Population Genetics Tools Reveal Disease Status and Treatment Outcomes

Timothy F. Kowalik, Ph.D.
Microbiology and Physiological Systems
UMass Medical School
Post-Session Community Engagement Forum

Shuttle buses leave Wed at 7:00 am
last bus departs at 7:15 am
breakfast provided on site*

CTA Architects Engineers
3601 South Congress Avenue
Building C
Austin, TX

Can join us even if you did not register yet!

*Yummy breakfast tacos!!!
0.5% – 0.7%

US Congenital CMV infection rate
1:150

US Congenital CMV infection rate

*CMV is the leading cause of infection-associated birth defects*

*CMV, the birth defect virus*

*CMV, the stealth virus*

*CMV is also the leading cause of juvenile nonfamilial sensorineural hearing loss (SNHL)*
Cytomegalovirus (CMV)

- The most complex human virus
  - ~235,000 bp DNA
  - >200 open reading frames (genes)
    - May be over 700!
    - Encodes miRNAs, IncRNAs

- Infection
  - Primary infection usually via mucosal surfaces
  - Spread through epithelial cells to fibroblasts and endothelial cells
  - Monocyte/macrophage infection
    - Source of in vivo dissemination
    - Latency (CD34+ stem cells and lineage)
  - Capable of infecting most cell types in vivo
  - Reinfection of “immune” individuals appears to be common
Congenital CMV (cCMV) Infection

- Active CMV replication during pregnancy
  - Plasma DNA (low)
  - Antibody
- Virus crosses placenta
- CMV invades tissues throughout the fetus
  - Including CNS
- Sustained viral replication and shedding
  - Months to years

- Viral DNA loads are often not informative
CMV, a genetically complex virus

Can an understanding of CMV genomics:

- provide a better understanding of in vivo infections and
- reveal therapeutic/vaccination opportunities?
Genotyping, the “go-to” method for sequencing regions of large genomes

Genotype analysis of cytomegalovirus in congenitally infected infants

40% of pts, >5 genotypes!

“CHIMES” study

Genotyping, the “go-to” method for sequencing large genomes

<table>
<thead>
<tr>
<th>TRl</th>
<th>UL</th>
<th>UL/b’</th>
<th>US</th>
</tr>
</thead>
</table>

[Diagram showing the distribution of markers across different regions]
Genotyping, the “go-to” method for sequencing large genomes

Breadth of published CMV genotypes ~5% coverage (40 papers)
Depth ~20-40 sequences/sample

What new insights would be revealed if we sequence the entire CMV genome?

What could we learn by sequencing many of the CMV genomes in a person?
Genome-wide coverage using deep sequencing

Breadth of published CMV genotypes ~5% coverage (40 papers)
Depth ~20-40 sequences/sample

Sequence coverage, urine specimens from three neonates
Two types of data emerge from deep sequencing

- **Consensus sequence**
  - Sequential listing of the most common nucleotide
    - “AgggCTTgCAAAg...” (similar to shotgun Sanger sequencing)

- **Sequence of the viral population**
  - Align to the consensus sequence, ID mismatches, insertion/deletions
  - Requires error correction to call true polymorphisms with high confidence

- **Deep sequencing does not provide:**
  - Linkage information of alleles
  - Measures of epistasis
Samples analyzed by 2015

<table>
<thead>
<tr>
<th>Specimen Type</th>
<th>Number of Specimens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urine</td>
<td>32</td>
</tr>
<tr>
<td>Amniotic Fluid</td>
<td>1</td>
</tr>
<tr>
<td>Plasma</td>
<td>6</td>
</tr>
<tr>
<td>Cord Blood</td>
<td>2</td>
</tr>
<tr>
<td>Saliva</td>
<td>6</td>
</tr>
<tr>
<td>GBM</td>
<td>1</td>
</tr>
</tbody>
</table>

Patients from the US, Europe, Brazil

<table>
<thead>
<tr>
<th>Sequencing summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Number of Patients</td>
</tr>
<tr>
<td>Total Number of Specimens</td>
</tr>
<tr>
<td>Total Number of HCMV Bases</td>
</tr>
<tr>
<td>Average coverage</td>
</tr>
<tr>
<td>Mean Depth</td>
</tr>
<tr>
<td>Total Number of SNPs Detected</td>
</tr>
<tr>
<td>Total Number of Unique SNPs</td>
</tr>
<tr>
<td>Total Number of Polymorphic Sites</td>
</tr>
<tr>
<td>Tri- or Quad-Allelic Sites (%)</td>
</tr>
</tbody>
</table>

~4,000~44,000 SNPs/sample*

75% of the cCMV genome is diverse may have implications in DNA-based diagnostics

*“standing genetic variation”
HCMV populations are as diverse as RNA virus populations

Diagram:

- EBV
- HHV-6A
- HHV-6B
- WNV
- HCMV
- DENV
- HCV

Scale:
0.1 % Intrahost Diversity

Renzette Curr Opin Virol 2014
J. Virol 2014
Bhattacharjee, submitted
Preexisting drug resistance alleles are common in cCMV!?
cCMV sequences differ by tissue compartment

Temporal and spatial compartmentalization of drug-resistant CMV in a child with CMV meningoencephalitis: implications for sampling in molecular diagnosis.
Franege et al, 2013. *J Clin Micro*
cCMV populations differentiate by compartment
Temporal changes in SNPs: largely stable...
Temporal changes in SNPs: largely stable...

