

# CURRICULUM VITAE

SCOTT J. EMRICH

---

## CURRENT ADDRESS

211B Cushing 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

2015–present	Director of Bioinformatics appointed by Notre Dame VP for Research
2015–present	Concurrent Associate Professor, Dept. of Biological Sciences
2014–present	Research Associate Professor, Dept. of Computer Science and Engineering
8/2007–2014	Assistant Professor, Dept. of Computer Science and Engineering

### Affiliations and Positions Held

University of Notre Dame

2011–present	Bioinformatics director, Notre Dame Genomics and Bioinformatics Core (GBC)
2011–present	Member, Environmental Change Initiative
2008–present	Member, Eck Institute for Global Health
2008–2011	Director, Notre Dame Bioinformatics Core

### International

Indian Institute of Technology (IIT), Bombay

Fall 2006	Visiting Scholar, Kanwal Rekhi School of Information Technology
-----------	---

---

## BRIEF SUMMARY OF RESEARCH IMPACT AND EXTERNAL FUNDING SUCCESS

- Over 75 peer-reviewed publications. Full report available via Google Scholar
- \$43.37 million in awards, of more than half of which was awarded after 2014. Two new NIH awards as PI (R21 and R01) in first half of 2016; co-PI and core lead of a new (2017) malaria genomics P01 led by Michael Ferdig (UND).
- Both of my students' 2016 CS conference papers invited and now accepted for journal publication.

Citation summary	All	Since 2012	Funding role	Last year (2016)	Details
Citations	5871	4505	PI	\$2,198,944 direct	\$643,541 to SJE
h index	27	22	co-PI	\$119,565 direct	shared postdoc to SJE
i10 index	43	42	Overall impact	largest award in COS in 2016 (VectorBase yr 3)	responsible for 5% of all 2016 awards to ND

---

## 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

---

## REFEREED PUBLICATIONS

Journal	<i>Graduate and Undergraduate students are underlined below, postdoctoral trainees in italics</i>
2017	Konar, A., <u>Choudhury, O.</u> , Bullis, R., Fiedler, L., Kruser, J.M., Stephens, M.T., Gailing, O., Schlarbaum, S., Coggeshall, M.V., Staton, M.E., Carlson, J.E., <b>Emrich, S.</b> and J. Romero-Severson. High-quality genetic mapping with ddRADseq in the non-model tree <i>Quercus rubra</i> , <i>BMC genomics</i> 18 (1), 417.
2017	<u>Choudhury, O.</u> , Chakrabarty, A., and <b>S.J. Emrich</b> . Highly Accurate and Efficient Data-Driven Methods For Genotype Imputation, <i>IEEE/ACM Transactions on Computational Biology and Bioinformatics</i> , in press.
2017	Chaney, J.L.*, <u>Steele, A.*</u> , Carmichael, R., Rodriguez, A., Specht, A.T., Ngo, K., Li, J, <b>Emrich, S.*</b> and P.L. Clark*. Widespread position-specific conservation of synonymous rare codons within coding sequences. <i>PLoS Computational Biology</i> , 13(5), e1005531.
2017	Adema, C., Hillier, L., Jones, C., Loker, E.S., Knight, M., Minx, P., Oliveira, G., Raghavan, N., Shedlock, A., Amaral, L. et al. Whole genome analysis of a schistosomiasis-transmitting freshwater snail. <i>Nature Communications</i> , 8, 15451 (Google: 5 citations).
2016	Siwo, G., <u>Rider, A.</u> , Tan, A., Pinapati, R., <b>Emrich, S.</b> , Chawla, N. and M. Ferdig. Prediction of fine-tuned promoter activity from DNA sequence. <i>F1000Res</i> , 5: 158. (Google: 3 citations)

