Shotgun Metagenomics: A Success Story of the rapid development and implementation of a genomic surveillance method during the 2022 human Monkeypox outbreak

Jared Johnson, PhD
Lead Bioinformatician, Genomics Section
Michigan Department of Health and Human Services
3/14/2023
Mpox background

- Potentially life-threatening disease caused by the mpox virus (formerly monkeypox virus) that causes rash lesions and flu-like symptoms.

Photo Credit: NHS England High Consequence Infectious Diseases Network

Classifying mpox viruses

2022 Outbreak
~0.08% mortality rate
Not select agent!

~10.6% mortality rate
Is select agent!

Species
mpox

Mpxo surveillance

Real-time PCR:

• Non-Variola Orthopoxvirus assay (VAC1)
• MPXV-Specific assay
  ➢ Primers in terminal regions – prone to mutations

High Throughput Sequencing:
• Capable of achieving clade-level or even lineage-level classification.
Mpox surveillance

**Tiled amplicon sequencing**
- High throughput but does not adapt well to evolving viruses.

![Diagram](image_url)
Outbreak and pipeline development timeline

Reasons for selecting shotgun approach:
1. Wanted to explore metagenomic approach for SARS-CoV-2
2. Have the appropriate sequencers
3. Anticipated large workload adopting tiled approach on wet-bench side
Overview of sequencing process

Clinical Sample → DNA Extraction (Qiagen EZ1) → qPCR (LRN Non-Variola Orthopoxvirus) → Positive Tests → Library Prep (Illumina DNA Prep) → DNA Sequencing (NextSeq 550; mid output v2.5) → Sequence Data (FASTQ)
Overview of bioinformatic pipeline

Read Processing
- Paired-End FASTQ
- Quality Filtering & Trimming
- Extract Mpox Reads w/ Clade I Ref

Assembly
- Consensus Assembly (iVar)
- De novo Assembly (SPAdes)

Classification
- Nextclade
- Mash

Results
- Lineage Classification
- Clade Classification
- Species Classification
Mash was fast and effective method for classifying mpox clades!
• Approximately 42% of Vac1 reactive samples were sequenced.
• Those not sequenced typically had high Ct values or pre-dated the pipeline creation.
Sequencing and pipeline metrics

Samples Per Run
• ~20 samples

Average Read Metrics
• ~19 million high quality reads per sample
• ~1.4 % mapped to the Clade I reference genome

Average Assembly Metrics
• ~86.3% genome fraction
• ~140-165X depth of coverage
Establishing Ct thresholds

De novo assembly generally resulted in higher genome recovery rates.

**Lineage-level:**
- Target: 95% genome recovery
- Threshold: $Ct \leq 21$

**Clade-level:**
- Target: 10% genome recovery
- Threshold: $Ct \leq 27.5$
Classification performance

Clade Level (Ct ≤ 27.5)

- Nextclade: 69 (84% pass, 13 (16%) fail)
- Mash: 72 (88% pass, 10 (12%) fail)

Lineage Level (Ct ≤ 21)

- Nextclade: 53 (93% pass, 4 (7%) fail)
Further use of lineage data
Summary

• Metagenomic approach allowed for rapid development of a surveillance method during an emerging threat.

• This was only possible because we had the appropriate sequencers to handle the low depth of metagenomic samples.

• Mash is a versatile tool that may be useful for classifying organisms during the early stages of an outbreak.
Next Steps

• Continue to refine metagenomic approach.

• Apply what we learned from mpox to our SARS-CoV-2 pipeline.

• Explore the use of Twist Bioscience controls for validation purposes.
Acknowledgements

MDHHS Genomics Section
Dr. Heather Blankenship
Dr. Arianna Miles-Jay
Becky Kramer
Dr. Kristin Jacobs
Elizabeth Burgess
Steve Dietrich
Ellie Ewald

MDHHS Virology Section
Dr. Diana Riner
Dr. Katie Margulieux
Laura Mosher
Kevin Rodeman