



PHoeNix: Platform-Agnostic Bioinformatic Pipeline for Strengthening CDC's Antimicrobial Resistance Laboratory Network

J. V. Hagey¹, N. Vlachos¹, A. G. Kent¹, K. A. Moser¹, P. Qi^{1,2}, T. Masters^{1,2}, A. L. Halpin^{1,3}

1. Division of Healthcare Quality and Promotion, US Centers for Disease Control and Prevention, Atlanta, GA 30329, USA

2. Goldbelt Professional Servies, Chesapeake, VA, 23320, USA

3. U.S. Public Health Service Commissioned Corps, Rockville, MD 20852, USA

Jill Hagey, PhD
Division of
Healthcare Quality
Promotion
jpk9@cdc.gov



Background

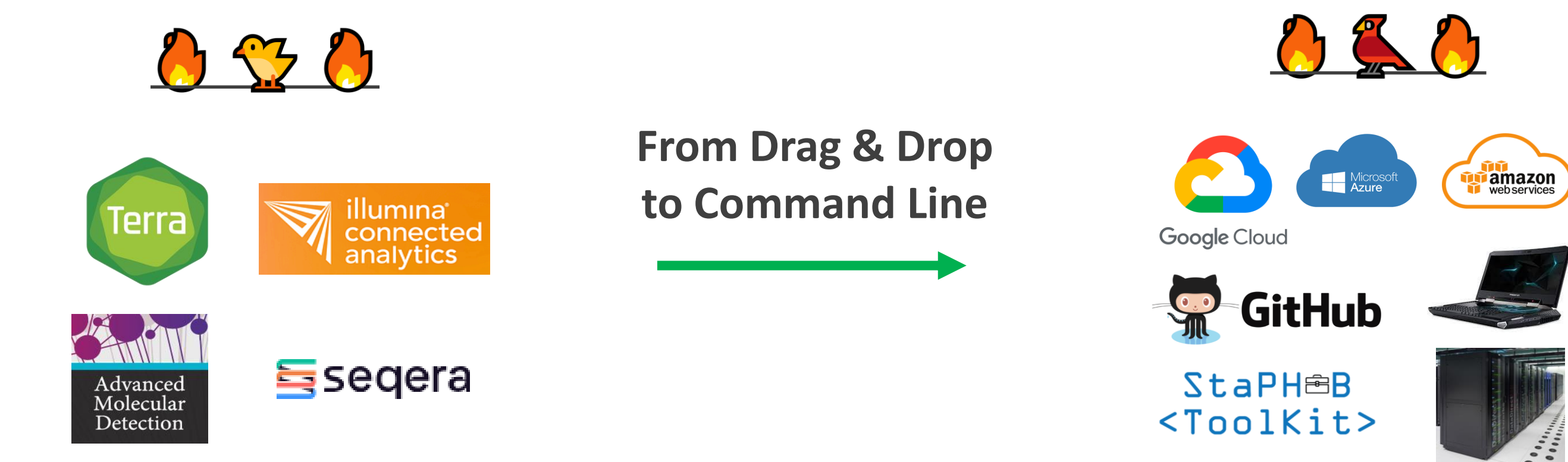
Annually in the United States over 35,000 people succumb to antibiotic-resistant infections, with over 2.8 million annual infections estimated in CDC's 2019 Antibiotic Resistance Threats Report. As sequencing of antimicrobial resistant pathogens causing healthcare-associated infections (HAIs) becomes increasingly prevalent in public health labs (PHLs), we are challenged by diverse IT requirements, computational resources, levels of bioinformatic expertise, and lack of free, accessible, standardized bioinformatic pipelines. These variations make analysis, reproducibility, and comparisons difficult, highlighting the critical need to bridge the gap between public health lab (PHL) capabilities and analyses of data to support public health action for HAIs and antimicrobial resistance (AR).