- 2016 Love, R.R., Steele, A., Mamadou B., Coulibaly, M.B., Traore, S., **Emrich, S.J.**, Fontaine, M.C. and N. J. Besansky. Chromosomal inversions and ecotypic differentiation in *Anopheles gambiae*: the perspective from whole-genome sequencing. *Molecular Ecology*, 25(23), 5889–5906 (Google: 2 citations).
- 2016 Behura, S.K., *Sarro, J.*, Li, P., Mysore, K., Severson, D.W., **Emrich, S.J.\*** and M. Duman-Scheel\*. High-throughput cis-regulatory element discovery in the vector mosquito *Aedes aegypti*. *BMC Genomics*, 17 (1), 341 (Google: 2 citations).
- 2016 Hall, A.B.\*, Papathanos, P.-A.\*, Sharma, A.\*, Cheng, C.\*, Akbari, O.S., Assour, L., Bergman, N.H., Cagnetti, A. Crisanti, A. Dottorini, T., Fiorentini, E., Galizi, R., Hnath, J., Jiang, X., Koren, S., Nolan, T., Radune, D., Sharakhova, M.V., Steele, A., Timoshvskiy, V., Windbichler, N., Zhang, S., Hahn, M.W., Phillippy, A.M., **Emrich, S.J.**, Sharakhov, I.V., Tu, Z.J., and N.J. Besansky. Radical remodeling of the Y chromosome in a recent radiation of malaria mosquitoes. *Proceedings of the National Academy of Sciences*, 113(15):E2114-E2123 (Google: 17 citations).
- 2015 Mesquita, R.D. *et al.* (120 authors total). The genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite transmission. *Proceedings of the National Academy of Science*, 112 (48), 14936-14941 (Google: 64 citations).
- 2015 Choudhury, O., Hazekamp, N., Thain, D. and **S. Emrich**. Accelerating comparative genomics workflows in a distributed environment with optimized data partitioning and workflow fusion. *Scalable Computing: Practice and Experience*, 16 (1), 53-70 (Google: 5 citations).
- 2015 Ramalho-Ortigao, M., Coutinho-Abreu, I.V., Balbino, V.Q., Figueiredo, C.A.S., Mukbel, R., Dayem, H., Hanafi, H.A., El-Hossary, S.S., Fawaz, E., Abo-Shehada, M., Hoel, D.F., Stayback, G., Wadsworth, M., Shoue, D.A., Abrudan, J., Lobo, N.F., Mahon, A.R., **Emrich, S. J.**, Kamhawi, S., Collins, F.H. and M. A. McDowell. *Phlebotomus papatasi* SP15: mRNA expression variability and amino acid sequence polymorphisms of field populations. *Parasites and Vectors*, 8(1):298 (Google: 4 citations).
- 2015 Egan, S.P.\*, *Ragland, G.\**, Assour, L., Powell, T.H.Q., Hood, G.R., **Emrich, S.**, Nosil, P. and J. L. Feder. Experimental evidence of genome-wide impact of ecological selection during early stages of speciation-with-gene-flow. *Ecology Letters*, 18(8):817–825 (Google: 34 citations).
- 2015 Mbengue, A.\*, Bhattacharjee\*, S., Pandharkar, T. , Haining, L., Estiu, G., Stahelin, R.V., Rizk, S., Njimoh, D.L., Ryan, Y., Chotivanich, K., Nguon, C., Ghorbal, M., Lopez-Rubio, J.J., Pfrender, M., **Emrich, S.**, Mohandas, N., Dondorp, A.M., Wiest, O. and K. Haldar. A molecular mechanism of artemisinin resistance in *Plasmodium falciparum* malaria. *Nature*, 520(7549):683–687 (Google: 153 citations).
- 2015 Warren, A.S., Aurrecochea, C., Brunk, B. Desai, P., **Emrich, S.**, *Giraldo-Calderon, G.*, Harb, O., Hix, D., Lawson, D., Machi, D., Mao, C., McClelland, M., Nordbergm E., Shukla, M., Wattam, A.R., Will, R., Yo, H.S. and B. Sobral. RNA-Rocket: An RNaseq analysis tool for infectious disease research. *Bioinformatics*, btv002 (Google: 7 citations).
- 2015 *Fontaine, M.C.\**, Pease, J.B.\*, Steele, A., Waterhouse, R.M., Neafsey, D.E., Sharakhovm I.V., Jiang, X., Hall, A.B., Catteruccia, F., Kakani, E., Mitchell, S.N., Wu, Y.-C., Smith, H.A., Love, R.R., Lawniczak, M.K., Slotman, M.A., **Emrich, S.J.**, Hahn, M.W. and N.J. Besansky. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*, 347(6217):1258524 (Google: 138 citations)
- 2015 The Anopheles Genomes Cluster Consortium. Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. *Science*, 347(6217):1258522 (Google: 161 citations).

