Studying RNA Virus Replication with Cryo-Electron Microscopy on HTC

Hong ZHAN

2019 May 20th
Positive-strand RNA virus: threats to public health

RNA Virus

Interaction

Host
(Plants, microbiomes, animals, humans)

Yellow fever virus

SARS outbreak in Asia

HBV-C & liver cancer
Outlines

1. Overview of Cryo-EM method in the study of nano-machinery of RNA virus genome replication complex

2. Data processing with HTC

3. Using HTC in study RNA viral replication machinery
Outlines

1. Overview of Cryo-EM method in the study of nano-machinery of RNA virus genome replication complex

2. Data processing with HTC

3. Using HTC in study RNA viral replication machinery
Nodavirus: RNA genome replication overview

Flock Horse Virus (model system)

Viral particle

Infection

Pass through host membrane

RNA1

RNA2

RNA3

Protein A

Protein B1/B2

Capsid

Mitochondria [“Energy Factory” for Cell]

New virus
Cryo-electron tomography/subtomogram averaging
High-resolution study of protein structures

Hole size: 2µm
Hole center distance: 4µm
Plunge freeze vitrification

Carbon film
⌀: 3mm
Hole size: 2µm
Hole center distance: 4µm
Plunge freeze vitrification

Tilt series from -60° to 60° (3° increment)
Back-projection

3D tomogram reconstruction

Averaged 3D crown
Example: Cryo-Tomography of isolated infected mitochondria
Major challenges in Cryo-EM study

Instrument imperfection

Irradiation damage

Sample characteristics
Solutions

Instrument imperfection: Perfect alignment/calibration; computational correction for retrieve degraded information

Sample damage: Dose-symmetric acquisition (from “best” to “worst”)

Thermo-drift due to irradiation: Dose-fractionation or take several frames instead of a single image

Sample characteristics: alternative approach to overcome (single particle)

Increase computational costs exponentially

How to process large cryo-EM data in a reasonable time period?
Outlines

1. Overview of Cryo-EM method in the study of nano-machinery of RNA virus genome replication complex

2. Data processing with HTC

3. Using HTC in study RNA viral replication machinery
Difficulties in HTCondor for Cryo-EM study

1. Human interaction
   a) Align frames
   b) Pick targets

2. Lack of graphic visualization ability
   a) Check quality
   b) No GUI

3. Software
   Open-source software
   Multiple programs for different steps
Difficulties in HTCondor for Cryo-EM study

Solutions

1. Human interaction
   a) Align frames
   b) Pick targets

   Find optimal parameters
   Parameter 1
   Parameter 2
   Parameter 3
   ...

   One sample

2. Lack of graphic visualization ability
   a) Check quality
   b) GUI

   GUI support
   Alternative way to visualize results locally

3. Software
   Open-source software
   Multiple programs for different steps

   Divide steps for using different software
Outlines

1. Overview of Cryo-EM method in the study of nano-machinery of RNA virus genome replication complex

2. Data processing with HTC

3. Using HTC in study RNA viral replication machinery
Overview of Cryo-EM work-flow on HTC

Submit node/home directory
Data, programs, other files

File system (Online)
(Squid [100MB-1GB])

"Mount"

CHTC CPU pools

Output

"download"

Local workstation

Final results

File system (Gluster [>1GB])
Queue jobs from a list with parameters

List of movies to align independently:

List of 3503

Executable script:
Using $1, $2, $3 ... to call different field in a list file as an input
Perform "alignment" more efficiently

Each movie alignment takes ~ 2-5 min on one cpu of a standalone workstation ~ 5 hr to 13 hr
Using multiple CPUs within one job

