Viral Metagenomics

Lori Holtz, MD
Pediatric Gastroenterology

Disclosures

• No conflicts of interest to disclose
What is “Metagenomics”? 

• The “Metagenome”  
  – Handelsman et al Chemistry and Biology (1998)  
  “…..advances in molecular biology and eukaryotic genomics, which have laid the groundwork for cloning and functional analysis of the collective genomes of soil microflora, which we term the metagenome of the soil.”

• Metagenomics  
  – Aggregate nucleic acid sequences in a sample  
  – Studies in the absence of specific culture or targeting  
  – “Whole community genome analysis”  
  – Bacterial  
  – Viral

Defining our metagenome

• Stool metagenomics

• Human microbiome project

• Stool viral metagenomics (stool virome)
Uses of metagenomics

• Pathogen discovery (Part 1)
• Define our microbial communities (Part 2)

Why Search for Novel Viruses?

1. Many unexplained diseases
2. Novel insights from novel viruses
Existing Viral Detection/Discovery Methods Have Limitations

- Viral culture
  - Many viruses not culturable, specific cell types needed
- Electron Microscopy
  - High titers, many viruses with similar morphology
- Immunoassays
  - Require special reagents, candidate viruses
- PCR
  - Limited breadth (candidate viruses)

- Solution: Develop comprehensive, unbiased, and high-throughput methods.
- Metagenomics offers key

Why study diarrhea?

Leading Causes of Death Due to Infectious Diseases

<table>
<thead>
<tr>
<th>Disease</th>
<th>Deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Respiratory Tract Infections</td>
<td>3.46 million</td>
</tr>
<tr>
<td>Diarrheal Diseases</td>
<td>2.46 million</td>
</tr>
<tr>
<td>HIV/AIDS</td>
<td>1.78 million</td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>1.34 million</td>
</tr>
</tbody>
</table>

A significant fraction of diarrhea cases are of unexplained etiology

4 Major viral causes of diarrhea:

- Rotavirus
- Norovirus
- Adenovirus
- Astrovirus

~40% of diarrhea is of unknown etiology

Are these unexplained cases due to novel viruses?
Could complex viromes be responsible for these unknown cases?

How do we identify novel viruses?

1. Filter sample
2. Extract nucleic acid
3. Randomly amplify by PCR
4. Next Generation Sequence
5. Custom bioinformatics pipeline

- Eukaryotic Viruses
- Bacteria
- Host
- Fungi
- Phage
- Other
- Unassigned

All viral reads

Novel
Known
Computational Pipeline

1. All the sequences from one sample
2. Quality control
3. BLASTn against Virus-only nucleotide database
   - E value < 10^-10
4. BLASTx against Virus-only protein database
   - E value < 10^-3
5. All potential virus sequences

Pipeline (continued)

1. All potential virus sequences
2. MegaBLAST vs. NCBI nt DB
   - E value < 10^-10
3. BLASTn vs. NCBI nt DB
   - E value < 10^-10
4. BLASTx vs. NCBI nt DB
   - E value < 10^-3
5. Classify using NCBI Taxonomy DB
   - Bacteria, Fungus, Virus, Phage, Other, Ambiguous, Host
Samples for viral discovery

Pediatric diarrhea
  – Negative for known pathogens
  – Global collaborations:

Evidence for many novel viruses in diarrhea specimens

• Prior to 2005 the rate of viral discovery was slow
• 1975-2005: <10 viruses discovered associated with diarrhea
• >1000 diarrhea samples sequenced from around the world
• Many novel viruses:
Evidence for many novel viruses in diarrhea specimens