- 2014 *Giraldo-Calderon, G.\**, **Emrich, S.\***, MacCallum, B., Maslen, G., Gesing, S., Dialynas, E., Topalis, P., Ho, N., Madey, G., Colins, F., Lawson, D. and the VectorBase Consortium. VectorBase: an updated bioinformatics resource for invertebrate vectors of human pathogens *Nucleic Acids Research*, gku1117 (Google: 92 citations).
- 2014 Tomchaney, M., Mysore, K., Sun, L., Li, P., **Emrich, S.J.**, Severson, D.W. and M. Duman-Scheel. Examination of the genetic basis for sexual dimorphism in the *Aedes aegypti* (dengue vector mosquito) pupal brain. *Biology of Sex Differences*, 5:10 (Google: 4 citations).
- 2014 Rider, A.K., Milenkovic, T., Siwo, G.H., Pinpati, R.S., **Emrich, S.J.**, Ferdig, M.T. and N.V. Chawla. Networks' characteristics are important for systems biology. *Network Science*, 2(2):139–161 (Google: 3 citations).
- 2014 Jiang, X., Peery, A., Hall, A.B., Sharma, A., Chen, X.-G. Waterhouse, R.M., Komissarov, A., Riehle, M.M. Shouche, Y., Sharakhova, M.V. *et al.*. Genome analysis of a major urban malaria vector mosquito, *Anopheles stephensi*. *Genome Biology*, 15(9):459 (Google: 47 citations).
- 2014 Dugan, V.G., **Emrich, S.J.**, *Giraldo-Calderon, G.I.*, Harb, O.S., Newman, R.M., Pickett, B.E., Schriml, L.M., Stockwell, T.B., Stoeckert, C.J. Sullivan, D.E., *et al.*, Standardized metadata for human pathogen/vector genomic sequences. *PLoS One*, 9(6):e99979 (Google: 18 citations).
- 2014 O'Neil, S.T., Dzurisin, J.D.K., Williams, C.M. , Lobo, N.F., Higgins, J.K., Deines, J.M., Carmichael, R.D., *Zeng, E.*, Tan, J.C., Wu, G.C. **Emrich, S.J.** and J.J. Hellmann. Gene expression in closely related species mirrors local adaptation: consequences for responses to a warming world. *Molecular Ecology*, 23(11):2686–2698 (Google : 13 citations).
- 2014 Thrasher, A., Musgrave, Z., Kachmark, B., Thain, D. and **S. J. Emrich**. Shifting the bioinformatics computing paradigm: A case study in parallelizing genome annotation using MAKER and Work Queue. *Int. J. Bioinformatics Research and Applications*, 10(4)447–460 (Google: 1 citation).
- 2014 Geng, P., Li, W., Lin, L., de Miranda, J.R., **Emrich, S.**, An, L. and O. Terenius. Genetic characterization of a novel Iffavirus associated with vomiting disease in the Chinese oak silkworm *Antheraea pernyi*. *PLoS One*, 9(3):e92107 (Google : 11 citations).
- 2014 *Zhang, W.*, *Zeng, E.*, *Livermore, J.*, Liu, D., Jones, S. and **S. Emrich**. Mapping genomic features to functional traits through microbial whole genome sequences. *Int. J. Bioinformatics Research and Applications*, 10(4):461–478 (Google: 5 citations).
- 2013 Meyer, P., Siwo, G., Zeevi, D., Sharon, E., Norel, R., Segal, E., Stolovitzky, G., *et al.* Inferring gene expression from ribosomal promoter sequences, a crowdsourcing approach. *Genome Research*, 23(11), 1928–1937 (Google: 4 citations).
- 2013 Evans, J.D., Brown, S.J., Hackett, K.J., Robinson, G., Richards, S., Lawson, D., Elsik, C., Coddington, J., Edwards, O., **Emrich, S.**, Gabaldon, T., Goldsmith, M., Hanes, G., Misof, B., Munoz-Torres, M., Niehuis, O., Papanicolaou, A., Pfrender, M., Poelchau, M., Purcell-Miramontes, M., Robertson, H.M., Ryder, O., Tagu, D., Torres, T., Zdobnov, E., Zhang, G. and X. Zhou. The i5K Initiative: Advancing Arthropod Genomics for Knowledge, Human Health, Agriculture, and the Environment. *Journal of Heredity*, 5: 595–600 (Google: 105 citations).
- 2013 O'Neil, S.T. and **S. J. Emrich**. Assessing *de novo* transcriptome assembly metrics for consistency and utility. *BMC Genomics*, 14(1):465 (Google: 68 citations).