Methods

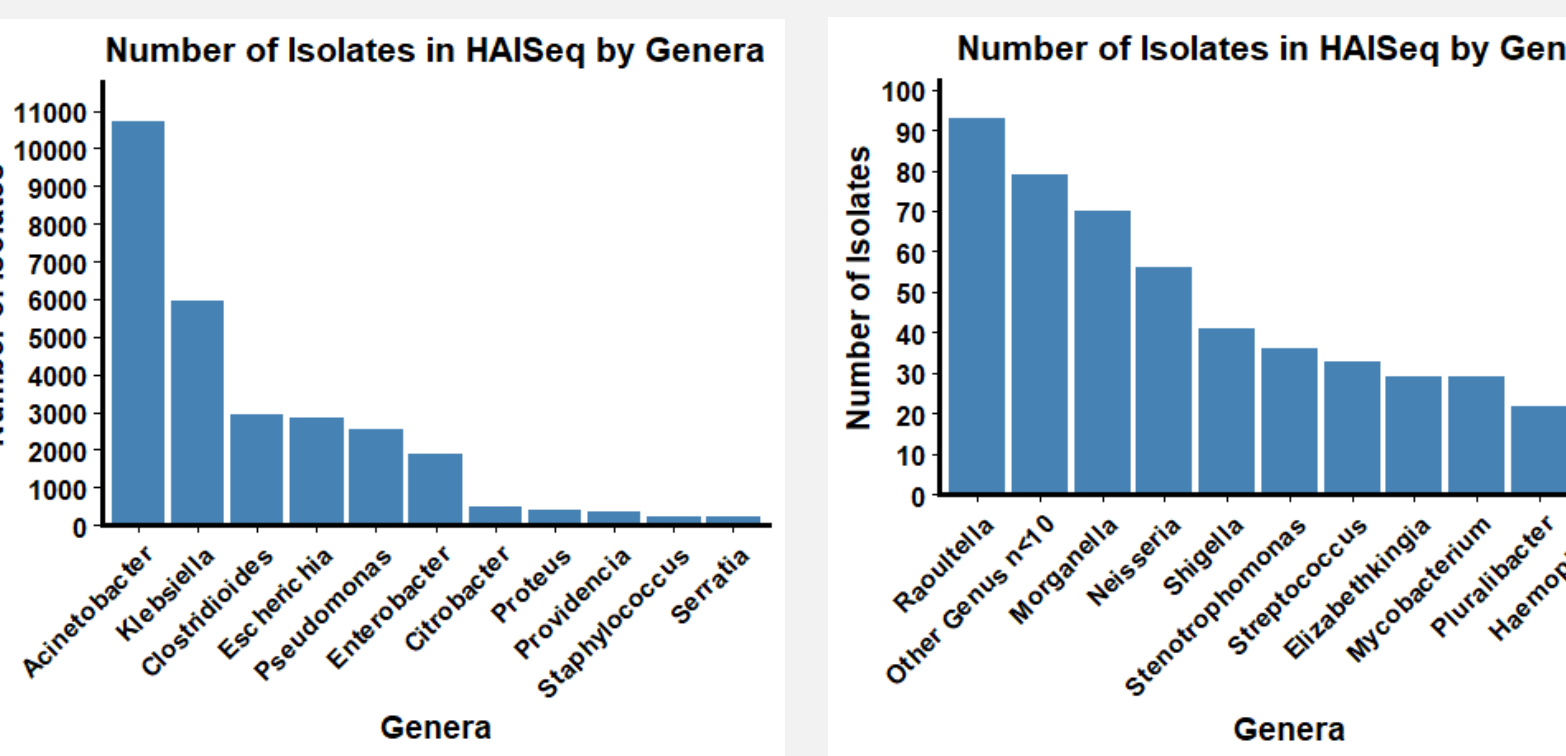
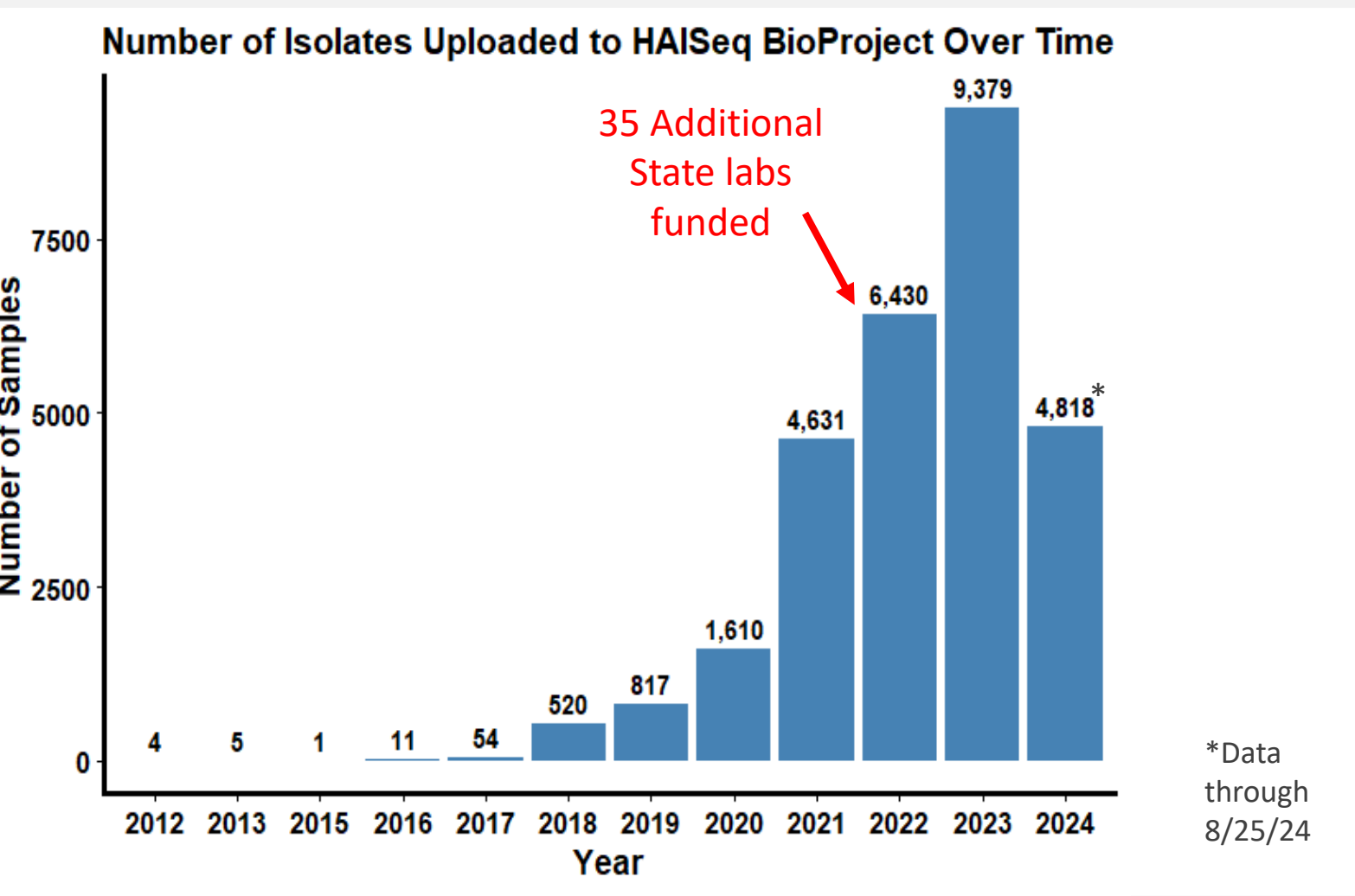
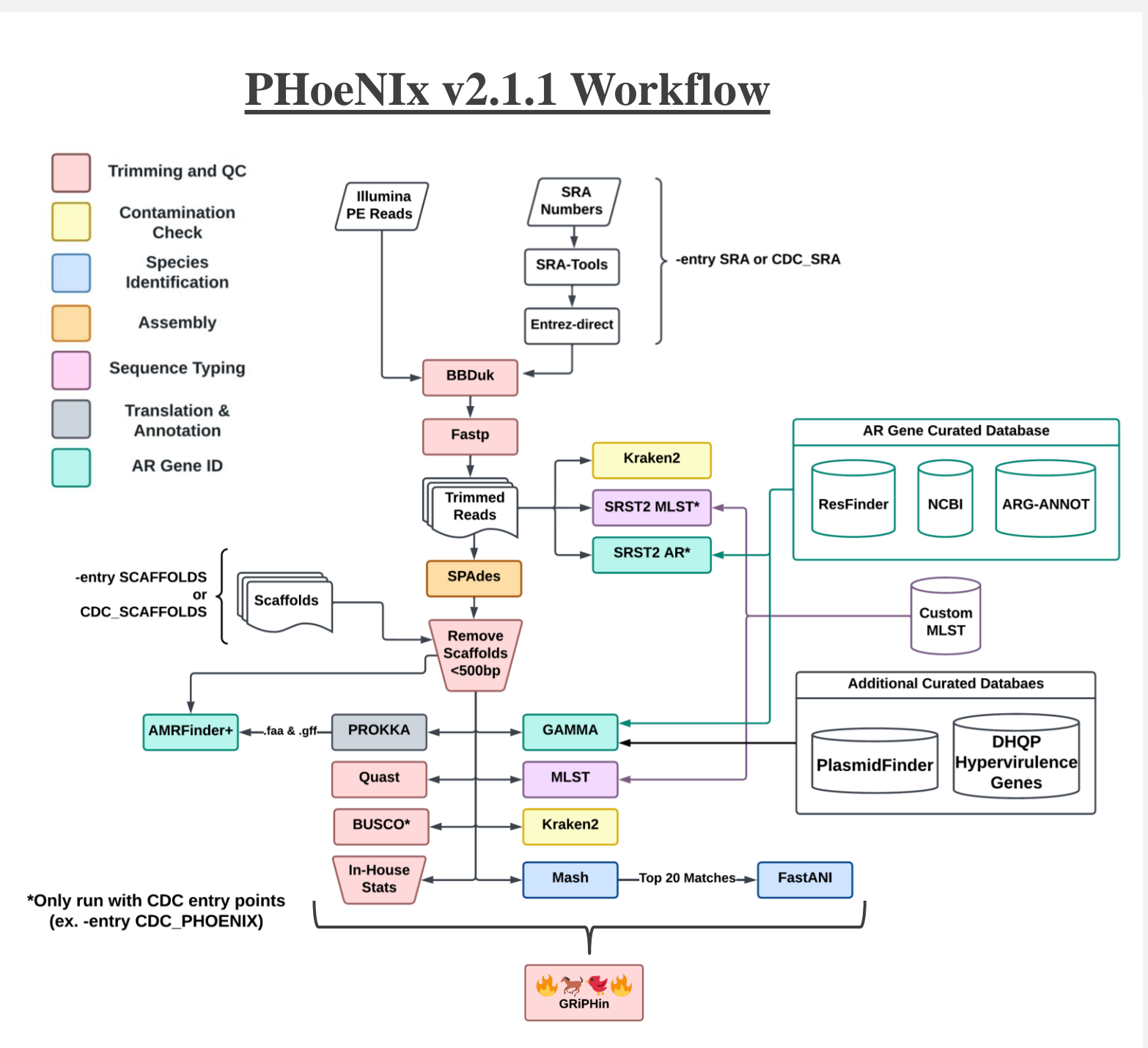
To address these challenges, we developed PHoeNix (Portable Healthcare Next generation Informatics), an open-source bioinformatic pipeline. It accepts Illumina paired-end reads or assemblies as input, and performs quality control, assembly (if needed), taxonomic identification, multilocus sequence typing (MLST), and identification of plasmid replicons, AR genes and select hypervirulence marker genes. PHoeNix also flags novel AR gene variants and MLST profiles for submission to their appropriate database and generates submission forms for NCBI. Developed using Nextflow and nf-core best practices, PHoeNix is platform-agnostic and utilizes containers for each step, ensuring easy installation and highly reproducible results across various computing infrastructures. Full documentation and code are available at <https://github.com/CDCgov/phoenix>.

