IMPLEMENTATION OF GHOST FOR HAV OUTBREAK DETECTION

Sumathi Ramachandran, MS, MPH, PhD

Molecular Epidemiology and Bioinformatics Team
Division of Viral Hepatitis
APHIL Annual Meeting, June 2018
Background

- Hepatitis A virus (HAV) is primarily a human pathogen, transmitted by person-to-person contact or ingestion of contaminated food or water.
- Hepatitis A vaccine given within 2 weeks of exposure can prevent the disease.
- In the USA, hepatitis A vaccination has been recommended for all children aged 12–23 months since 2006, and consequently the number of acute cases has greatly fallen.
- In recent years, large hepatitis A outbreaks have been observed in several states.
Objectives

- Determine the genetic characteristics of HAV circulating in the United States
- Provide phylogenetic background of HAV for outbreaks investigation and molecular tracking of hepatitis A in the United States
- Transition from traditional Sanger sequencing to NGS & GHOST system for hepatitis A outbreaks investigations
<table>
<thead>
<tr>
<th>HAV Specimen Source (1996-2018)</th>
<th># Positive / Tested Specimens</th>
<th># Unique Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentinel Counties (1996-2006)</td>
<td>1234/1510</td>
<td>404</td>
</tr>
<tr>
<td>EIP Sites (2007-2013)</td>
<td>472/755</td>
<td>309</td>
</tr>
<tr>
<td>Border Infectious Disease Sites (2000-2005)</td>
<td>378/450</td>
<td>154</td>
</tr>
<tr>
<td>Various Outbreaks (2000-2018)</td>
<td>2495/2913</td>
<td>426</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>4684/6044</strong></td>
<td><strong>1348</strong></td>
</tr>
</tbody>
</table>
# Genotype distribution of HAV circulating in USA

<table>
<thead>
<tr>
<th>HAV Specimens</th>
<th>HAV Genotype (%)</th>
<th>Total #</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IA (%)</td>
<td>IB (%)</td>
</tr>
<tr>
<td>Surveillance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996-2006</td>
<td>1573 (97.6)</td>
<td>34 (2.1)</td>
</tr>
<tr>
<td>2007-2018</td>
<td>480 (82.3)</td>
<td>92 (15.9)**</td>
</tr>
<tr>
<td>Sub-total</td>
<td>2053 (93.8)</td>
<td>126 (5.8)</td>
</tr>
<tr>
<td>Outbreaks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996-2006</td>
<td>600 (99.7)</td>
<td>1 (0.2)</td>
</tr>
<tr>
<td>2007-2018**</td>
<td>439 (23.2)</td>
<td>1453 (76.8)**</td>
</tr>
<tr>
<td>Sub-total</td>
<td>1039 (41.6)</td>
<td>1454 (58.3)**</td>
</tr>
</tbody>
</table>
HAV Genotype and Phylogenetic Sub-cluster

Legend
- US & BIDS
- Mexico
- Brazil
- Israel
- Spain
- Egypt & Jordan
- Europe & Japan
- Tunisia

International travel

US

Brazil
Moldova, Kazakhstan
Italy, Norway, Japan

IDU

MSM

Genotype I

Sub-genotype IIIA

Nucleotide variation

3%

2.0 %

Nucleotide Variation

HAV genotype and sub-genotype
# Rising HAV Genotype IB in USA

## HAV Outbreaks

<table>
<thead>
<tr>
<th>Year</th>
<th>IA (%)</th>
<th>IB (%)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>39 (56.5)</td>
<td>30 (43.5)</td>
<td>69</td>
</tr>
<tr>
<td>2015</td>
<td>26 (96.3)</td>
<td>1 (3.7)</td>
<td>27</td>
</tr>
<tr>
<td>2016</td>
<td>110 (46.2)</td>
<td>127 (53.4)</td>
<td>238</td>
</tr>
<tr>
<td>2017</td>
<td>217 (16.7)</td>
<td>1082 (83.3)</td>
<td>1298</td>
</tr>
<tr>
<td>2018</td>
<td>58 (12.2)</td>
<td>417 (87.8)</td>
<td>475</td>
</tr>
<tr>
<td>Total</td>
<td>450 (21.4)</td>
<td>1657 (78.6)</td>
<td>2107</td>
</tr>
</tbody>
</table>

**IA**

- 13NM
- 16MI
- A16VA2
- A16VA1

**IB**

- A17CA, UT & KY

**Middle East References**
Schematic Representation of the HAV Genome & Region Used for Molecular Analysis

- CDC SOP for HAV Molecular Analysis
  - Traditional Sanger Sequencing (Up to April 2018)
  - Next Generation Sequencing/ GHOST (Since May 2018)
Hepatitis A Outbreaks (2016 -2018)

(VP1/P2B region, 315 bp in length)

June 5th 2018 update

2300 Samples processed at CDC
1924 Samples were PCR positive

Lab Technical Assistance:
San Diego
CA
MI
FL

CA Clusters (A-D)

MI Clusters (1 & 2)
HAV Transmission Clusters (2016 -2018)

