Expanded sequencing of measles virus genomes for improved strain discrimination and molecular epidemiology

Sara B. Griesemer[1], Patrick Bryant[1], Tugba Yildirim[1], Dylan Johns[2], Rachel Wester[2], Martha Iwamoto[3], Bisram Deocharan[4], Jennifer Rakeman-Cagno[4], Noelle Bessette[5], Elizabeth Zaremski[5], Barbara Montana[5], Rebecca J. McNall[6], Bettina Bankamp[6], Daryl Lamson[1], Kirsten St.George[1]

[1] Laboratory of Viral Diseases, Wadsworth Center, New York State Department of Health, Albany, NY [2] Bureau of Immunization, New York State Department of Health, Albany, NY

[3] Bureau of Immunization, NYC Department of Health and Mental Hygiene, New York, NY [4] Public Health Laboratory, NYC Department of Health and Mental Hygiene, New York, NY

[5] Public Health and Environmental Laboratories, New Jersey Department of Health, Ewing, NJ [6] Division of Viral Diseases, Centers for Disease Control and Prevention, Atlanta GA USA

NEW YORK STATE OF OPPORTUNITY. | Department of Health | Wadsworth Center

BACKGROUND

October 2018: the largest measles outbreak in the US in 25 years began in NY State

September 2019: more than 800 cases have been confirmed in NY State, all D8 genotype

- Genetic characterization is useful for tracking importation and transmission pathways
- Current genotyping method uses sequencing of 450 nucleotides of the nucleoprotein gene (N-450)
- Extended sequencing, including the M-F non-coding region (MF-NCR) and whole genome (WGS) have been suggested for enhanced molecular epidemiology and improved phylogenetic resolution.
- In this study, the three methods were compared for their ability to differentiate measles strains.

METHODS

- Measles-positive specimens from New York State (NYS), New York City (NYC) and New Jersey (NJ), were retrieved from frozen storage
- Conventional RT-PCR and Sanger sequencing produced N-450 and MF-NCR sequence data **(Figure 1)**
- A subset of 29 samples were chosen for WGS, using two different methods:
  - **1)** CDC amplicon-based (kindly provided by CDC Division of Viral Diseases, adapted from Penedos AR et al, PLos One 2015)
  - **2)** Custom-designed AmpliSeq panel from ThermoFisher Scientific

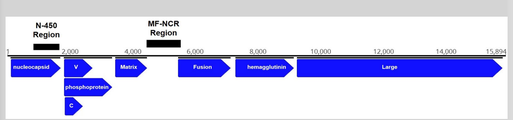

Figure 1: Measles virus genome showing the N-450 and MF-NCR regions sequenced for genotyping and molecular epidemiology.

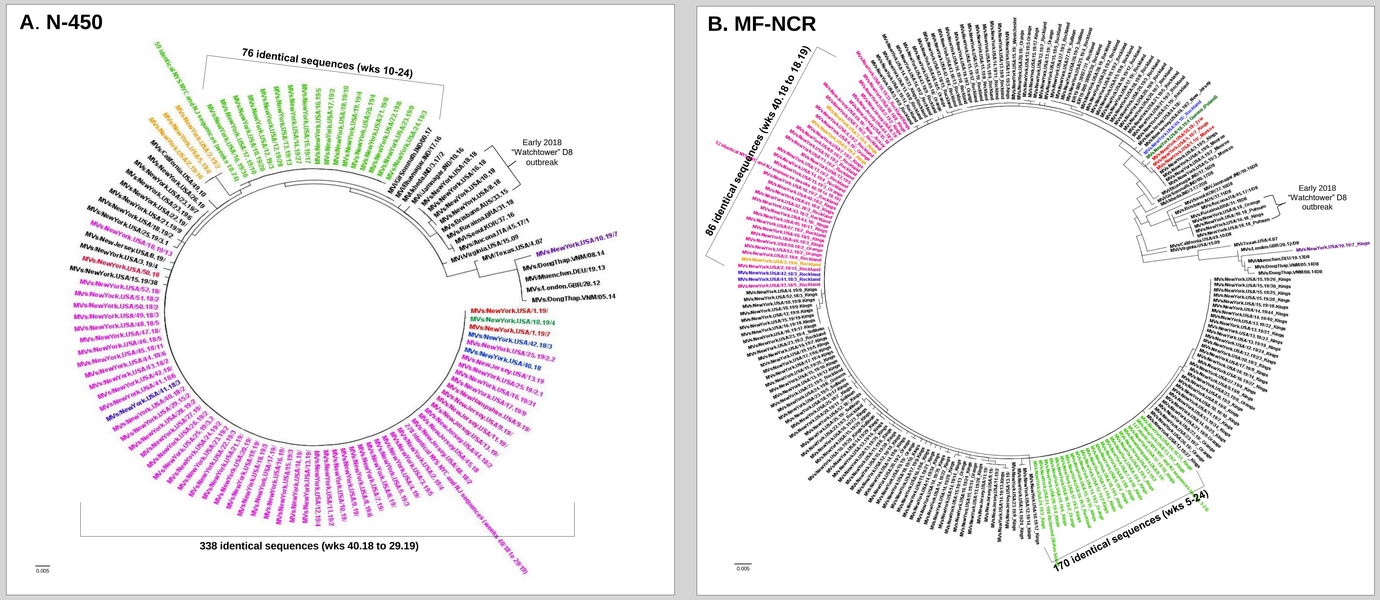

Figure 2. Phylogenetic analysis of N-450 (A) and MF-NCR (B) sequence data from >400 NYS, NYC and NJ measles samples collected during the 2018-2019 outbreak. Maximum likelihood trees were obtained using MEGAX with the GTR model and 500 bootstraps. ■ =identical sequences representing early to late outbreak samples. ■ =identical sequences representing mid-late outbreak samples. Index and known import cases are colored as in **Figure 3**.

