

# Summary Report: Lessons learned from a one-year pilot of population-based genomic surveillance of *Shigella* for local public health response in King County, Washington

**Public Health – Seattle & King County**

**Report date: September, 2026**

## Summary

Public Health – Seattle & King County (PHSKC) partnered with the Washington State Department of Health (WADOH) and the Washington State Public Health Lab (WAPHL) to implement a one-year genomic surveillance pilot for *Shigella* in King County. Between October 1, 2024 and September 30, 2025, we attempted whole-genome sequencing (WGS) on all successfully cultured *Shigella* isolates and utilized sequencing data and genomic clustering to inform local public health response.

The program highlighted that local genomic surveillance is feasible with dedicated resources, including funding and a collaborative team of lab, state, and local public health professionals. Among participating teams, implementing genomic surveillance led to increased skills and experience using genomic similarity to inform case linkage decisions, insight into baseline local genomic diversity of *Shigella*, development of a pipeline to predict antimicrobial resistance from sequencing, and expanded genomic surveillance capacity, both for *Shigella* and more broadly.

## Background

Public Health – Seattle & King County (PHSKC) initiated a population-based genomic surveillance pilot for *Shigella* with the Washington State Department of Health (WADOH) and the Washington State Public Health Lab (WAPHL). The pilot aimed to explore the feasibility and public health utility of sequencing and analyzing all *Shigella* isolates identified within King County while expanding PHSKC's skills and capacity to incorporate genomic epidemiology into routine surveillance and outbreak investigations. We selected *Shigella* because of concerns with increasing incidence, antimicrobial resistance (AMR), and disproportionate impact on people experiencing homelessness (PEH)<sup>1</sup>. *Shigella* was relatively straightforward to expand from ad-hoc sequencing to population-based genomic surveillance because case volume was sufficient for surveillance and appropriate for staff capacity, healthcare facilities were already required to send PCR-positive specimens to WAPHL for culture, and WAPHL already had the ability to perform whole-genome sequencing (WGS) on *Shigella* isolates. This allowed the pilot to focus on developing interagency collaboration, data sharing, and integration into public health response.

The pilot ran between October 1, 2024 and September 30, 2025 and aimed to sequence all successfully cultured *Shigella* isolates from King County, then share sequencing data back with PHSKC epidemiology and investigation teams to inform public health response.

## Approach

Between April 2024 and July 2024, the PHSKC Genomics team met with each team across all three agencies involved in conducting genomic surveillance to leverage existing sequencing, bioinformatics, and genomic epidemiology infrastructure and to determine workflow priorities and needs. The teams were PHSKC Foodborne and Enteric Disease (FB&E), WADOH Molecular Epidemiology Program (MEP), WADOH FB&E, WADOH Data Integration and Quality Assurance (DIQA), WAPHL Enterics laboratory, WAPHL Advanced Molecular Detection (AMD) laboratory, and WAPHL Bioinformatics (Figure 1).

| Public Health Seattle & King County |                               | Washington State Department of Health |                               |                                        | Washington State Public Health Lab |                                         |                |
|-------------------------------------|-------------------------------|---------------------------------------|-------------------------------|----------------------------------------|------------------------------------|-----------------------------------------|----------------|
| Genomics                            | Foodborne and Enteric Disease | Molecular Epidemiology                | Foodborne and Enteric Disease | Data Integration and Quality Assurance | Enterics laboratory                | Advanced Molecular Detection laboratory | Bioinformatics |

**Figure 1. Pilot participant teams by agency**

The PHSKC Genomics team drafted a data and communications workflow based on those conversations and convened the first all-pilot partner call in August 2024 to review and agree on the workflow (Figure 2). Specific team priorities addressed during development included prioritization of specimens related to outbreak investigations as needed, development of genomic data capacity in the Washington Disease Reporting System (WDRS), and communication of select metadata and task completion statuses between teams.