- CA Clusters
- CA Cluster <15 cases
- MI Cluster

CA Cluster A (multiple states, n=12)
CA Clusters B, C, D
MI Cluster 1
MI Cluster 2
Next-Generation Sequencing
Two Strategies for Viral Pathogens

Whole-Genome Sequencing
- NGS reads cover the entire genome
- Variable coverage of different genomic regions
- Reads stagger
- Consensus sequence obtained
- Heterogeneity at individual positions

Single-Amplicon Sequencing
- NGS reads cover a single region
- Sequences of individual variants
- Deep coverage
- Minority variants
- Reads strictly aligned
- Heterogeneity of the region
- Intra-host population structure
- Strong genetic strain identity
- Increased multiplexing
- Reduction in cost per specimen

HAV Genome

NGS of VP1-2A gene variants
- Dominant major variant
Clinical samples → Amplicon Deep Sequencing: Miseq (Illumina) → FASTQ/FASTA

**GHOST Portal**
(https://webappx.cdc.gov/GHOST)

- Upload
- Web-based Models
- Not linked
- Transmission cluster
- Source?

**Outbreak**

- Uploaded NGS files are automatically analyzed and the results are visualized in the form of transmission networks
- GHOST web-site contains bioinformatics tools for outbreak investigation and molecular surveillance
GHOST module for HAV Outbreak Investigation

A17KY clusters (N=314 tested):

<table>
<thead>
<tr>
<th>A17KY Cluster</th>
<th>Case # (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA IB Cluster A</td>
<td>227 (85.7)</td>
</tr>
<tr>
<td>CA IB Cluster B</td>
<td>1</td>
</tr>
<tr>
<td>MI IB Cluster 2</td>
<td>4</td>
</tr>
<tr>
<td>others IB</td>
<td>33</td>
</tr>
<tr>
<td>IA/IIIA</td>
<td>3</td>
</tr>
</tbody>
</table>

KY HAV cases
GHOST Networks

A17CA cluster A

A17KY, A17UT & 2A17CA Sequences Analysis
(YP19P2B region, 215 bp in length; updated 6/12/16)

A17CA Cluster A
HAV Intra-host variants - Kstep networks

A17KY early cases

A17CA Cluster A

A17KY recent cases

<table>
<thead>
<tr>
<th>Min Frequency = 10, Max Frequency = 6853</th>
</tr>
</thead>
<tbody>
<tr>
<td>shared</td>
</tr>
<tr>
<td>A17KY132_S16_L001_1b</td>
</tr>
<tr>
<td>A17KY166_S17_L001_1b</td>
</tr>
<tr>
<td>A17KY214_S2_L001_1b</td>
</tr>
<tr>
<td>A17KY245_S10_L001_1b</td>
</tr>
<tr>
<td>A17KY286_S24_L001_1b</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Shared Sequences</th>
<th>487</th>
</tr>
</thead>
<tbody>
<tr>
<td>A17KY18_S3_L001_1b</td>
<td>0</td>
</tr>
<tr>
<td>A17KY19_S4_L001_1b</td>
<td>0</td>
</tr>
<tr>
<td>A17KY23_S8_L001_1b</td>
<td>4</td>
</tr>
<tr>
<td>A17KY51_S6_L001_1b</td>
<td>0</td>
</tr>
<tr>
<td>A17KY63_S2_L001_1b</td>
<td>1</td>
</tr>
<tr>
<td>A17KY81_S21_L001_1b</td>
<td>0</td>
</tr>
</tbody>
</table>
## HAV Intra-host variants - Kstep networks