>20,000 SNPS
Temporal changes in SNPs: largely stable...

>20,000 SNPS
Temporal changes in SNPs: largely stable...
...until treated with antivirals...
...and viral loads rebound.

Maternal, CMV(pp65/IE1) stimulated T cell infusion
...and viral loads rebound.

CMV evolution can be rapid and dynamic

Presumably due to standing genetic variation
HCMV loads are not always clinically informative

All samples (whole blood, plasma, urine)

Viral loads over time

Urine only

Viral Load (copies/mL)

Day of Sampling

In asymptomatic patients, viral load remains low and stable over time. In symptomatic patients, viral load may increase during reactivation and then decrease during a symptomatic episode. Extinction of viral load indicates resolution of infection.
Two ways to measure population size

**Census size:**
number of individuals in a population

**Effective population size**
number of parents generating offspring

Typically, Census size >> Effective pop size

Can $N_e$ inform clinical disease/treatment outcomes?

Selection coefficients, bottleneck sizes & timing estimated from modeling of infection history
Effective population size ($N_e$) & clinical outcomes
Effective population size ($N_e$) & clinical outcomes

Patient treated, Recovered (extinction)
Patient treated, symptomatic
Patient not treated, asymptomatic
Patient not treated, viral population rebounds

Biomarker for disease management?
Effective population size ($N_e$) & clinical outcomes

Patient treated, Recovers (extinction)

Patient not treated symptomatic

Patient treated, viral population rebounds

Patient not treated asymptomatic

Biomarker for disease management?
Selection versus Demography

Selection

Bottleneck-Expansion
Selection coefficients \((s)\), a measure of fitness

- \(s\) is estimated for every allele in a population
- A "distribution of fitness effects" is determined by curve fitting the plotted \(s\) values
- \(s=0\) is neutral, i.e., no fitness advantage or disadvantage
Selection coefficients \((s)\), a measure of fitness

- **Asymptomatic**
- **Symptomatic, no treatment**
Selection coefficients \((s)\), a measure of fitness

![Graphs showing distribution of allele fitness effects for different scenarios: Asymptomatic, Symptomatic, no treatment, and Symptomatic, treated.](image)
Selection versus Demography

Selection

Bottleneck-Expansion
Demography

Bottleneck-Expansion

Guntenkunst, Diffusion approximation dadi.googlecode.com
Bottlenecks associated with fetal infection & kidney colonization

Fetal Infection: Bottleneck (50-150 genomes) at 13-18 weeks gestational age

Kidney Infection: Bottleneck (10-20 genomes) at 24-29 weeks gestational age

Relatively large transmission bottleneck

Renzette PLoS Gen 2013
Bottlenecks are rational therapeutic targets
Bottlenecks are rational therapeutic targets

Can inferred dating of fetal infection be used to exclude false negatives?
In sum...

• DNA viruses can be as diverse as RNA viruses in intrahost infections
  — Key is large founder sizes & resultant elevated standing genetic variation

• CMV populations can rapidly differentiate when colonizing new compartments.
  — Population differentiation appears to be driven bottlenecks, positive selection, other mechanisms?

• Intrahost CMV is significantly different from secreted virus.
  — Implications for transmission, vaccine design, and monitoring strategies

• Popgen modeling reveals clinically significant information about viral infections including treatment outcome.
Acknowledgements

Kowlab members:

Nicholas Renzette
Xiaofei E
SeongAe Kwak
Bornali Bhattacharjee
Alexander Lagadinos
Joel O’Bryan
Pallavi Gandhi

Ecole polytechnique fédérale de Lausanne
Jeff Jensen and lab members
University of Massachusetts Medical School
Laura Gibson
University of Minnesota Medical School
Mark Schleiss
University of Alabama School of Medicine
Bill Britt
Suresh Roppala

Most importantly—CMV Parents and cCMV Infants & Children

Maria Massi Pimentel
Aparecida Y. Yamamoto
University Hospital of Tuebingen
Klaus Hamprecht
Faculté de Médecine-Université de Limoges
Sophie Alain
Children’s National Health System
Michael Keller
Catherine Bollard
National Institute of Allergy and Infectious Diseases
Jeffrey Cohen
Princeton University
Tom Shenk