The final workflow was as follows. WAPHL Enterics identified all cultured isolates from King County residents and transferred them to WAPHL AMD for WGS on an Illumina MiSeq. WAPHL AMD uploaded assemblies to NCBI and shared sequence accessions with three teams: WAPHL Bioinformatics processed sequence data using CDC's PHoeNix pipeline<sup>2</sup>; WADOH MEP conducted phylogenetic analyses using WAPHL's BigBacter pipeline and interpreted genomic findings<sup>3</sup>; and WADOH DIQA linked sequences to cases and uploaded metadata into WDRS. WADOH MEP provided PHSKC Genomics with reports containing phylogenetic analyses, genomic clustering interpretations, and sequence metadata. Data shared with PHSKC included interactive SNP matrices and phylogenetic trees via Microreact which enabled the PHSKC genomic epidemiologist to overlay local data on the visualizations and download underlying data for further analysis (Figure 2).

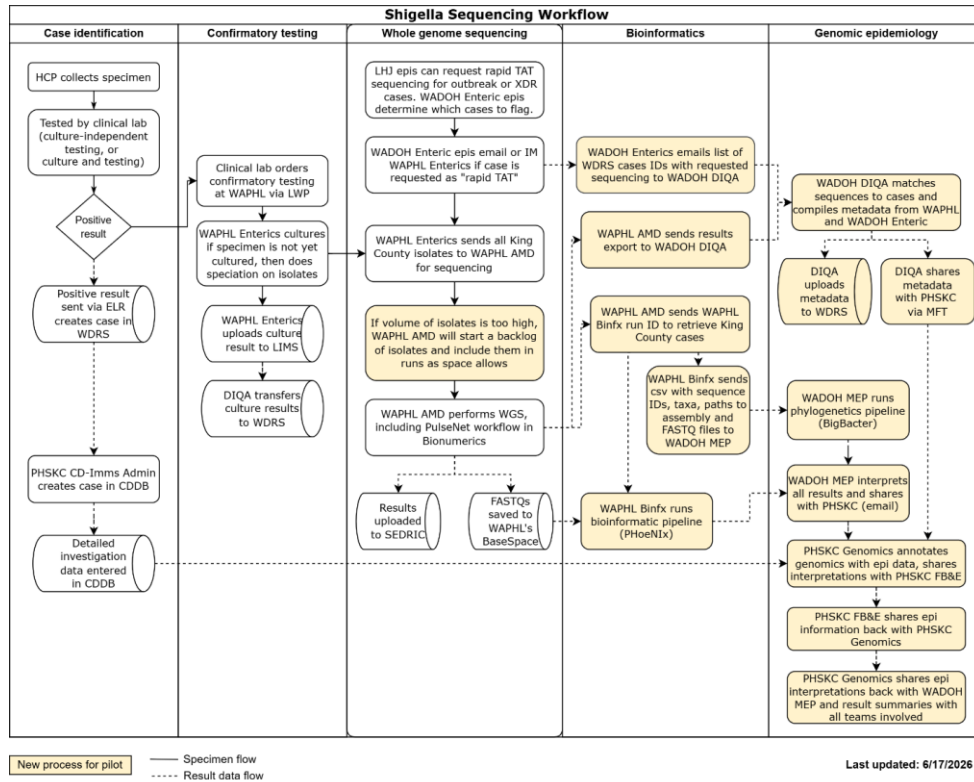


**Figure 2. *Shigella* sequencing pilot – testing, analysis, and data sharing workflow**

Initially WADOH MEP classified genomic linkages as strong (0–10 SNPs) and intermediate (11–50 SNPs). While literature indicates all genomic linkages of 0-10 SNPs may suggest recent *Shigella* transmission<sup>4</sup>, PHSKC FB&E did not have capacity to review all cases that met that threshold. To prioritize the genomic linkages most likely to have an identifiable epidemiologic linkage, PHSKC and WADOH MEP agreed to create a third category to identify very strong genomic linkages (0–5 SNPs) that could signal recent transmission and likely epidemiologic linkage between cases. The PHSKC genomic epidemiologist reviewed all strong and very strong linkages using the genomic data shared by WADOH MEP and the sequence-to-case linkage from WDRS and flagged a more conservative set of cases for the PHSKC FB&E team to review. Cases were flagged if they were in a cluster of at least 3 cases with sequences within 0–2 SNPs and collected within 60 days of each other. PHSKC FB&E re-assessed the available data to look for epidemiological linkage between those cases (Figure 3).

