Composition and dynamics of the human virome

Frederic Bushman
The global virome

- $10^7$ viruses per ml in sea water
- Viruses outnumber hosts by ~10-fold in sea water
- $10^{31}$ viral particles on Earth
- Numerically most successful biological entities

Data from Lita Proctor, Forrest Rohwer, Curtis Suttle and others

Proctor, 1997
# The Human Virome

## Persistent/latent infections

<table>
<thead>
<tr>
<th>Virus</th>
<th>% of population seropositive</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBV</td>
<td>100%</td>
</tr>
<tr>
<td>VZV</td>
<td>95%</td>
</tr>
<tr>
<td>Herpes</td>
<td>80%</td>
</tr>
<tr>
<td>HSV1</td>
<td>68%</td>
</tr>
<tr>
<td>Papilloma</td>
<td>60%</td>
</tr>
<tr>
<td>CMV</td>
<td>59%</td>
</tr>
<tr>
<td>HSV2</td>
<td>22%</td>
</tr>
<tr>
<td>HIV</td>
<td>1%</td>
</tr>
<tr>
<td>HCV</td>
<td>1%</td>
</tr>
</tbody>
</table>

## Transient infections with animal cell viruses

- Endogenous retroviruses: 8% of human DNA

## Bacteriophage predators of bacteria and archaea

- 10^{10}-10^{11} per gram of stool
Virome analysis by deep sequencing

Find new pathogens
Watch viral evolution
Characterize viral DNA integration into genomes
Characterize complex uncultured communities

Solexa/Illumina HiSeq

Rohwer, Suttle, Lipkin, Hahn, Weinstock, Storch, Wang, Virgin, Gordon, Reyes, Proctor, Bushman many others
What is the composition of the human gut microbiome, and how does it change over time?

Support: HMP demonstration project “Diet, Genetic Factors, and the Gut Microbiome in Crohn’s Disease”. PIs Wu, Lewis and Bushman


Methods for virus purification from stool

Homogenize

Spin down solid waste

Filter at 0.2um to remove cells

Isopycnic centrifugation (CsCl) to isolate 1.35-1.7g/mL

Chloroform to rupture membranes

Degrade unprotected DNA (DNase)

Digest proteins to remove capsid

Column purify DNA (DNeasy)

Adapted from Thurber, et al. Nat. Protocols (2009)
Solexa/Illumina HiSeq data
12 subjects, ~40 BILLION bases

~500-1000 types per individual

Most sequence new

Individuals show little or no resemblance to each other

One of 12 subjects showed strongly-supported eukaryotic cell virus

Minot et al., PNAS, 2012
De Novo analysis of gene types discloses cassettes

Only 25% of phage ORFs have matches to database ORFs (permissive threshold)

Compare genes for similarity within the data set without reference to previous annotation. 58% have at least one match (30% identity)

Check conservation of gene order and orientation

Reveals 28 types of cassettes containing from 2-8 protein coding families.

Mean proportion of contigs covered by cassettes is 27%

Minot et al., PLoS ONE, 2012
Gene content in bacterial versus viral communities

Viruses are parasites

Sam Minot et al., Genome Res. 2011
How does the human gut virome change over time?

- Examine change over 2.5 years by dense time-series analysis of a single individual
- Purified DNA viruses
- 57 billion base pairs of sequence (Illumina HiSeq), assemble with deBruijn Graph method
- Also HiSeq shotgun data on total stool DNA
- 478 virome contigs, average of 82-fold sequence coverage

Minot et al., PNAS 2013
Viral community membership and persistence

Sixty circular assemblies, suggestive of completed sequences
Most lineages persist over 2.5 years
Rarefaction suggests saturation (by this method)

Minot et al., PNAS 2013
Deeper analysis combining Illumina and PacBio reads

Acquired 138 Mb of single molecule sequencing data. Only 30% overlap between Illumina and PacBio data sets.

PacBio contig links two Illumina contigs

Illumina contig links several PacBio contigs

Improved coverage and assembly combining Illumina and PacBio reads
Accumulation of base substitutions:
Rapid evolution of Microviridae in the human gut

• Some Microviridae species separated by as little as 3.1% substitution
• Evolution of novel bacteriophage species over the 2.5 years studied?

Minot et al., PNAS 2013

Up to 4% substitution
Longitudinal changes associated with CRISPRs

Six viral contigs targeted by bacterial CRISPRs
As many as 27 repeats targeting a single viral contig

Possible viral escape mutant

Change associated with possible escape mutation
Change associated with spacer substitution in viral CRISPR arrays
Variation associated with Diversity Generating Retroelements (DGRs)

Reverse transcriptase mediates targeted hypervariation in DNA phage genomes
Hypervariable loci in gut bacteriophages

~100 bp regions where nearly every read is different from every other read
Associated with an abundant class of reverse transcriptases

Minot et al., 2012
## Activity of Diversity Generating Retroelements

<table>
<thead>
<tr>
<th>Contig with DGR</th>
<th>Significant change over 2.5 years?</th>
<th>ORF Length</th>
<th>mutagenized domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>d23-6_1160627</td>
<td>No</td>
<td>381</td>
<td>Clec (MTD)</td>
</tr>
<tr>
<td>38</td>
<td>No</td>
<td>381</td>
<td>Clec (MTD)</td>
</tr>
<tr>
<td>42</td>
<td>Yes</td>
<td>351</td>
<td>Clec</td>
</tr>
<tr>
<td>d03-_1820004</td>
<td>No</td>
<td>603</td>
<td>Clec (MTD)</td>
</tr>
<tr>
<td>166</td>
<td>No</td>
<td>592</td>
<td>Ig-superfamily</td>
</tr>
<tr>
<td>90</td>
<td>No</td>
<td>365</td>
<td>Clec (MTD)</td>
</tr>
</tbody>
</table>

- Compare within time point diversity to between time point diversity.
- Compare distances between communities over short time points and long time points, look for significantly greater distances over longer time points.
- Only one DGR clearly active.
- What about others? Activity down-regulated?
Why do humans harbor such huge viral populations?

Why are humans so different from each other?

Bacterial hosts of bacteriophages differ among individuals.

At least some of the phage are changing really fast (steady substitution, DGRs, CRISPR, etc).
Gaps, needs and challenges

1) We need better methods for connecting molecular data on microbes in patients with disease causality.

2) We need much more information on predators (phage) in the gut microbiome, starting with basic information on predator-prey ratios and predation rates.

3) The gut virome shows extensive epigenetic features in the form of DNA modification. These are mostly uninvestigated in metagenomic data.
Credits

Sam Minot  
Emily Charlson  
Christian Hoffmann  
Scott Sherrill-Mix  
Serena Dollive  
Kyle Bittinger  
Rohini Sinha  
Stephanie Grunberg  
Alexandra Bryson  
Christel Chehoud  
Jacques Young

*Gary Wu  
*James Lewis  
Tyson Clark  
Ron Collman  
Hongzhi Li  
Eric Chen  
Bob Baldassano  
Sue Keilbaugh  
Ying-Yu Chen  
Lisa Nessel  
Jun Chen

* co-PIs