027- RIPS: Routine Intuitive Pathogen Surveillance With Whole Genome Sequence Data

Tim Muruvanda, Arthur Pightling, James Pettengill, and Hugh Rand
Center for Food Safety and Applied Nutrition, 5001 Campus Dr, College Park MD 20742

Abstract

DNA sequence data from foodborne bacteria are monitored for emerging outbreaks. However, these activities are largely restricted to those with computational skills. We present RIPS (Routine Intuitive Pathogen Surveillance), an interactive dashboard for routine surveillance, and provide an example of its use for detecting a Salmonella enterica outbreak.
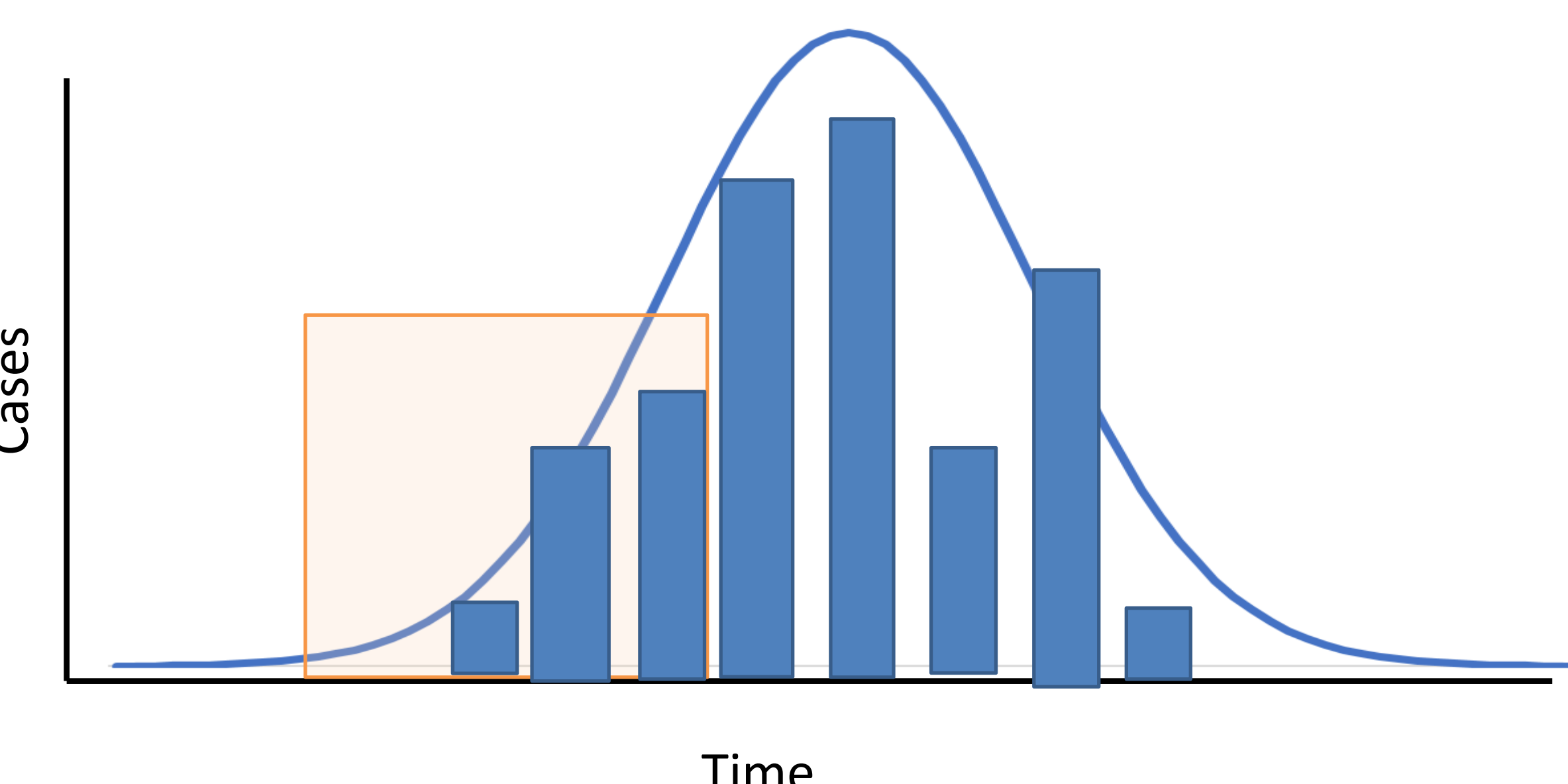
Figure 1. Epidemiological curve for foodborne outbreak

Introduction

Genomes from bacterial pathogens, provided by whole genome sequence (WGS) collection programs, are monitored for emerging outbreaks by finding matches between new and previously submitted data originating from foods, environments, and patients. The FDA CFSAN's Biostatistics and Bioinformatics Staff (BBS) uses specialized tools for WGS surveillance, which limits their use. Here, we present an accessible tool for routine WGS surveillance (RIPS – Routine Intuitive Pathogen Surveillance). RIPS makes routine surveillance of whole genome sequence data accessible to a wide range of end-users, democratizing the detection of emerging outbreaks while they are small, which may help to reduce the numbers of illnesses. RIPS provides a means to apply uniform and objective criteria that define emerging outbreaks. The criteria are tunable, which allows the user to ensure optimal performance for their needs.