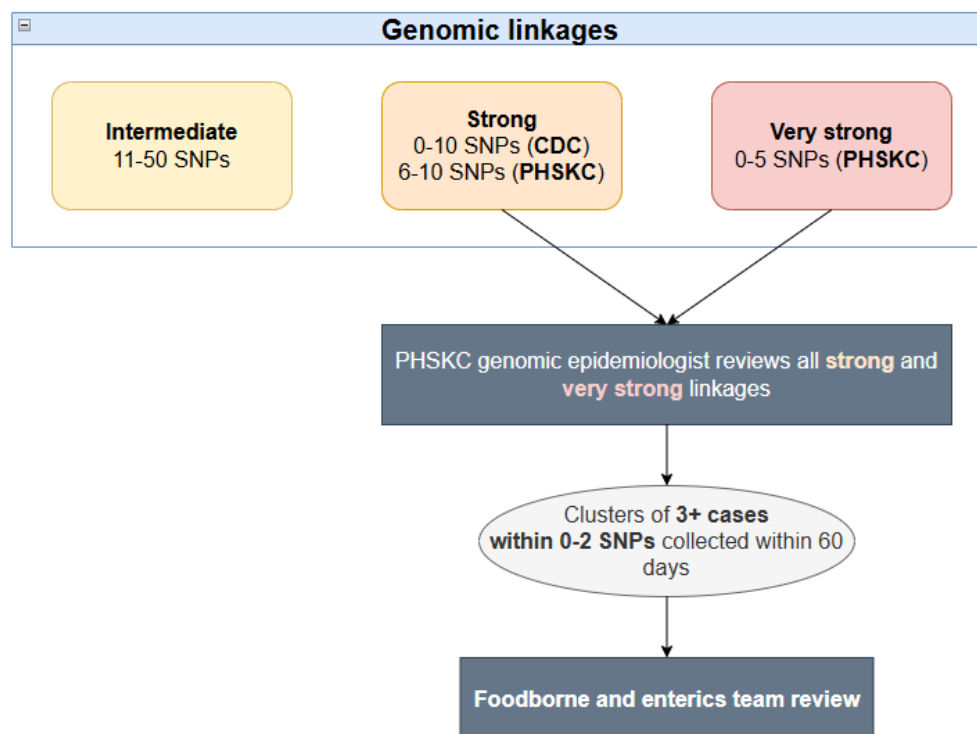


**Figure 3. *Shigella* genomic linkage review thresholds at PHSKC**

Pilot partner calls were facilitated monthly with the participating teams mentioned above; teams discussed changes in capacity and needs for workflow revision and participated in collaborative data interpretation.

## Findings

PHSKC conducted post-pilot interviews to assess the utility and feasibility of the pilot. Partners shared that the pilot was appropriate and useful. Workload was 2–3x higher in the first few months while processes were being streamlined, then regulated to a range of 1–12 hours/week per team by Month Four. Teams with the highest workload were PHSKC Genomics, WADOH MEP, and WAPHL Bioinformatics (8–12 hours/week per team). Across teams, participants felt the work led to broader

benefits to genomic surveillance capacity (including knowledge, data infrastructure, established workflows and more) beyond *Shigella*, which increased the acceptability of the workload.

Genomic surveillance capacity building was a major success from the pilot. WAPHL Bioinformatics improved *Shigella* reference genome selection and processes used across pathogens for internal data tracking and partitioning methods. WADOH MEP used the pilot meeting format as a model for other teams and added the very strong genomic linkage category to other pathogen reports. WADOH DIQA used the new *Shigella* data package as a framework for other pathogens, speeding up the subsequent addition of genomic data storage for multi-drug-resistant organisms (MDROs), *C. auris*, and 7 additional enteric diseases in WDRS, after four years of only storing genomic data for SARS-CoV-2.