- 2013 Neafsey, D.E., Christophides, G.K., Collins, F.H., **Emrich, S.J.**, Fontaine, M.C., Gelbart, W., Hahn, M.W., Howell, P.I., Kafatos, F.C., Lawson, D., Muskavitch, M.A.T., Waterhouse, R.M., Williams, L.J. and N. J. Besansky. The evolution of the *Anopheles* 16 genomes project. *G3*, 3:(7), 1191–1194 (Google: 48 citations).
- 2013 Livermore, J.A., **Emrich, S.J.**, Tan, J. and S. E. Jones. Freshwater bacterial lifestyles inferred from comparative genomics. *Environmental Microbiology*, 16(3):746–758 (Google: 15 citations).
- 2013 Abrudan, J., Ramalho-Ortigao, M., O’Neil, S., Stayback, G., Wadsworth, M., Bernard, M., Shoue,D., **Emrich, S.**, Lawyer, P., Kamhawi, S., Rowton, E.D., Lehane, M.J., Bates, P.A., Valenzeula, J.G., Tomlinson, C., Appelbaum, E., Moeller, D., Thiesing, B., Dillon, R., Clifton, S., Lobo, N.F., Wilson, R.K., Collins, F.H. and M.A. McDowell. The characterization of the *Phlebotomus papatasi* transcriptome. *Insect Molecular Biology*,22(2):211–232 (Google: 11 citations).
- 2013 Bradnam, K.R., Fass, J.N., Alexandrov, A., Baranay, P., Bechner, M., Birol, I., Boisvert, J., Chapman, J.A., Chapuis, G., Chikhi, R. *et al.* Assemblathon 2: evaluating *de novo* methods of genome assembly in three vertebrate species, *GigaScience*, 2(1):1–31 (Google: 316 citations).
- 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*, 23(12):2189–2197 (Google: 22 citations).
- 2012 O’Neil, S. T. and **S. J. Emrich**. Haplotype and minimum-chimerism consensus determination using short sequence data. *BMC Genomics* 13(Suppl 2), S4 (Google: 13 citations).
- 2012 Megy, K.\*, **Emrich, S. J.\***, Lawson, D., Campbell, D., Dialynas, E., Hughes, D.S., 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 (Google: 148 citations).
- 2012 Lanc, I., Bui, P., Thain, D. and **S.J. Emrich**. Adapting bioinformatics applications for heterogeneous systems: a case study. *Concurrency and Computation: Practice and Experience* (Google: 13 citations).
- 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 (Google: 6 citations).
- 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 (Google: 11 citations).
- 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 (Google: 25 citations).
- 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: 207 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 (Google: 4 citations).
- 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 (Google: 19 citations).
- 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*, 13:243–246 (Google: 58 citations).
- 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 propertius* and *Papilio zelicaon*. *BMC Genomics*, 11:310 (Google: 113 citations).
- 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*, 184(1):19–26 (Google: 64 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: 2347 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. Global gene expression analysis of the shoot apical meristem of maize (*Zea mays* L.). *Plant Journal*, **52**:391–404 (Google: 140 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: 418 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: 37 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: 332 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: 46 citations; featured in *Science*, Vol. 315, No. 5810, pp. 302 in Editor’s Choice: Highlights of recent literature). (Google: 64 citations)
- 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: 79 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: 45 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. (Google: 65 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. (Google: 74 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: 57 citations)