Materials and Methods

RIPS was written with the Shiny package in R. It downloads Rapid Reports data from the Pathogen Detection database maintained by the National Center for Biotechnology Information. Results of a preliminary analysis that compares newly submitted sequence data to all other sequences in the database are further filtered using metrics, such as the numbers of sequences for bacteria collected from patients (clinical isolates) within a stated timeframe and the genomic distances between clinical and non-clinical isolates. The information is displayed in an easy-to-understand format, along with additional information, such as serotypes and geography.
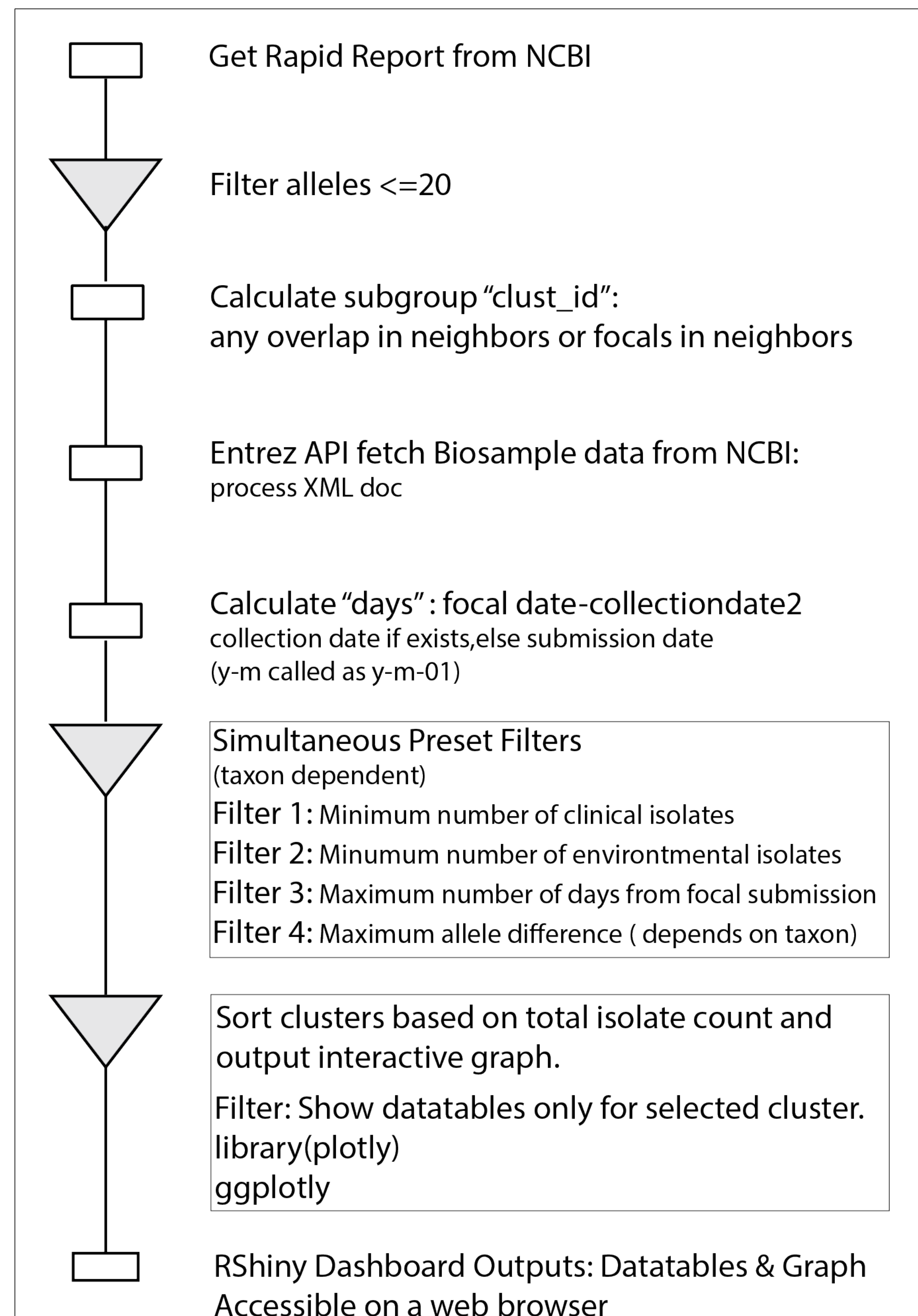