<table>
<thead>
<tr>
<th>PID</th>
<th>USER</th>
<th>PR</th>
<th>NI</th>
<th>VIRT</th>
<th>RES</th>
<th>SHR</th>
<th>S</th>
<th>%CPU</th>
<th>%MEM</th>
<th>TIME+</th>
<th>COMMAND</th>
</tr>
</thead>
<tbody>
<tr>
<td>5673</td>
<td>hzhan3</td>
<td>20</td>
<td>0</td>
<td>300552</td>
<td>299852</td>
<td>772</td>
<td>R</td>
<td>100.0</td>
<td>0.2</td>
<td>122:48.81</td>
<td>CTFcorrect.exe</td>
</tr>
<tr>
<td>5684</td>
<td>hzhan3</td>
<td>20</td>
<td>0</td>
<td>277044</td>
<td>270124</td>
<td>772</td>
<td>R</td>
<td>100.0</td>
<td>0.2</td>
<td>122:46.93</td>
<td>CTFcorrect.exe</td>
</tr>
<tr>
<td>12492</td>
<td>hzhan3</td>
<td>20</td>
<td>0</td>
<td>204608</td>
<td>204384</td>
<td>772</td>
<td>R</td>
<td>100.0</td>
<td>0.2</td>
<td>108:26.46</td>
<td>CTFcorrect.exe</td>
</tr>
<tr>
<td>30561</td>
<td>hzhan3</td>
<td>20</td>
<td>0</td>
<td>143204</td>
<td>142952</td>
<td>772</td>
<td>R</td>
<td>100.0</td>
<td>0.1</td>
<td>47:35.06</td>
<td>CTFcorrect.exe</td>
</tr>
<tr>
<td>34735</td>
<td>hzhan3</td>
<td>20</td>
<td>0</td>
<td>313996</td>
<td>313804</td>
<td>772</td>
<td>R</td>
<td>98.4</td>
<td>0.2</td>
<td>256:54.91</td>
<td>CTFcorrect.exe</td>
</tr>
<tr>
<td>34748</td>
<td>hzhan3</td>
<td>20</td>
<td>0</td>
<td>285200</td>
<td>285016</td>
<td>772</td>
<td>R</td>
<td>96.8</td>
<td>0.2</td>
<td>256:37.02</td>
<td>CTFcorrect.exe</td>
</tr>
<tr>
<td>12504</td>
<td>hzhan3</td>
<td>20</td>
<td>0</td>
<td>198876</td>
<td>195472</td>
<td>772</td>
<td>R</td>
<td>95.5</td>
<td>0.1</td>
<td>102:42.80</td>
<td>CTFcorrect.exe</td>
</tr>
<tr>
<td>30545</td>
<td>hzhan3</td>
<td>20</td>
<td>0</td>
<td>146076</td>
<td>145820</td>
<td>772</td>
<td>R</td>
<td>91.9</td>
<td>0.1</td>
<td>42:16.78</td>
<td>CTFcorrect.exe</td>
</tr>
<tr>
<td>19552</td>
<td>hzhan3</td>
<td>20</td>
<td>0</td>
<td>233404</td>
<td>231096</td>
<td>772</td>
<td>R</td>
<td>88.4</td>
<td>0.2</td>
<td>106:44.20</td>
<td>CTFcorrect.exe</td>
</tr>
<tr>
<td>19592</td>
<td>hzhan3</td>
<td>20</td>
<td>0</td>
<td>213240</td>
<td>213024</td>
<td>772</td>
<td>R</td>
<td>70.3</td>
<td>0.2</td>
<td>135:52.54</td>
<td>CTFcorrect.exe</td>
</tr>
<tr>
<td>19540</td>
<td>hzhan3</td>
<td>20</td>
<td>0</td>
<td>246836</td>
<td>246208</td>
<td>772</td>
<td>R</td>
<td>68.4</td>
<td>0.2</td>
<td>105:11.63</td>
<td>CTFcorrect.exe</td>
</tr>
<tr>
<td>19565</td>
<td>hzhan3</td>
<td>20</td>
<td>0</td>
<td>225836</td>
<td>223696</td>
<td>772</td>
<td>R</td>
<td>54.8</td>
<td>0.2</td>
<td>105:21.76</td>
<td>CTFcorrect.exe</td>
</tr>
</tbody>
</table>
Cryo-tomography/subtomogram averaging using HTC

1. Align frames
   - 2hrs on HTC

2. Align tilts (projections)
   - 4hr with GPU for 31 tomograms

3. Picking targets
   - HPC system
   - Workstation
   - 4 days

Scale bar: 20nm
Future direction ...

HTC usage: **8,812** total HTC hours

High resolution crown structure: More data, better pre-processing
Acknowledgements

**Dr. Paul Ahlquist’s lab**
- Paul Ahlquist
- Megan Bracken
- James Bruce
- Zach Coleman
- Johan den Boon
- Reza Djavadian
- Mark Horswill
- Maskaki Nishikiori
- Janice Pennington
- Nuruddin Unchwaniwala

**Cryo-EM facility at Janelia**
- Rick Huang
- Chuan Hong

**Cryo-EM facility Pacific Northwest Cryo-EM center**
- Claudia Lopez
- Craig Yoshioka

**Janelia Research Campus**
- Dr. Nikolaus Gregorieffl’s lab
  - Tim Grant
  - Benjamin Himes

**UW-Madison Data-hub**
- Sarah Stevens
- Steven Goldstein

**UW-Madison Center for High Throughput Computing**
- Lauren Michael
- Christian Koch

**Morgridge Core Computation**
- Miron Livny
- Brian Bockelman

**Morgridge Virology**
- Anthony Gitter