\* denotes equal contribution

### Peer-reviewed Conference

- 2016 Zhu, S., Chen, D. and **S.J. Emrich**. Single molecule sequencing-guided scaffolding and correction of draft assemblies. *6th IEEE International Conference on Computational Advances in Bio and Medical Sciences*, Atlanta.
- 2016 Choudhury, O.C., Chakrabarty, A. and **S.J. Emrich**. HAPI-Gen: Highly accurate phasing and imputation of genotype data. *ACM International Conference on Bioinformatics and Computational Biology (ACM-BCB)*, 78–87.
- 2015 Zeng, E., Zhang, W., **Emrich, S.**, Liu, D., Livermore, J. and S. Jones. A computational framework for integrative analysis of large microbial genomics data. *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*.
- 2015 Hazekamp, N., Sarro, J., Choudhury, O., Gesing, S., **Emrich, S.**, and D. Thain. Scaling up bioinformatics workflows with dynamic job expansion: a case study using Galaxy and Makeflow. *IEEE 11th International Conference on e-Science (e-Science)* (Google: 1 citation).
- 2015 Assour, L.A. and **S.J. Emrich**. Global Maximum-parsimony based ancestral reconstruction with non-universal genes. *5th IEEE International Conference on Computational Advances in Bio and Medical Sciences*, Miami.
- 2015 Choudhury, O., Rajan, D., Hazekamp, N., Gesing, S., Thain, D. and **S. Emrich**. Balancing thread-level and task-level parallelism for data-intensive workloads on clusters and clouds. *Proceedings of IEEE Cluster*, 2015 (Google: 1 citation).
- 2015 Assour, L. and **S.J. Emrich**. Multi-genome synteny for assembly improvement. *Proceedings of the 7th International Conference on Bioinformatics and Computational Biology (BICoB)* (Google: 2 citations).
- 2015 Steele, A. and **S.J. Emrich**. pbSandwich: Scaffolding draft genomes with long reads. *Proceedings of the 7th International Conference on Bioinformatics and Computational Biology (BICoB)*.
- 2014 Choudhury, O., Hazekamp, N.L., Thain, D. and **S.J. Emrich**. Accelerating comparative genomics workflows in a distributed environment with optimized data partitioning. *Proceedings of C4Bio (part of 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid))* (Google: 4 citations)
- 2014 Hazekamp, N.L., Choudhury, O., Gesing, S., **Emrich, S.** and D. Thain. Expansion of logical workflows into series of independent encapsulated workflows. *Proceedings of the 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)*.

- 2013 Lanc, I. and **S.J. Emrich**. An unsupervised learning approach to assembly validation. *3rd IEEE International Conference on Computational Advances in Bio and medical Sciences*, New Orleans (Google: 1 citation).
- 2013 Lumibao, C.Y., McLachlan, J., Kuch, M., **Emrich, S.J.**, Jackson, S.T. and H. Poinar. Unlocking the Past: Genetic Clues of Historical Shifts in Forest Community Composition from Ancient DNA in Lake Sediments. *Proceedings of the 98th Annual Meeting of the Ecological Society of America (ESA)*, Minneapolis, MN.
- 2013 Zhang, W., Zeng, E., Livermore, J., Liu, D., Jones, S. and **S. Emrich**. Predicting bacterial functional traits from whole genome sequences Using random forests. *3rd IEEE International Conference on Computational Advances in Bio and medical Sciences*, New Orleans (Google: 1 citation)
- 2013 Rajan, D., Thrasher, A., Abdul-Wahid. B., Izaguirre, J.A., **Emrich, S.** and D. Thain. Case studies in designing elastic applications. *13th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)* (Google: 1 citation)
- 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. (Google: 6 citations)
- 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 (Google: 5 citations).
- 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 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*. (Google: 6 citations)
- 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: 9 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. (Google: 2 citations)
- 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. (Google: 5 citations)
- 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*. (Google: 18 citations)
- 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)*. (Google: 9 citations)
- 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)*. (Google: 54 citations)