Figure 2. Processing flowchart

Get Rapid Report from NCBI

Filter alleles <=20

Calculate subgroup "clust_id":
any overlap in neighbors or focals in neighbors

Entrez API fetch Biosample data from NCBI:
process XML doc

Calculate "days" : focal date-collectiondate2
collection date if exists, else submission date
(y-m called as y-m-01)

Simultaneous Preset Filters
(taxon dependent)
Filter 1: Minimum number of clinical isolates
Filter 2: Minumum number of environtmental isolates
Filter 3: Maximum number of days from focal submission
Filter 4: Maximum allele difference (depends on taxon)

Sort clusters based on total isolate count and output interactive graph.

Filter: Show datatables only for selected cluster.
library(plotly)
ggplotly

RShiny Dashboard Outputs: Datatables & Graph
Accessible on a web browser

Figure 3. R.I.P.S. Output for July 22, 2022

Results and Discussion

A typical example surveillance run for Salmonella enterica for July 21, 2022 finds six clusters of interest (at least seven clinical isolates submitted in the prior 90 days with no more than seven allele differences from a non-clinical isolate). One cluster (PDS000113505) includes a sequence from S. enterica isolated from a sample that was collected by FDA and seven clinical isolates that are part of an outbreak (now closed).
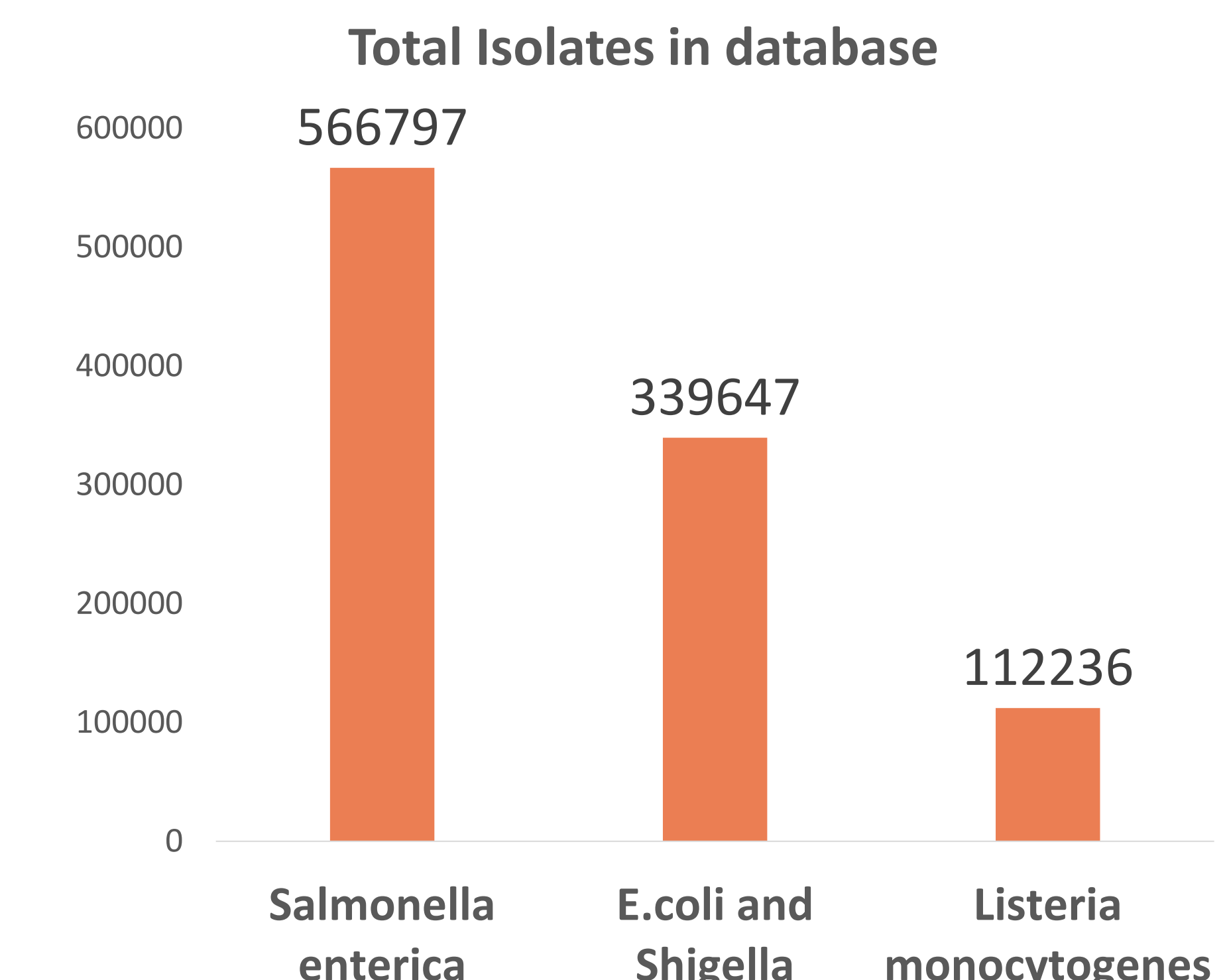