Considerations for pipeline development:

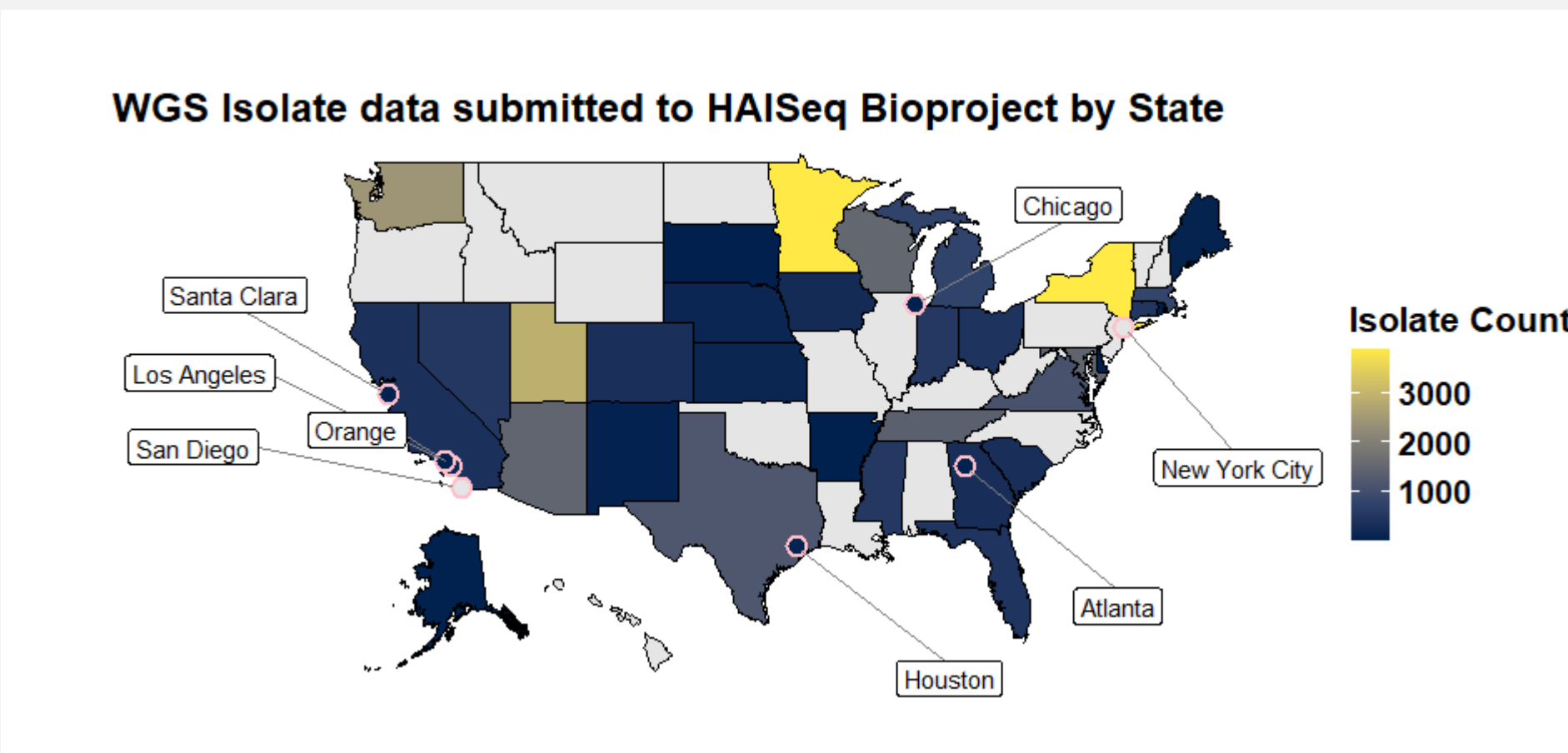
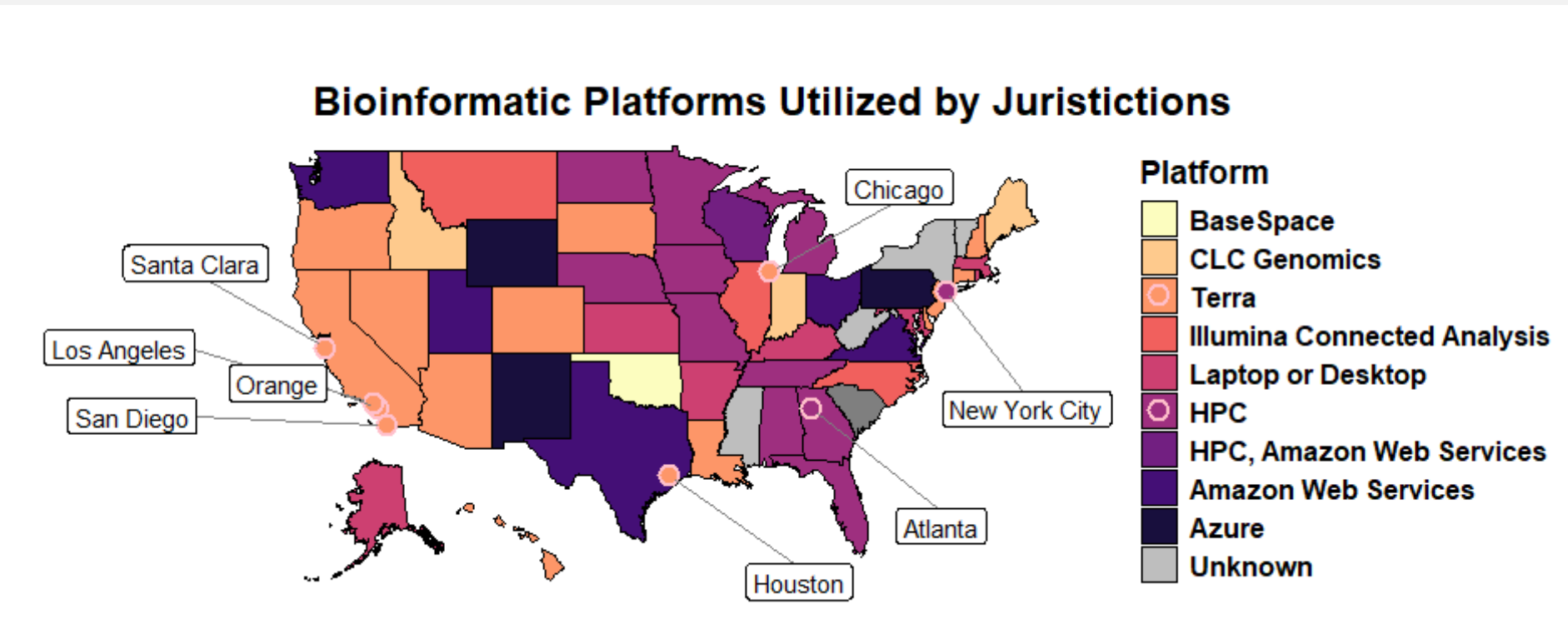
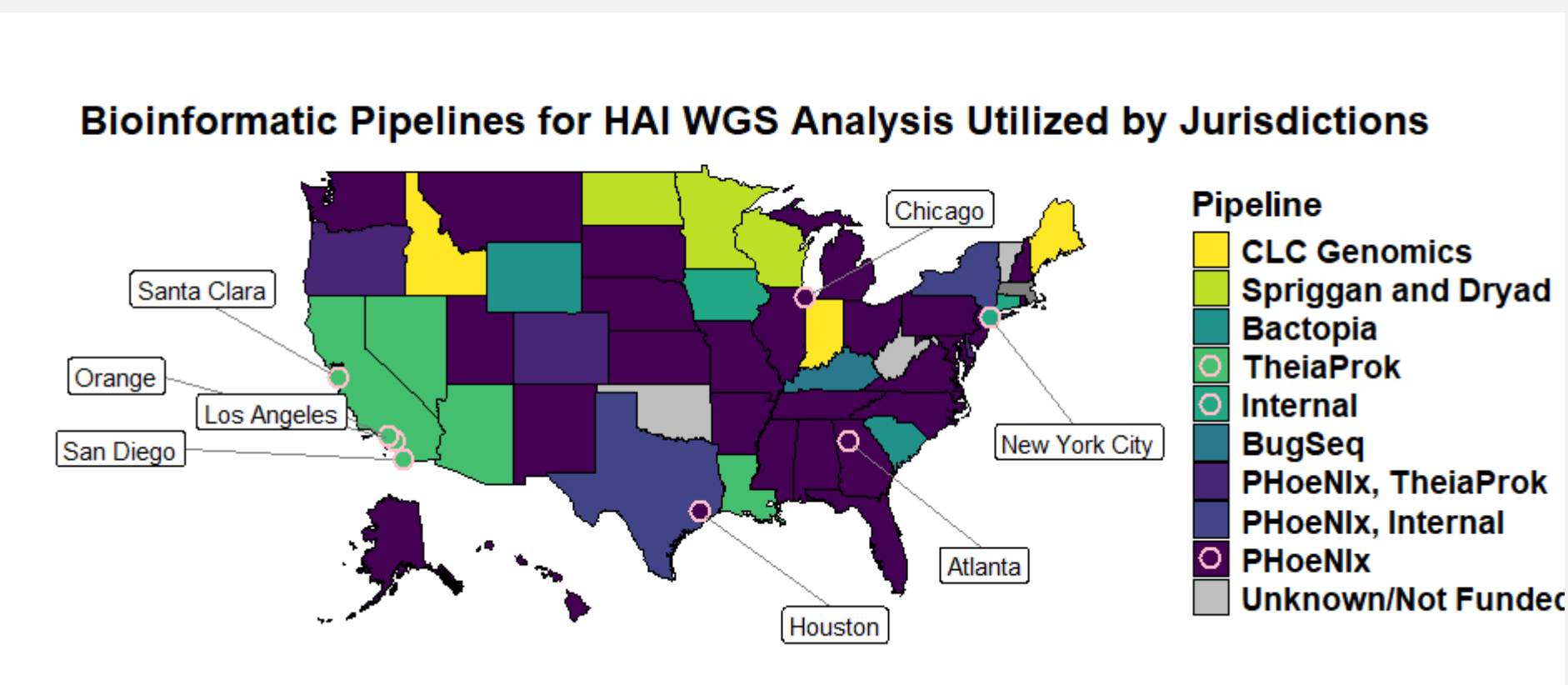
- Portable and reproducible
- Open source
- Accessible on HPC/Cloud/Terra/CLI/ICA
- Easy to use
- Low \$\$ for cloud users
- Clear guidance on limitations and documentation on use.



