



HIVE Pathogen Detection Pipeline

Ekaterina Osipova¹, Alin Voskanian¹, Vahan Simonyan¹



¹Food and Drug Administration, Center for Biologics Evaluation and Research, Silver Spring, MD 20993

Whole genome sequencing (WGS) allows for more rapid, cost effective, accurate and efficient pathogen identification in a variety of environments. Next generation sequencing (NGS) allows for this to happen in a massively parallel manner and as such it is becoming more heavily relied upon in prediction of new trends, identification of food contaminants, microbiology studies and other relevant fields. To allow for precise, rapid analysis of the NGS data the High-Performance Integrated Virtual Environment (HIVE) was developed and implemented at the FDA through a partnership with George Washington University Having access to such a platform has allowed the development of tools that facilitate identification of metagenomic content.

NGS facilitates global outbreak investigations thus, protecting food supply and public health. The proposed pipeline allows for pathogen identification through specimen sampling and subsequently links to external resources to provide further information. This pipeline, implemented in HIVE, leverages the internally developed tool: CensuScope which allows an investigator to take a 'sneak peak' into the population distribution and estimate taxonomic composition then link out to other relevant sources. This allows more rapid accurate identification of food contaminants thereby ensuring greater food safety.



Food:
Food in question is gathered for analysis

Sample Collection:
Samples are collected from the foods to be analyzed.



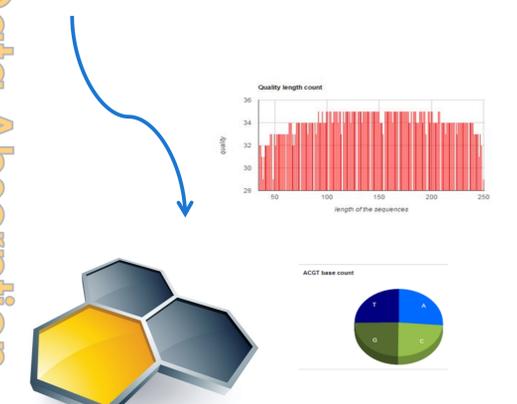
Sequencing:
The samples are sequenced using different machines/techniques like Illumina. Resulting in output files that are a digital representation of the samples' sequences.

Background Abstract

With the decreasing costs of sequencing, NGS data is quickly becoming the preferred typed of data for pathogen detection analysis. The HIVE pathogen pipeline attempts to detect pathogens present in any sequenced samples. This pipeline has specifically been applied to foodborne pathogen detection but may have applicability in other spheres. The pipeline leverages the HIVE infrastructure and internally developed metagenomic identification tool: CensuScope; allowing it to identify and link out to NCBI and other pertinent databases for detailed information on pathogen presence in the sample of interest. The pipeline also produces results and graphics which can be easily interpreted.

Conclusion

Data Absorption



HIVE:
Data is uploaded to HIVE and goes through automatic Quality Control and Validation

Identification

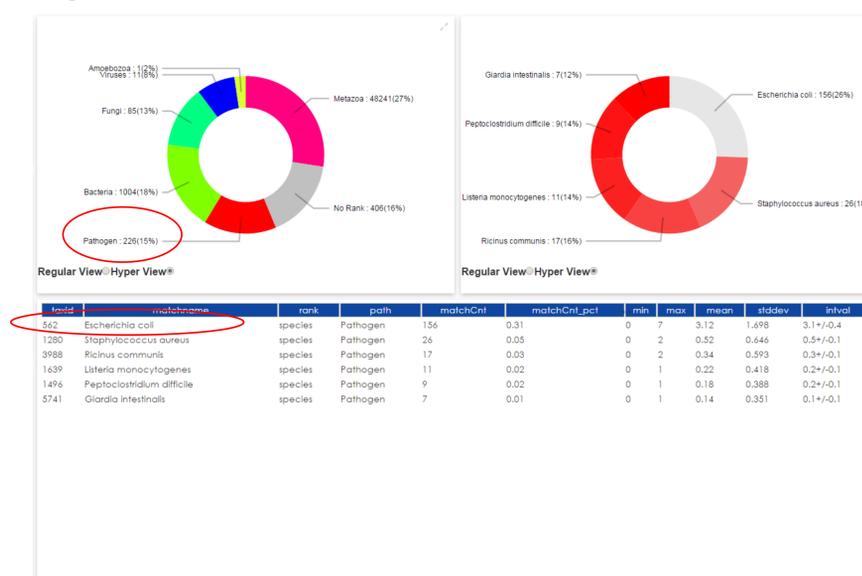
External Sources:

Can easily jump to a variety of external sources to get an in depth look at the different species detected. NCBI taxonomy for the selected item is shown by default; however linkage to WebMd and CDC also exists.

Pathogen Interface

Pathogen Pipeline:

Graphs showing the results from the Pathogen Pipeline. The user can get a general idea and an in depth look at the information generated.



ACKNOWLEDGEMENTS

Luis Santana-Quintero,CBER,FDA

For more information please email vahan.simonyan@fda.hhs.gov
HIVE FDA website: <https://minihive.fda.gov>
HIVE public website: <https://hive.biochemistry.gwu.edu>