The pilot led to nearly 95% of cases eligible for sequencing having at least one isolate sequenced, a dramatic increase compared to the previous year during which 7.6% of eligible cases had at least one isolate sequenced. By sequencing isolates from all eligible cases instead of only on request, the pilot increased the representativeness of key epidemiologic characteristics among cases with sequencing, giving better insight into transmission across the population. The increase in genomic data also gave better insight into the genomic diversity of *Shigella* circulating in the county. The pilot revealed King County sequences fell into three highly divergent genomic groups – two groups of *S. flexneri* and one group of *S. sonnei*. Nearly two-thirds of sequences (65%, n=106) belonged to a genomic cluster, defined as 0–10 SNPs. This proportion varied by species, with a higher proportion of *S. sonnei* sequences (82%, n=54) belonging to a genomic cluster than *S. flexneri* sequences (55%, n=52). Characterizing local genomic diversity also improved our ability to identify cases that represent new introductions to the county.

Additionally, we evaluated how well genomic linkage correlated with epidemiologic linkage. There were six pairs of cases with sequenced isolates and an identified epidemiological linkage during the pilot. All pairs of sequences diverged by less than 5 SNPs, falling into the very strong genomic linkage category. In other words, when sequencing was available, it identified all the case linkages identified by epidemiological investigation (6 out of 6) (Figure 4). We could not assess the rate of “false positives” – cases with a genomic link but no epidemiological link – because our epidemiological data for those cases are likely incomplete. We suspect the number of epidemiological links are under-counted due to inability to reach all cases for interview and interview limitations like recall bias. In addition to the six pairs with concordant genomic and epidemiological linkage, there were 183 more pairs of cases with strong genomic linkage and no identified epidemiological linkage. Some of these likely represent undetected epidemiological linkage, but how many remains unknown. Additional studies with greater resources for case follow up are needed to determine the proportion of strong genomic linkages that do not correlate to epidemiological linkage. In the meantime, the observed genomic similarity of the cases with known epidemiological linkage is encouraging.