- 2009 Regier, A., Olson, M. and **S.J. Emrich**. Alignment and analysis of closely related genomes, *International Conference on Bioinformatics and Computational Biology (BiCoB)* (Google: 1 citation)
- 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). (Google: 33 citations).
- 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: 8 citations).

#### Invited papers

- 2012 **S.J. Emrich**. Opportunities and challenges of non-model ecoinformatics. *2nd IEEE International Conference on Computational Advances in Bio and medical Sciences*, Las Vegas
- 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. (Google: 19 citations)

---

#### BOOK CHAPTERS

- 2013 Rider, A.K, Chawla, N.V. and **S.J. Emrich**. A survey of current integrative network algorithms for systems biology. Ales Prokop and Bela Csukas (Eds.) Springer, *Systems Biology*, 479–495. (Google: 3 citations)
- 2012 Thain, D., Albrecht, 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. (Google: 3 citations)
- 2007 Barbazuk, W.B., **Emrich, S.J** and P.S. Schnable. SNP mining from maize 454 EST sequences. *Cold Spring Harbor Protocols* (Google: 15 citations)
- 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 (Google: 8 citations).

---

#### INVITED PRESENTATIONS OVER THE LAST FIVE YEARS

- 9/29/16 Single molecule sequencing to improve genomes  
Invited talk, International Congress of Entomology (ICE)
- 4/5/16 VectorBase, your improved arthropod bioinformatics resource  
Invited talk, Entomological Society of America (ESA) Pacific branch, Honolulu, HI
- 2/20/16 Genome assembly improvement as applied to mosquitoes  
Invited seminar, Virginia Tech (Entomology)
- 6/17/15 Assembly assesement and improvement in the Anophelines  
Invited talk, 9th Annual Arthropod Genomics Symposium, Manhattan, KS.

4/28/15	Non-model genome assembly improvement Invited seminar, University of Georgia (Institute for Bioinformatics)
8/23/12	Challenges and opportunities of non-model ecoinformatics <b>Keynote: Bioinformatics and the next-gen revolution symposium</b> International Congress of Entomology (ICE2012)
2/15/12	Challenges and opportunities of non-model informatics CANGS bioinformatics conference, Las Vegas, NV

## EXTERNAL SUPPORT

### Current

S. Emrich (PI). Improving the *Culex quinquefasciatus* genome for improved population-level analysis. National Institutes of Health, \$391,312, 3/1/16–2/28/18

S. Emrich (PI w/ Milenkovic, Clark and Li). Integrative Computational Framework for Pattern Mining in Big -omics Data: Linking Synonymous Codon Usage to Biogenesis. National Institutes of Health, \$1,094,400, 6/1/16–5/31/21.

S. Emrich and F. Collins (PIs): VectorBase: A Bioinformatics Resource Center for invertebrate vectors of human pathogens. National Institutes of Health, \$14,714,077 with all options, 9/15/14–09/14/19.

### Completed

N. Besansky (PI) and S. Emrich. Anchoring and uniting the *An. funestus* assembly for improved vector analysis. National Institutes of Health, \$427,052, 07/01/14–06/30/17.

