Computer Science Colloquia

Date: Thursday, March 30th, 2017
Time: 3:40 p.m.
Location: 2019 Morrill Hall

Integrating Diverse Data for Improved Computational Genomics

Genomics-driven analysis of many important species, which we have called "non-models", remains challenging. My group is funded by the NIH to computationally leverage newer higher-throughput sequencing and domain expert-provided metadata (biological traits like drug resistance, protein folding, community-sourced data) to tackle problems mostly related to arthropod-borne diseases (e.g., malaria).

For this talk I will focus on updates to our 2016 ACM BCB paper, which has been submitted at the request of the organizers to IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB). Previous computational approaches for imputation of missing geno-type data have relied on a linear order of markers and a genotype panel, both of which are not common in non-models. We address this limitation with our ADDIT (Accurate Data-Driven Imputation Technique) approach, which is composed of two data integration-focused algorithms: a non-model variant that employs statistical inference, and a model organism variant that better leverages reference data using a supervised learning-based approach. I will show that ADDIT is more accurate, faster and requires less memory than state-of-the art methods using model (human) and non-model (maize, apple, grape) datasets. I also may present emerging genomics results from three other funded projects, two involving closely related mosquito species complexes (Anopheles funestus and Culex quinquefasciatus) and a new R01 looking at sequence and network patterns linked to protein folding. These methods integrate sequence analysis with graph and sketch-based methods for integrating diverse data types with variable levels of uncertainty.

Part of the Computer Science Seminar Series

IOWA STATE UNIVERSITY
Department of Computer Science
www.cs.iastate.edu

Scott Emrich

Prior to joining the faculty of the University of Notre Dame, Scott Emrich obtained a BS in Biology and Computer Science from Loyola College (MD) and a PhD in Bioinformatics and Computational Biology from Iowa State University. He was the 2008 recipient of the Iowa State Zaffrano Prize for Graduate Research, and now focuses mostly in computational genomics/sequence analysis and related informatics. He has published over 75 peer-reviewed publications including venues such as Science (4, 2 covers), PNAS (3), Nature and Genome Research. Most of his work at ND involved his ten PhD students (8 grads) in a leadership role. In fact, two of his students co-won interdisciplinary submissions for both the Assemblethon2 and regulatory prediction (DREAM). He has three active awards from the NIH and serves as Notre Dame's Director of Bioinformatics.