Figure 4. Total number of isolates in NCBI as of September 27,2023

Conclusion

RIPS makes it easier to perform surveillance of organisms within the NCBI's Pathogen Detection database by allowing detection of signals before they turn into outbreaks. Additionally, by simplifying surveillance, this tool encourages the use of WGS for pathogen surveillan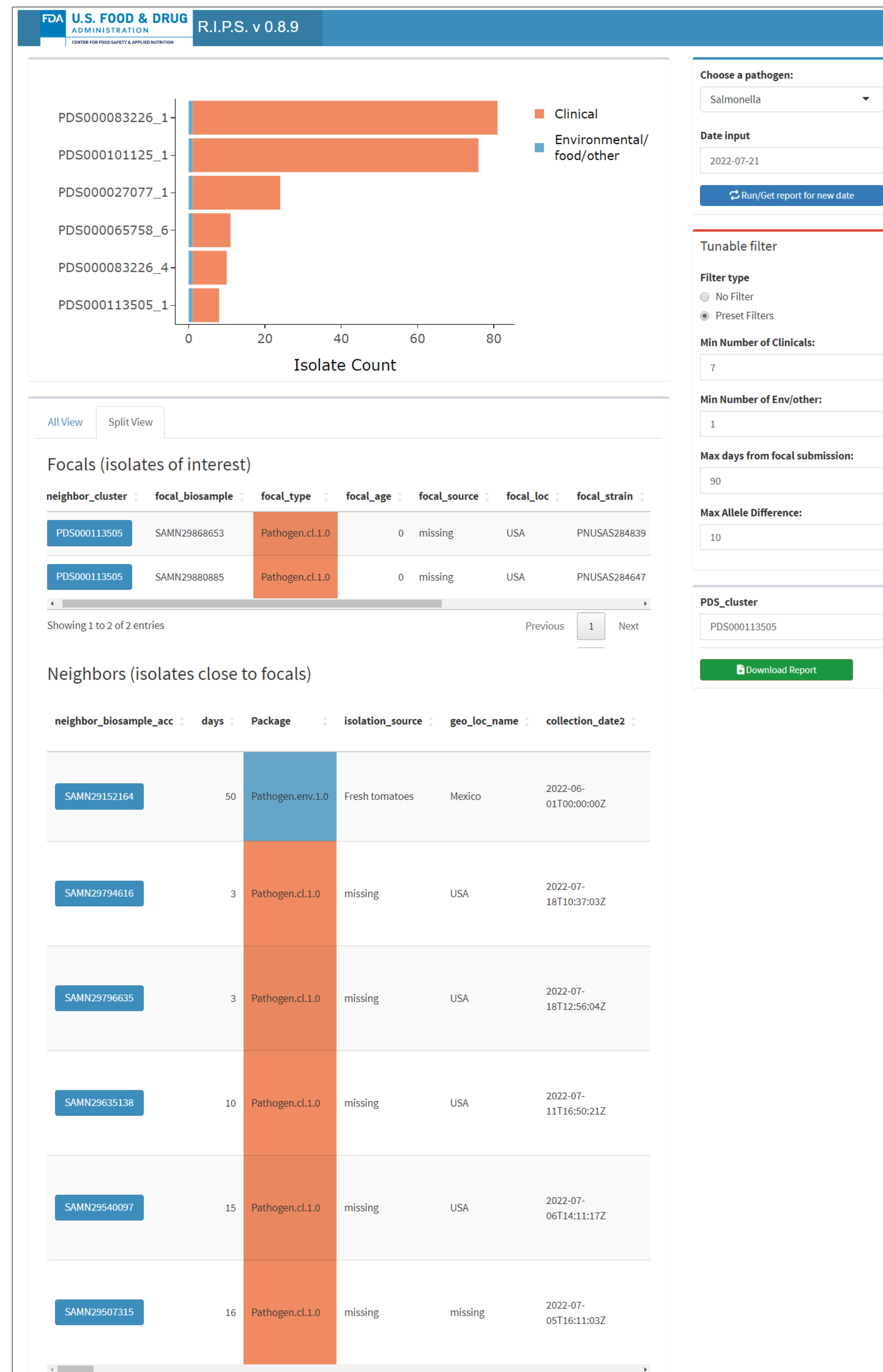ce, in foodborne pathogens and in other areas.