<table>
<thead>
<tr>
<th>Novel virus</th>
<th>Viral family</th>
<th>RNA or DNA virus</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Astrovirus MLB1</td>
<td>Astroviridae</td>
<td>RNA</td>
<td>Finkbeiner et al, 2008</td>
</tr>
<tr>
<td>Astrovirus VA1</td>
<td>Astroviridae</td>
<td>RNA</td>
<td>Finkbeiner et al, 2009</td>
</tr>
<tr>
<td>Astrovirus MLB3</td>
<td>Astroviridae</td>
<td>RNA</td>
<td>Jiang et al, 2013</td>
</tr>
<tr>
<td>Astrovirus VA4</td>
<td>Astroviridae</td>
<td>RNA</td>
<td>Jiang et al, 2013</td>
</tr>
<tr>
<td>Astrovirus VA5</td>
<td>Astroviridae</td>
<td>RNA</td>
<td>Meyer et al, 2015</td>
</tr>
<tr>
<td>Cosavirus E1</td>
<td>Picornaviridae</td>
<td>RNA</td>
<td>Holtz et al, 2008</td>
</tr>
<tr>
<td>Klassevirus</td>
<td>Picornaviridae</td>
<td>RNA</td>
<td>Holtz et al, 2009</td>
</tr>
<tr>
<td>Rosavirus 2</td>
<td>Picornaviridae</td>
<td>RNA</td>
<td>Lim et al, 2014</td>
</tr>
<tr>
<td>Novel Amdovirus</td>
<td>Parvoviridae</td>
<td>DNA</td>
<td>Unpublished</td>
</tr>
</tbody>
</table>

Flowchart Following Novel Virus Discovery

Screen Samples for Novel Viruses

→ Novel Virus Identified

Epidemiologic Characterization
- Diagnostic assays
- Prevalence
- Case-control association studies
- Seroepidemiology
- Disease association

Virologic Characterization
- Genome sequencing/comparisons
- Characterize individual proteins (e.g. transformation potential, toxin activity)
- Culture virus
- Animal models
Astrovirus background

• Known cause of human diarrhea

• 1 human species (1975)
  – Eight closely related serotypes

• Extent of infection in human populations varies greatly by serotype
  – Astrovirus serotype 1 (90%)
  – Astrovirus serotype 7 (10%)

• We have discovered 8 new astroviruses

8 novel astroviruses found in human diarrhea

Meyer et al, 2015
Discovery of novel astrovirus MLB1

• Stool sample:
  – collected in 1999
  – 3 yr old boy
  – acute diarrhea
  – Melbourne, Australia
  – negative for known pathogens

Characterization of novel viruses

Determine spectrum of viruses in sample

↓

Characterization of novel virus(es)

Epidemiological studies of novel viruses

Is the virus associated with disease?

Molecular studies of novel viruses

How frequently is the virus found in the population?

How prevalent are antibodies to the virus?
Astrovirus MLB1 is present globally

Overall prevalence is in diarrhea is 1-2%

What are the potential roles of MLB1?

• Agent of diarrhea?
• Dietary finding?
• Is a pathogen outside of the GI tract, but is shed in stool (i.e. polio)?
Is MLB1 associated with diarrhea?

- Volunteer ingestion is no longer allowable
- 2\textsuperscript{nd} best alternative is case control study
  - India
  - Africa

Stool samples: cases and controls

- 3 year longitudinal birth cohort study (n=370)
- 27,000 surveillance samples (controls)
- 1800 diarrhea samples (cases)

- Samples cannot leave India

Gagandeep Kang
Each subject serves as their own control

- 27,000 controls and 1800 diarrhea samples available
- Power calculation: 400 diarrhea samples and 400 controls

At least 6 weeks between diarrhea and control sample

Results of case control study of astroviruses

<table>
<thead>
<tr>
<th></th>
<th>Classic Astrovirus Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases (n=400)</td>
<td>14</td>
</tr>
<tr>
<td>Controls (n=400)</td>
<td>4</td>
</tr>
<tr>
<td>p-value (McNemar)</td>
<td>0.03</td>
</tr>
</tbody>
</table>

2\textsuperscript{nd} MLB1 Case Control Study

- Samples collected from the Global Enteric Multicenter Study (GEMS)
  - Field sites: Kenya and The Gambia
- Cases: moderate-severe diarrhea
  - $\geq 3$ loose stools in prior 24 hours
  - New episode (onset after $> 7$ diarrhea free days)
  - +1 of following: sunken eyes, loss of skin turgor, IV hydration prescribed, blood in stools or admission to the hospital due to diarrhea
- Controls:
  - Excluded if episode of diarrhea in prior 7 days

