AI-POWERED DIAGNOSTICS

Tactical Bioinformatics With Secure Bloom-filter Analysis And Compression (SB-FAC)

Jacob Beal RTX BBN **Aaron Adler** RTX BBN **Bryan Bartley** RTX BBN **Aaron Helsinger** RTX BBN **Susan Katz** RTX BBN **Tom Mitchell** RTX BBN **Steven Murphy** RTX BBN **Miles Rogers** RTX BBN **Allison Taggart** RTX BBN **Dan Wyschogrod** RTX BBN

Rapid advancements in fieldable technology for the collection of biological ‘omics data offer great promise for addressing a variety of DoD-related biosecurity objectives. Nanopore sequencers, such as the ONT MinION, allows target-agnostic DNA or RNA sequencing in the field at the point-of-need, while portable tandem mass spectrometry, such as the BaySpec Continuity, can enable the same for metabolomics or proteomics.

The practical deployment of such data collection, however, is limited by bioinformatics processing: large ‘omics data collections must either be transmitted to a datacenter for processing or must be processed on-site using scarce computational resources. BBN addresses these challenges with Secure Bloom-Filter Analysis and Compression (SB-FAC), a lightweight platform for secure, real-time function-specific pre-processing and compression of genomic, transcriptomic, and metabolomic data for DoD-relevant mission goals, including pathogen-agnostic detection in environmental samples and diagnostics in clinical samples from military personnel.

The SB-FAC architecture operates in two stages, by pre-processing data on-site to compress the data to only those portions containing “signatures of interest,” then transferring the compressed data of interest for conventional bioinformatic analysis at an off-site data center. In the on-site testing phase, individual bioinformatics functions are broken into Bloom filter-based computations that identify regions of interest from within the raw biological data. This sequence detection technology has been deployed commercially as BBN’s FAST-NA Scanner system by some of the largest companies in the DNA synthesis screening industry, easily lending itself to transition in the field.

Bioinformatics functions implemented within the SB-FAC architecture may be broadly grouped into classes: matching functions and subtraction functions. Matching functions, such as toxin detection or rRNA metagenomics, retain data that contains a known “signature of interest.” Subtraction functions, such as background organism removal, retain data that contains no “signatures of disinterest.” This offers the flexibility to construct a wide variety of bioinformatics functions including screening for sequences of interest, data pre-processing and filtering, metagenomics-based analyses, and biological experiment design. We have demonstrated the efficacy and generality of this approach by applying it to diverse bioinformatics functions for both DNA sequencing–toxin detection, rRNA metagenomics analysis, and mixed microbial background removal–and RNA sequencing–RNA virus detection, ribodepletion, and biomarker quantification. With DNA sequencing functions, SB-FAC demonstrates high levels of compression (from 12x to >700x improvement vs. gzip), while retaining fidelity in bioinformatic results and requiring only laptop-scale computational resources. Early results with RNAseq indicate similar performance (>100x mean compression for RNA virus detection), and preliminary study indicates similar performance should be possible for proteomic functions as well, such as detection of infection response biomarkers.

By addressing these computational challenges of fieldable ‘omics, SB-FAC increases the ability for DoD/IC organizations to field biosecurity ‘omics-based systems, thereby increasing national biosecurity and reducing potential risks relating to biological attacks, accidents, and emergent diseases.

This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), and the Naval Information Warfare Center - Pacific (NIWC-P) via contract N66001-23-C-4502. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein. This document does not contain technology or technical data controlled under either the U.S. International Traffic in Arms Regulations or the U.S. Export Administration Regulations.