Results



Currently, 36 domestic PHLs and 4 international partners use PHoeNix on 3 of the major platforms, demonstrating its widespread adoption.



Conclusions

PHoeNix is a portable, reproducible pipeline that can be run on all major platforms that public health labs use.

Development of PHoeNix continues and aims to reduce any barrier for PHLs to meet their local public health needs and the objectives of the AR Laboratory Network (AR Lab Network).

PHoeNix plays a pivotal role in achieving CDC's mission of supporting PHLs with a resource, backed by CDC technical expertise and support, for rapidly detecting emerging HAI/AR threats as it offers a standardized approach for identifying and characterizing HAI bacterial pathogens along with subsequent data sharing.

Future Directions

The PHoeNix Universe of software is expanding to include:

- Long-read sequence data as input
- Species-specific sub-workflows (*C. difficile*, *S. aureus*, *E. coli*)
- 🦉🌿🐦🔥 PhyloPHoeNix: pipeline for relatedness determination. This pipeline is meant to be run in tandem with PHoeNix, supporting outbreak investigations through phylogenomic analyses of isolates suspected to be part of an outbreak.

These tools enhance local, national and international capacity for utilizing sequence data for HAI/AR surveillance and outbreak investigations.

Affiliates / Partners

We thank the labs of AR Lab Network for their tireless work identifying and tracking antimicrobial resistant threats in the United States. We appreciate the feedback from our beta testers, PHoeNix community at-large, and StaPH-Bs hosting of containers that has greatly improved the functionality, output, and utility of the pipeline.