\textit{O. Colin Stine}

Results of astrovirus case control study

<table>
<thead>
<tr>
<th></th>
<th>Kenya</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases n=181</td>
<td>Controls n=181</td>
<td>p-value</td>
</tr>
<tr>
<td>HAstV</td>
<td>$7$ (3.9%)</td>
<td>$6$ (3.3%)</td>
<td>$p=0.327$</td>
</tr>
</tbody>
</table>

- Classic astroviruses not associated with diarrhea
- MLB1 is associated with diarrhea in Kenya

\textit{Meyer et al, 2015}
Case control studies fail these pathogens

- *Campylobacter*
- *Giardia*
- Diarrheagenic *E.coli*

Does MLB1 have a role in human disease?

- Agent of diarrhea?
- Protective effect?
- Dietary finding?
- Has a role outside of the GI tract, but is shed in stool (i.e. polio)?
Characterization of novel viruses

Determine spectrum of viruses in sample

Characterization of novel virus(es)

Epidemiological studies of novel viruses

Molecular studies of novel viruses

How frequently is the virus found in the population?

Is the virus associated with disease?

How prevalent are antibodies to the virus?

What is needed to determine seroprevalence of MLB1?

• Develop assay:
  – Indirect ELISA
  – Capture antigen (protein of interest)
    • Recombinant capsid protein
      (Dr. Schultz-Cherry)
  – Positive and negative controls
    • Pre and post immune rabbit serum

• Human sera to test
  – 400 age stratified healthy volunteer serum samples (Dr. Belshe)
First evidence of human antibody to MLB1

100% seropositivity by adulthood!

Holtz et al, Clin Vacc Immunol, 2014

Summary (Part 1)

• Viral metagenomics can be used to identify novel viruses
• 8 new astroviruses discovered
• MLB1 present globally
• MLB1 may be associated with diarrhea
• Humans develop antibodies to MLB1, a ubiquitous infection
Uses of metagenomics

- Pathogen discovery (Part 1)
- Define our microbial communities (Part 2)

Why do we care about metagenomics?

- Traditional thought: 1 organism=1 disease
- Change in paradigm: Metagenome role in disease
- Bacterial Stool Microbiome: Changes with Disease
- Dysbiosis=shift in relative abundance of bacterial taxa

![Pie charts showing changes in bacterial taxa in healthy control vs patient] by Spor et al, 2011
Is there an enteric virome?

• When do viruses populate the human gut?
• Do these populations change over time?
• Does the virome vary between people?
• What factors affect the virome?

Why study the infant enteric virome?

• Acquisition and development of gut bacterial microbiome thought to affect host later outcomes:
  – Growth and habitus
  – Infection resistance
  – Inflammatory and metabolic disorders
  – Maturation and function of immune system
• Does the enteric virome also have similar consequences?
Why study the enteric virome?

• Enteric virome is a source for potential pathogens
• Healthy individuals are chronically infected with ~10 viruses (herpes, polyoma, papilloma, etc)
  – These affect our innate and adaptive immunity
  – Animal models show protection against bacterial infections
  – Unknown if chronic enteric viruses
• Viruses interact with bacteria
  – Gut bacteria enhance enteric eukaryotic virus replication
  – Bacteriophage modulate bacterial pathogenicity, antibiotic resistance, and metabolism

Challenges to studying the enteric virome

• Virome=eukaryotic viruses + bacteriophage

• Unlike studying the bacterial microbiome there is no conserved sequence across viruses
• Virome sequencing to date largely biased – RNA portion ignored
• Majority of sequences in purified viral samples defy annotation:
  – Not related to currently identified virus
  – Have rapidly evolved from known viruses
  – Inadequate reference databases
Limited understanding of the infant virome

