (ISCB)

Recent reviewer for the following journals:

Bioinformatics

Nucleic Acids Research

BMC Bioinformatics

BMC Genomics

Heredity