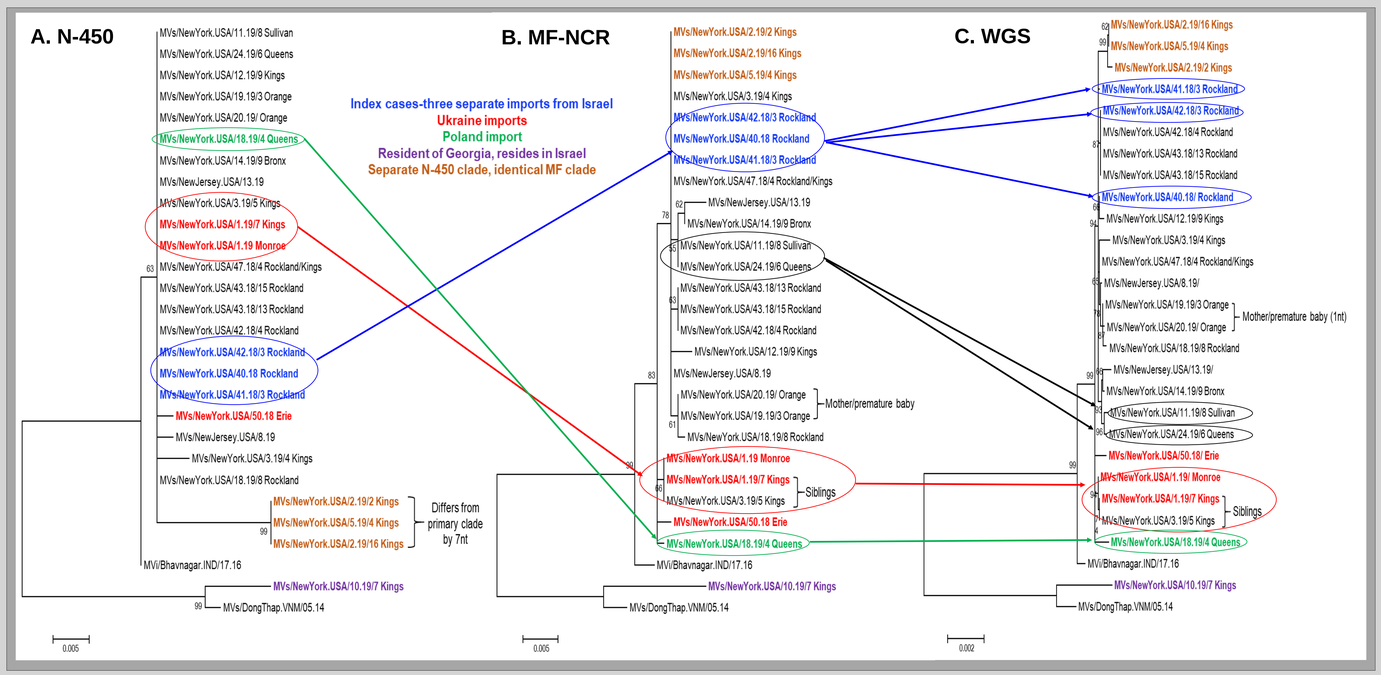

Figure 3. Phylogenetic analysis of 26 NYS, NYC and NJ measles viruses using A) N-450 B) MF-NCR and C) WGS. Maximum likelihood trees were obtained using MEGAX with the GTR model and 500 bootstraps.

RESULTS

- N-450 sequences from current outbreak samples were virtually identical and separated into two primary groups differing by only 1 nucleotide (**Figure 2A**)
- MF-NCR analysis showed higher resolution than N-450 genotyping, separating samples into multiple clusters and distinguishing strains imported from the Ukraine and Poland **(Figure 2 and 3**)
- 26 near complete genomes were obtained and compared to the N-450 and MF-NCR sequences (**Figure 3**)
- WGS increased the resolution of outbreak samples further than MF-NCR, separating cases into clusters not seen in either N-450 or MF-NCR trees (**Figure 3C**)

CONCLUSIONS

- N-450 sequencing can provide genotype, but does not differentiate between strains from imported cases and those from local transmissions
- MF-NCR analysis can be valuable for distinguishing new importations and may identify some lines of transmission
- WGS provides the highest resolution between strains for the analysis of phylogenetic relationships, to monitor importations and assess measles transmission
- Deeper epidemiological investigations are ongoing to further identify relationships between and within clusters of cases

ACKNOWLEDGEMENTS

The authors thank the Applied Genomics Technology Core for performing the MiSeq and Sanger sequencing, Gabe Solis for assisting with MF sequencing and analysis, and the ThermoFisher White Glove Team for assisting in the design and development of the AmpliSeq panel.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.