Tracking the Few and Far Between: Computational Strategies to Speed the Discovery of Low-frequency Genomic Variation in COVID-19

Nancy Amato, Head of the Department of Computer Science and Abel Bliss Professor of Engineering, University of Illinois at Urbana-Champaign

To date, the vast majority of COVID-19 genomic research has been focused on a high-level view of SARS-CoV-2 diversity, overlooking the diversity of the viral population that exists within each COVID-19 positive patient. While viral load of SARS-CoV-2 in an individual can exceed hundreds of thousands of copies, available genomic databases only contain a single consensus version of this diverse population, discarding any low-frequency mutations. The goal of our project, CoVariants, is to develop novel computational approaches to recover these discarded variants and allow for rapid characterization of within-host diversity of SARS-CoV-2 across tens of thousands of samples. Commonly used computational tools either are not designed for the detection of low-frequency variants within viral populations, or require significant computational resources per sample. In this talk, we will describe how new parallelization strategies and approximate statistical methods can reduce the computational requirements of a widely used existing approach by up to 400 percent while preserving 100 percent of the low-frequency genomic diversity. We will end our talk by highlighting how we plan to use these improved computational methods to provide insight into the biological underpinnings of SARS-CoV-2 transmissibility and severity compared to other coronaviruses.

Nancy M. Amato is Abel Bliss Professor and Department Head of Computer Science at the University of Illinois. She received undergraduate degrees in Mathematical Sciences and Economics from Stanford, and M.S. and Ph.D. degrees in Computer Science from UC Berkeley and the University of Illinois, respectively. Her research focuses on robotics motion planning, computational biology and geometry, and parallel computing. Amato received the 2019 IEEE RAS Leadership Award in Robotics and Automation, the 2014 CRA Habermann Award, and the inaugural NCWIT Harrold/Notkin Research and Graduate Mentoring Award in 2014. She is a Fellow of the AAAI, AAAS, ACM, and IEEE.

Lawrence Rauchwerger is a professor in the Department of Computer Science at the University of Illinois. Previously, he was the Eppright Professor of Computer Science and Engineering at Texas A&M University and co-Director of the Parasol Lab. He received an engineering degree from the Polytechnic Institute Bucharest, a M.S. in Electrical Engineering from Stanford University and a Ph.D. in Computer Science from the University of Illinois. His approach to auto-parallelization, thread-level speculation and parallel code development has influenced industrial products at corporations including IBM, Intel, and Sun. Rauchwerger is a Fellow of the AAAS and IEEE and has received an NSF CAREER Award, awards from IBM and Intel.

Todd J. Treangen, Ph.D. is an Assistant Professor in the Department of Computer Science at Rice University and co-lead of the COVID-19 International Research Team. Before joining Rice, Dr. Treangen was a Research Scientist at the University of Maryland College Park. He received his Ph.D. in Computer Science in 2008 from the Polytechnic University of Catalonia (Barcelona, Spain). His research group focuses on solving large-scale computational problems specific to computational biology, with a focus on developing robust software tools targeted towards biothreat screening, infectious disease monitoring, and microbial forensics.