<table>
<thead>
<tr>
<th>MI Local ID</th>
<th>MI Sequences</th>
<th>Genotype</th>
<th>CDC_ID</th>
<th>CDC Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAV18-000016</td>
<td>MI Cluster 2</td>
<td>IB</td>
<td>A16MI526</td>
<td>MI Cluster 2</td>
</tr>
<tr>
<td>HAV18-000020</td>
<td>MI Cluster 2</td>
<td>IB</td>
<td>A16MI527</td>
<td>MI Cluster 2</td>
</tr>
<tr>
<td>HAV18-000021</td>
<td>MI Cluster 2</td>
<td>IB</td>
<td>A16MI528</td>
<td>MI Cluster 2</td>
</tr>
<tr>
<td>HAV18-0045</td>
<td>Unique</td>
<td>IB</td>
<td>A16MI529</td>
<td>MI Cluster 2</td>
</tr>
<tr>
<td>HAV18-0046</td>
<td>MI Cluster 2</td>
<td>IB</td>
<td>A16MI530</td>
<td>MI Cluster 2</td>
</tr>
<tr>
<td>HAV18-0050</td>
<td>MI Cluster 2</td>
<td>IB</td>
<td>A16MI531</td>
<td>MI Cluster 2</td>
</tr>
<tr>
<td>HAV18-0067</td>
<td>MI Cluster 2</td>
<td>IB</td>
<td>A16MI532</td>
<td>MI Cluster 2</td>
</tr>
<tr>
<td>HAV18-0073</td>
<td>MI Cluster 2</td>
<td>IB</td>
<td>A16MI533</td>
<td>MI Cluster 2</td>
</tr>
<tr>
<td>HAV18-001111</td>
<td>Negative</td>
<td>IB</td>
<td>A16MI534</td>
<td>MI Cluster 2</td>
</tr>
<tr>
<td>HAV18-001112</td>
<td>Negative</td>
<td>IB</td>
<td>A16MI535</td>
<td>Unique</td>
</tr>
<tr>
<td>HAV18-0121</td>
<td>MI Cluster 1</td>
<td>IB</td>
<td>A16MI536</td>
<td>MI Cluster 1</td>
</tr>
<tr>
<td>HAV18-0128</td>
<td>MI Cluster 2</td>
<td>IB</td>
<td>A16MI537</td>
<td>MI Cluster 2</td>
</tr>
<tr>
<td>HAV18-0129</td>
<td>MI Cluster 2</td>
<td>IB</td>
<td>A16MI538</td>
<td>MI Cluster 2</td>
</tr>
<tr>
<td>HAV18-0132</td>
<td>MI Cluster 2</td>
<td>IB</td>
<td>A16MI539</td>
<td>NGS in progress</td>
</tr>
<tr>
<td>HAV18-0139</td>
<td>MI Cluster 1</td>
<td>IB</td>
<td>A16MI541</td>
<td>MI Cluster 1</td>
</tr>
<tr>
<td>HAV18-0140</td>
<td>MI Cluster 2</td>
<td>IB</td>
<td>A16MI542</td>
<td>MI Cluster 2</td>
</tr>
<tr>
<td>HAV18-0154</td>
<td>MI Cluster 1</td>
<td>IB</td>
<td>A16MI543</td>
<td>MI Cluster 1</td>
</tr>
<tr>
<td>HAV18-0160</td>
<td>MI Cluster 2</td>
<td>IB</td>
<td>A16MI544</td>
<td>MI Cluster 2</td>
</tr>
<tr>
<td>HAV18-0162</td>
<td>MI Cluster 2</td>
<td>IB</td>
<td>A16MI545</td>
<td>MI Cluster 2</td>
</tr>
<tr>
<td>HAV18-0178</td>
<td>MI Cluster 2</td>
<td>IB</td>
<td>A16MI546</td>
<td>MI Cluster 2</td>
</tr>
<tr>
<td>HAV18-0183</td>
<td>Unique</td>
<td>IB</td>
<td>A16MI547</td>
<td>HAV18-0183</td>
</tr>
<tr>
<td>HAV18-0186</td>
<td>MI Cluster 2</td>
<td>IB</td>
<td>A16MI548</td>
<td>MI Cluster 2</td>
</tr>
</tbody>
</table>

The diagram illustrates the shared sequences with a total of 799 sequences, including:

- A16MI438_S2_L001: 1
- A17MI547_S23_L001: 190
- A16MI427_S3_L001: 4
- A16MI432_S3_L001: 4
- A16MI435_S6_L001: 3
- A17MI545_S21_L001: 4
- A17MI548_S24_L001: 3
GHOST Outreach Activities (2016 - ongoing)

GHOST HCV Pilot States
- TN
- NY
- NH
- MI
- AK

GHOST Trained States (22 states)
- 2016 & 2017 Workshops

GHOST HAV Testing sites
- CA & SD County
- IN
- MI
- NY

CDC-NIH-NIDA-SAMHSA
- Rural Opioid FOA (10 states)

GHOST Center
- MA (Harvard University, 2018)

CDC-APHL
- Phase 1 (2016 - ongoing)
- Phase 2 (2017- ongoing)

GHOST HCV Pilot States
- TN
- NY
- NH
- MI
- AK

GHOST HCV Implementation
- IA
- IN
- CO
- GA
- MT

(2018)
Conclusions

- In the 2016-2018 HAV outbreaks, genotype 1B was widely occurring among drug users/homeless, and genotype 1A among MSM.
- NGS sequences represent complex intra-host HAV populations.
- NGS sequencing coupled with GHOST eliminates the effects of population complexity and many technological artifacts on assessment of transmission links during outbreak investigations.
- GHOST module for a more accurate identification of HAV strains in surveillance is in development.
- Use by state labs is expected to help develop an improved GHOST module applicable to identification of HAV strains in both outbreak and surveillance settings.
Acknowledgments

- CDC Division of Viral Hepatitis
  - Jeff Efird
  - Monique Foster
  - Megan Hofmeister
  - Eyasu Teshale
  - Noele Nelson
  - Cynthia Jorgensen
  - Brigette Ulin
  - Blythe Ryerson
  - Cecily Campbell
  - Philip Spardling
  - Alaya Koneru
  - Tonya Hayden
  - Saleem Kamili
  - John Ward
  - Paul Weidle

- County of San Diego Health and Human Services Agency
- California Department of Public Health
- Utah Department of Health
- Kentucky Department of Health
- Michigan Department of Health and Human Services
- All other State Departments of Health
- APHL

- Viral Hepatitis Laboratory (GHOST)
  - Yulin Lin
  - Lili Punkova
  - Guo-Liang Xia
  - Garrett Longmire
  - Seth Sims
  - Walker Gussler
  - Amanda Poe
  - Yury Khudyakov
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.