NSF BIOGRAPHICAL SKETCH

NAME: Warren, Andrew

ORCID: 0000-0003-2660-0103

POSITION TITLE & INSTITUTION: Research Assistant Professor, University of Virginia

(a) PROFESSIONAL PREPARATION -(see PAPPG Chapter II.C.2.f.(a))

INSTITUTION	LOCATION	MAJOR / AREA OF STUDY	DEGREE (if applicable)	YEAR YYYY
Virginia Polytechnic Institute and State University	Blacksburg, VA	Biochemistry	BS	2004
Virginia Polytechnic Institute and State University	Blacksburg, VA	Computer Science	MS	2007
Virginia Polytechnic Institute and State University	Blacksburg, VA	Computer Science	PHD	2017

(b) APPOINTMENTS -(see PAPPG Chapter II.C.2.f.(b))

2018 - present Research Assistant Professor, University of Virginia, Biocomplexity Institute, Charlottesville, VA

2017 - 2018 Research Scientist, Virginia Tech, Biocomplexity Institute NDSSL, Blacksburg

2015 - 2017 Senior Software Engineer, Virginia Tech, Biocomplexity Institute, Blacksburg

2011 - 2015 Software Engineer, Virginia Tech, Virginia Bioinformatics Institute, Blacksburg, Virginia

(c) PRODUCTS -(see PAPPG Chapter II.C.2.f.(c))

Products Most Closely Related to the Proposed Project

1. Muthiah Sathappan, Warren Andrew, Butler Patrick, Datta Debanjan, Islam Mohammad Raihanul, Ramakrishnan Naren. ProtTox: Toxin identification from Protein Sequences. Machine Learning in Computational Biology; 2019; Vancouver; c2019.
2. Eubank Stephen, Warren Andrew. Can we tell whether short DNA sequences are threatening?. ASM Biothreats; 2019; Crystal City, VA; c2019.
3. Dickerman Allan, Mao Chunhong, Wattam Rebecca, Warren Andrew. Evolutionary Models for Recognizing Toxin Protein Threats. ASM Biothreats; 2019; Crystal City, VA; c2019.
4. Adiga A, Venkatramanan S, Schlitt J, Peddireddy A, Dickerman A, Bura A, Warren A, Klahn B, Mao C, Xie D, Machi D, Raymond E, Meng F, Barrow G, Mortveit H, Chen J, Walke J, Goldstein J, Wilson M, Orr M, Porebski P, Telionis P, Beckman R, Hoops S, Eubank S, Baek Y, Lewis B, Marathe M, Barrett C. Evaluating the impact of international airline suspensions on the early global spread of COVID-19. [Preprint]. 2020 February 23. DOI: 10.1101/2020.02.20.20025882

Other Significant Products, Whether or Not Related to the Proposed Project

1. Lew JM, Mao C, Shukla M, Warren A, Will R, Kuznetsov D, Xenarios I, Robertson BD, Gordon SV, Schnappinger D, Cole ST, Sobral B. Database resources for the tuberculosis community. Tuberculosis (Edinb). 2013 Jan;93(1):12-7. PubMed Central PMCID: PMC3592388.

2. Antonopoulos D, Assaf R, Aziz R, Brettin T, Bun C, Conrad N, Davis J, Dietrich E, Disz T, Gerdes S, Kenyon R, Machi D, Mao C, Murphy-Olson D, Nordberg E, Olsen G, Olson R, Overbeek R, Parrello B, Pusch G, Santerre J, Shukla M, Stevens R, VanOeffelen M, Vonstein V, Warren A, Wattam A, Xia F, Yoo H. PATRIC as a unique resource for studying antimicrobial resistance. Briefings in Bioinformatics. 2019 July; 20(4):1094-1102. Available from: https://academic.oup.com/bib/article/20/4/1094/4056412 DOI: 10.1093/bib/bbx083
3. Outten J, Warren A. Methods and Developments in Graphical Pangenomics. Journal of the Indian Institute of Science. 2021 August 24; 101(3):485-498. Available from: https://link.springer.com/10.1007/s41745-021-00255-z DOI: 10.1007/s41745-021-00255-z
4. VanOeffelen M, Nguyen M, Aytan-Aktug D, Brettin T, Dietrich E, Kenyon R, Machi D, Mao C, Olson R, Pusch G, Shukla M, Stevens R, Vonstein V, Warren A, Wattam A, Yoo H, Davis J. A genomic data resource for predicting antimicrobial resistance from laboratory-derived antimicrobial susceptibility phenotypes. Briefings in Bioinformatics. 2021 November; 22(6):-. Available from: https://academic.oup.com/bib/article/doi/10.1093/bib/bbab313/6347947 DOI: 10.1093/bib/bbab313

(d) SYNERGISTIC ACTIVITIES -(see PAPPG Chapter II.C.2.f.(d))

1. I currently co-lead a UVA effort to create biosurveillance strategies for the state of Virginia in collaboration with the Virginia Department of Health. The purpose of this effort is to create sensor deployment design to maximize sensitivity for estimation of local disease burden and minimize uncertainty for disease forecasting and contagion characterization.
2. As the PI for a Functional Genomic Computational Assessment of Threats I helped to create multiple methods for the recognition of threats using deep learning and statistical learning algorithms. This included but was not limited to the prediction of taxonomic class, threat level, molecular function, coding regions, viral read identification, and toxin prediction. This work continues to have an impact on the application of machine learning to genome annotation, metagenomics processing for surveillance, and identification of potentially threatening DNA synthesis requests. My role was in architecting the problems to be investigated, the approach, and data necessary to improve the state of the art.
3. As a research scientist on the PATRIC/BV-BRC project I have helped to create algorithms, backend infrastructure, and interfaces for the analysis of bacterial and viral biological data. PATRIC has thousands of daily returning users and hundreds of citations. It is consistently used in the investigation of infectious disease. This resource continues to be a leader onl annotation, experimental data processing, and prediction of antimicrobial resistance.
4. I have created a novel graph representation, of genomes and their alignments, called synteny graphs. This model can be used for synteny block analysis, whole genome sequence alignment, pan-genome analysis, multiple sequence alignment, and genome rearrangement analysis. This approach was originally developed from a pan-genome perspective for prokaryotes. Novel elements include the contextualization of synteny analysis both between and within multi-contig genomes, and the ability to quickly model distant and complex evolutionary relationships between thousands of genomes. We are currently working to adopt this approach to characterize entropy of SARS-CoV-2 surveillance to estimate the number of transmission chains in a given area.