Pathogenwatch: A Global Platform for Genomic Surveillance

Background: Rapid processing of genome data can deliver information to enable decision-making at a number of scales: Local - does a genome contain signatures of resistance to antibiotics? National - is there a circulating resistant lineage that requires targeted infection control, and International - to understand continental transmission and clonal dissemination. We require a platform to rapidly analyse genomic data, visualise large datasets in aggregate, and provide open access to these results.

Methods and Results: Pathogenwatch is a web application consisting of a Node.js back-end and React front-end, with an analysis pipeline based on a novel system we call the “Runner architecture”. This architecture uses Docker and standard streams to allow analytical programs to be easily integrated without modification. Users can drag and drop genome assemblies into Pathogenwatch and receive rapid speciation, AMR prediction, and cgMLST-based clustering for major bacterial pathogens. For a selected number of species, core SNP-based trees are provided. Results for single genomes are summarised in a simple report, and collections of multiple genomes are clustered and presented through interactive maps and phylogenies to assess the spread of pathogens at a regional and national level. The Runner architecture allows for analyses to be added to the system without knowledge of the architecture and technologies used. Adding an analysis can be as simple as wrapping a command-line tool in a Docker image and updating a configuration file. Using Docker allows for integration of programs written in any language, and the architecture allows the system to evolve without needing to change the analytical programs in any way. We will demonstrate the architecture and the application live with genomic epidemiology use cases for major pathogenic species, with a focus on near real-time, rapid processing of genomic data via the web.

Conclusions: Pathogenwatch offers a scalable, flexible platform for genomic epidemiology and surveillance for public health utility.