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 and S. Emrich (PIs): Influence of Translation on Protein Folding. National Institutes of Health collaborative supplement to 5R01GM074807-07: \$290,704 (2 years)

## RECENT INTERNAL COMPETITIVE FUNDING (1/2013 – PRESENT)

S. Emrich (PI) and M. Pfrender, J. Nabrzyski (co-PIs): Replacement of the Notre Dame Bioinformatics data analysis rig, \$98,000 (inc. cost share from the Eck Institute for Global Health and ND Environmental Change Initiative)

N. Besansky and S. Emrich (PIs): Building multi-disciplinary teams for Global Health research and training: “Decoding the Y chromosome of a killer: the African malaria mosquito,” \$24,12, Eck Institute for Global Health.

J. McLachlan and S. Emrich (PIs): “Using ancient DNA to resolve longstanding questions in global change ecology,” \$63,025, ND Environmental Change Initiative

## COURSE EXPERIENCE AT NOTRE DAME

CSE40532/60532 Bioinformatics Computing, Fall 2008, 2010, 2012–2013, 2015, 2017  
– Developed from scratch. One of the most multidisciplinary classes in CSE with 1/3 Biology, 1/3 CSE grad and 1/3 CSE undergrad.

- CSE20312      Data Structures, Spring 2016  
– C++ Object-oriented programming look at data structures with a capstone team project
- CSE20212      Fundamentals of Computing, 2008–2015  
– C++ Object-oriented programming with a capstone team project
- CSE66560      Advanced Bioinformatics Computing, Spring 2012  
– Active learning for bioinformatics algorithms and implementations at a graduate-only level. Developed from scratch.
- BIOS/CSE60132    Basic Computing for Bioinformatics, 2011–2016  
– Introduction to bioinformatics tailored for non-Computer Science majors built based on my and my students’ collaborations by Shawn O’Neil (now at Oregon State doing similar training). Was the highest rated fall CSE class, now taught by BIOS (Stuart Jones)
- 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

- Shenglong Zhu    MS/PhD; joined in 2014 (co-advised by Danny Chen)
- Gabriel Wright   MS/PhD; joined in 2017
- David Molik      MS/PhD; joined in 2017
- 

#### OTHER MEMBERS OF RESEARCH GROUP

- Joseph Sarro     Senior analyst, Notre Dame Genomics and Bioinformatics Core
- Robert Wieck     Lead Programmer, VectorBase
- Antelmo Aguilar   Programmer, VectorBase
- DK Kwon          Programmer, VectorBase
- 

#### UNDERGRADUATE RESEARCH ASSISTANTS

- Mara Staines     2016–present (**Engineering honors student**)
- Will Markley     2015–present
- 

#### THESES SUPERVISED

- 2017              Olivia Choudhury  
“Expediting analysis and improving fidelity of ‘Big Data’ genomics”  
Doctoral Thesis, Computer Science and Engineering, University of Notre Dame
- 2016              RJ Nowling  
“Replacing domain-specific methods in bioinformatics with machine learning techniques”  
Doctoral Thesis, Computer Science and Engineering, University of Notre Dame

2015	Lauren Assour "Species divergence: Computational methods to explore and resolve biological questions related to the evolution of species" Doctoral Thesis, Computer Science and Engineering, University of Notre Dame
2014	Irena Lanc "Novel methods for assembly validation and improvement" Doctoral Thesis, Computer Science and Engineering, University of Notre Dame
2013	Andrew Thrasher "Leveraging large-scale distributed systems for massive biological datasets" Doctoral Thesis, Computer Science and Engineering, University of Notre Dame
2013	Wei Zhang "Data mining for biological data learning: Algorithms and application" Doctoral Thesis, Computer Science and Engineering, University of Notre Dame
2013	Andrew Rider (co-advised with Nitesh Chawla) "Data and network science for noisy heterogeneous systems" Doctoral Thesis, Computer Science and Engineering, University of Notre Dame
2013	Rory Carmichael "Scaling collaborative bioinformatics" Masters Thesis, Computer Science and Engineering, University of Notre Dame
2012	Shawn O'Neil "Non-model transcriptomics: Applications, assessments, and algorithms" Doctoral Thesis, Computer Science and Engineering, University of Notre Dame
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