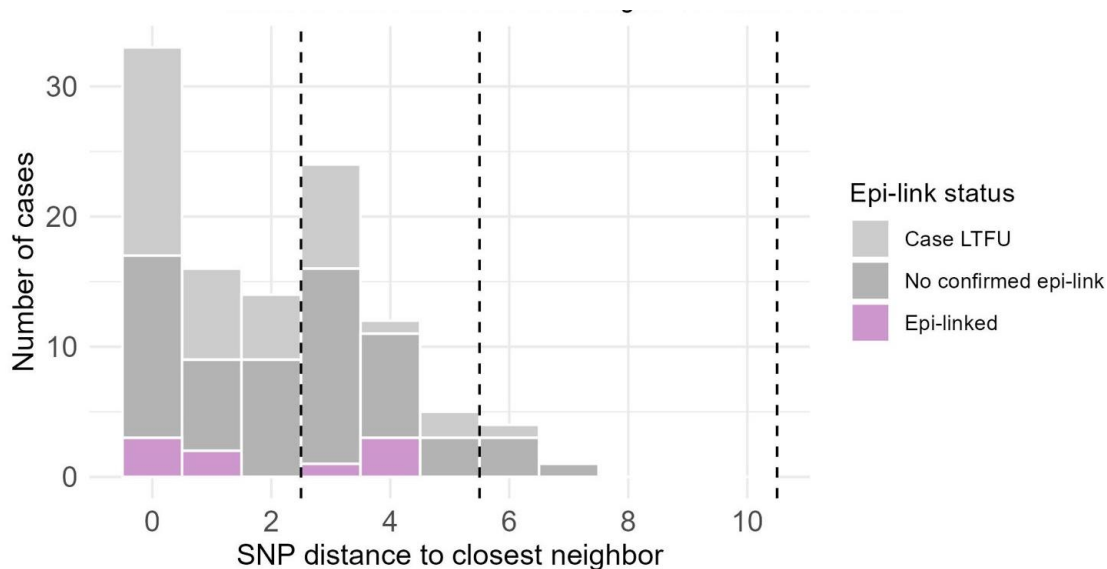


**Figure 4. Number of cases by SNP distance to nearest neighbor, October 2024-2025, King County**

We found genomic data useful for ruling cases out of clusters as well, especially in the context of insufficient epidemiologic data. For example, one suspect cluster of cases among PEH (who we were unable to interview) was found to be three unrelated genetic strains. Understanding these as unrelated cases, rather than a community outbreak spanning multiple facilities, informed the prioritization of time and resources from PHSKC. The data to inform prioritization of work is increasingly important in the context of limited public health resources.

Finally, in addition to clustering insight and expanded capacity, the pilot produced an analytic pipeline to predict AMR from sequencing (specifically from PHoeNix outputs). Preliminary results demonstrated near perfect agreement between cases that were phenotypically extensively drug resistant (XDR) and isolates predicted to be XDR by sequencing. Predicting XDR by sequencing led to a 222% increase in cases predicted to be XDR over the status quo of phenotypic testing (9 cases to 29 cases).

Pilot participants identified data infrastructure as a primary challenge of the pilot workflow. The various databases used to store and share data had limitations with storage of sequencing data, clustering data, and result status tracking. A shared tracking spreadsheet among WADOH teams was a temporary solution, and transition away from this tool was named a priority goal for future work. Another limitation of the pilot was that 45% of reported *Shigella* cases did not have an isolate available and therefore were ineligible for sequencing. Turnaround time (TAT) was also a limitation; median TAT for sequencing analyses was 17 weekdays from specimen collection, compared to 4 weekdays for completing an epidemiologic investigation, meaning results often arrived after an investigation was complete and were only used for retrospective insight. Reducing TAT would be another goal of future work.

## Conclusions

The *Shigella* genomic surveillance pilot demonstrated that local genomic surveillance for this pathogen was feasible and useful. Sequencing offers opportunities for improved cluster detection and resolution, case rule-out, and an unprecedented opportunity for AMR surveillance. Developing this workflow

resulted in a snowball effect of new capacity developments which extended beyond the pilot, including creation of new genomic data storage in WDRS for the first time in over four years.

We attribute much of the pilot's success to the creativity and collaboration of partners from all three agencies. Monthly meetings enabled teams to build familiarity with other teams' processes and solve problems collaboratively. Communication with interest holders throughout the design, implementation and evaluation of this pilot was integral to meet multiple team priorities. Dedicated PGCoE funding enabled staff to focus time on this work, without which the pilot would not have been possible. This pilot exemplifies what can be achieved when dedicated funding supports lab, state, and local partners together toward a shared goal.

## References

1. Public Health - Seattle & King County. 2022 Communicable Diseases and Epidemiology Annual Report - King County, Washington [Internet]. [cited 2026 May 29]. Available from: <https://www.kingcounty.gov/en/dept/dph/health-safety/disease-illness/annual-report>
2. Hagey JV, Vlachos N, Kent AG, Masters T, Qi P, and Halpin AL. CDCgov/phoenix: v2.2.0. Zenodo; 2025 [cited 2026 May 29]. Available from: <https://zenodo.org/records/8147510> doi:10.5281/zenodo.8147510
3. Johnson J. BigBacter. Available from: <https://github.com/DOH-JDJ0303/bigbacter-nf>
4. Mitchell HD, Mikhail AFW, Painset A, Dallman TJ, Jenkins C, Thomson NR, et al. Use of whole-genome sequencing to identify clusters of *Shigella flexneri* associated with sexual transmission in men who have sex with men in England: a validation study using linked behavioural data. *Microb Genomics*. 2019;5(11):11. doi:10.1099/mgen.0.000311