- Two prior studies on eukaryotic virome in healthy infants:
  - PCR based study of two infants starting at 4 months of age- near constant shedding different viruses (Kapusinszky et al, 2012)
  - Sanger sequencing (477 reads) of DNA viral community in one stool (Breitbart et al, 2008)
- One prior study on DNA phage virome in healthy infant
  - One infant’s stool at four time points suggested that phage virome changes quickly in the first week of life and then stabilizes (Breitbart et al, 2008)

Enteric microbiome and virome in adults

- Human microbiome dynamic until age 3 when it stabilizes
  - Influenced by nutrition, delivery route, geography, and antibiotic use
- DNA phage virome:
  - Significant diversity between individuals and temporally stable (Reyes et al, 2010)
- DS DNA virome:
  - Viruses detected in stools of healthy adults (Wylie et al, 2014)
Question

• What is the profile of healthy human microbial maturation?

Cohort

• Warner and Tarr: STL Twin Study
• 4 pairs of twins (=8)
• Gestational age >35 weeks
• Delivery route: 3 C/S, 1 Vaginal
• Zygosity: 3 Dizygotic, 1 Monozygotic
• Race: 3 Caucasian, 1 African American
Microbiome and comprehensive virome of 4 pairs of healthy infant twins over 2 years

First comprehensive view of infant virome
Infant viromes are more similar within twin pairs

Eukaryotic virome expands with age

$R^2 = 0.192$
Parechovirus and Enterovirus most frequently detected RNA eukaryotic viruses

Near identical parechovirus strains shared within twin pairs

<table>
<thead>
<tr>
<th>Sample</th>
<th>Sequencing reads</th>
<th>qRT-PCR copy no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1-6</td>
<td>2,10</td>
<td>10,110</td>
</tr>
<tr>
<td>B2-6</td>
<td>5,299</td>
<td>12,058</td>
</tr>
<tr>
<td>C1-3</td>
<td>307</td>
<td>26,191</td>
</tr>
<tr>
<td>C2-3</td>
<td>1,246</td>
<td>72,328</td>
</tr>
<tr>
<td>A1-3</td>
<td>31</td>
<td>2,178</td>
</tr>
<tr>
<td>A2-3</td>
<td>0</td>
<td>730</td>
</tr>
<tr>
<td>D2-12</td>
<td>0</td>
<td>675</td>
</tr>
<tr>
<td>D1-12</td>
<td>0</td>
<td>15</td>
</tr>
</tbody>
</table>
Anelloviruses are the most frequently detected DNA eukaryotic viruses

Anelloviridae

Circular ssDNA virus
Genome: 2.8 – 4.0kb

Alphatorquevirus (TTV)
Betatorquevirus (TTMV)
Gammatorquevirus (TTMDV)

Highly ubiquitous in humans and mammalian species

Anelloviridae expansion observed upon immunosuppression, and increased prevalence in pediatric patients with fever
Anelloviruses are diverse and can persist

Anelloviruses peak at 6-12 months

* $p < 0.05$
** $p < 0.01$
Bacteriophage richness and diversity decreases with age

Increase in Microviridae abundance during development
Do Changes in Phage Correlate with Changes in Bacterial Microbiome?

• First need to define the bacterial microbiome in these samples

Bacteria Microbiome Richness and Diversity Increases with Age
Bacterial Microbiome Shifts with Age

Phage and Bacteria Have Inverse Relationship
Summary (Part 2)

- Eukaryotic virome acquired over time
- Twins share similar viromes and strains
- Immense diversity of anelloviruses which peaks at 6-12 months
- Rich phage virome at birth that contracts towards *Microviridae*
- Bacterial microbiome expands in infancy
- Propose a predator-prey model between phage and bacteria in infant gut
## Characteristics of Cohort

<table>
<thead>
<tr>
<th>Age (days)</th>
<th>0</th>
<th>200</th>
<th>400</th>
<th>600</th>
<th>800</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Male</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stool sampling</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Solid food</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fever/vomiting/diarrhea</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Antibiotic use</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Zygosity
- **Dizygotic**
- **Monozygotic**

### Delivery mode
- Cesarean section
- Vaginal

### Breastmilk vs. Formula
- Breastfed only
- Formula only
- Breastfed and formula