---

#### ALUMNI (GRADUATE AND POSTDOCTORAL\*)

Olivia Choudhury	2012 – 2017 ( <b>Eck Global Health fellow</b> )
Lauren Assour	2010 – 2015 ( <b>Naughton Fellow</b> , coadvised by Ken Wolfe at Trinity College Dublin)
Colin Teberg	2014–2015 – currently a software engineer in Chicago
Erliang Zeng*	2009-2014 – Former managing director of ND Bioinformatics Core, now Assistant Professor at University of South Dakota
Irena Lanc	2009–2014 – researcher at WashU (St. Louis)

Wei Zhang            2010-2013 – Now at AdCents (recently bought by Walmart)

Josh Livermore\*   2011-2013 - Environmental Change postdoctoral fellow; now at a biotech firm in Iowa

Andrew Thrasher   2009–2013 – Senior Software Engineer at St. Jude Children’s Research Hospital

Andrew Rider       2008–2013 (**Global Health Fellow**; coadvised with Nitesh Chawla)

Rory Carmichael   2011–2012 – Now at QuantCast

Shawn O’Neil       2009–2012 (**Schmidt Fellow, Global Health Fellow**); currently a non-tenure track teaching and research faculty member at Oregon State

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

Chris Ray            2014–2016

Xuanyi (Lexi) Li    Fall 2013–2016

Brian Bishop        Summer 2016 REU (back at U. of New Hampshire)

Kim Ngo             2014–2015

Nick LaRosa         Summer 2012–2015 (MS at Notre Dame, job at start up)

Brian Kachmarck    Summer 2010–2012; job at Amazon (Seattle)

Paul Baranay        Summer 2010–2012; **NSF Graduate fellow**, doing PhD at Yale

Pengqui Cheng (Tsinghua/ND-SURE) 2011; M.S. at University of California – Berkeley

Benjamin Drda      2007–2010; software job at Epic (Madison)

---

#### 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), Peter Bui (2012 PhD), Hoang Bui (2012 PhD), James Gentile (2013 PhD), Li Yu (2014 PhD), Dinesh Rajan (2015 PhD), Patrick Donnelly (2016 PhD), Fattaneh Bayatbabolghani (2017 PhD)

Biology: Upeka Samaroon (2012 PhD), Jenica Abrudan (2013 PhD), Changde Cheng (2014 PhD), Masha Unger (2014 PhD), Candice Lumibao (2014 PhD), Sheri Sanders (2016 PhD), Rachel Love (2017 PhD), Yiyuan Li (PhD candidate)

*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)

---

EXTERNAL SERVICE ACTIVITIES

**Journal Editorial Boards**

2013–present Associate editor, *BMC Genomics*

**Grant proposal reviewing**

2009, 2011, 2012 ad hoc, NSF BIOS directorate

2015 Panelist, NSF BIOS directorate

**Steering Committee Member**

2010–present Arthropod Genomics Symposium

**Program Chair**

2016 Program Co-chair  
6th IEEE International Conference on Computational Advances in Bio and Medical Sciences, Atlanta, GA.

2013 Sequence Analysis Co-chair  
ACM Conference on Bioinformatics, Computational Biology, and Biomedical Informatics, Washington DC.

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**

2011–2017 IEEE International Conference on Computational Advances in Bio and medical Sciences (ICCABS)

2009–2017 International Conference on Bioinformatics and Computational Biology (BiCoB)

2010–2012 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)

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

2011, 2013, 2016, 2017 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, 2013 HICOMB Workshop, IEEE International Parallel and Distributed Processing Symposium (IPDPS)

### **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)*

### **Recent reviewer for the following journals: (roughly past year)**

*Nature*

*IEEE/ACM Transactions on Computational Biology and Bioinformatics*

*Bioinformatics*

*PLoS One*

*Nucleic Acids Research*

*BMC Bioinformatics*