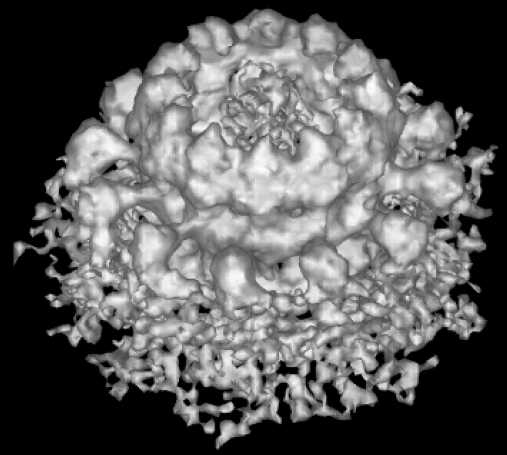
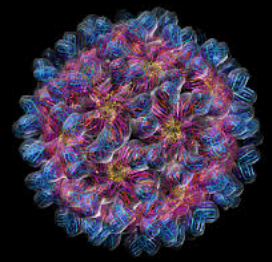
Studying RNA Virus Replication with Cryo-Electron Microscopy on HTC



Hong ZHAN

2019 May 20th

Positive-strand RNA virus: threats to public health



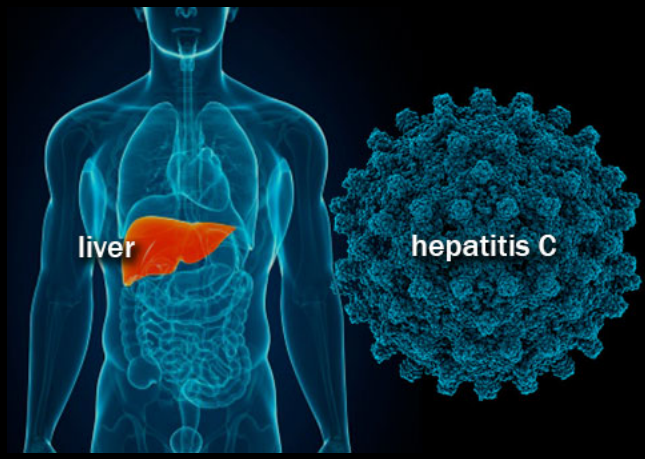
RNA Virus

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

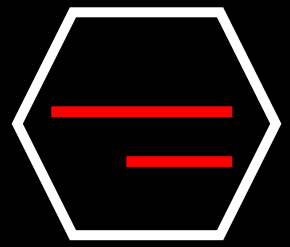
1. Overview of Cryo-EM method in the study of nano-machinery of RNA virus genome replication complex

2. Data processing with HTC

3. Using HTC in study RNA viral replication machinery

Nodavirus: RNA genome replication overview

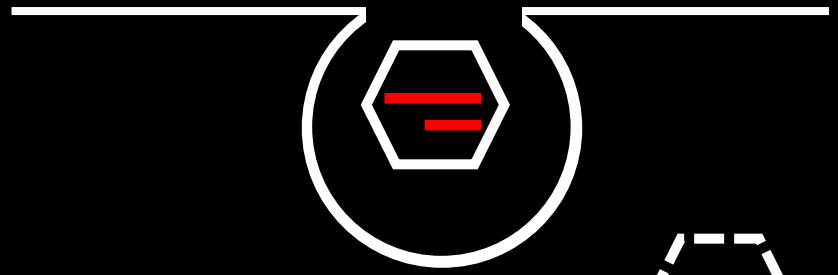
Flock Horse Virus (model system)

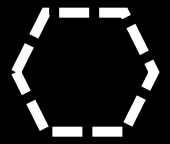


Viral particle

Infection

Pass through host membrane





RNA1

RNA3

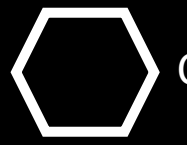
RNA2

RNA1

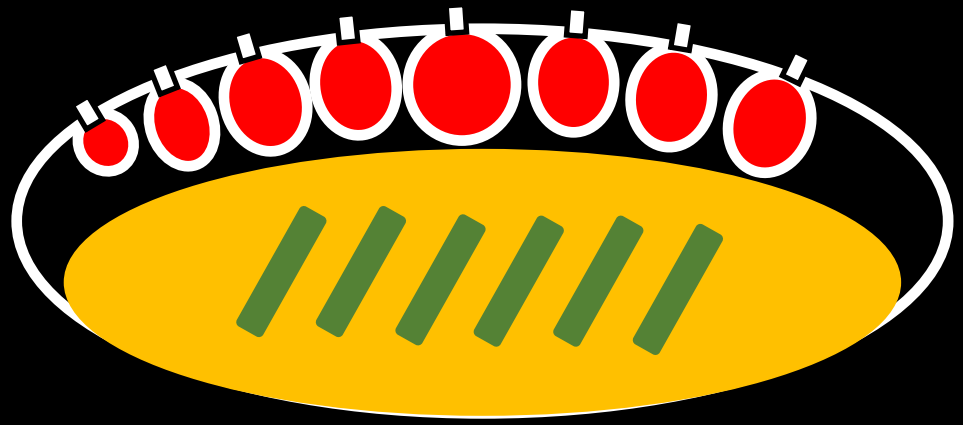
RNA2

Protein A

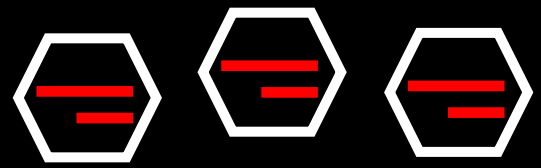
Protein B1/B2



Capsid

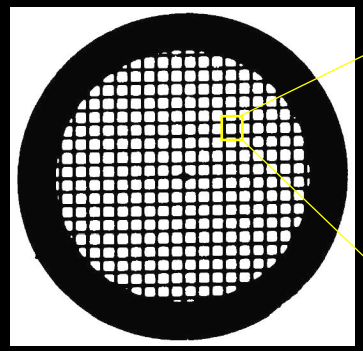


Mitochondria [“Energy Factory” for Cell]

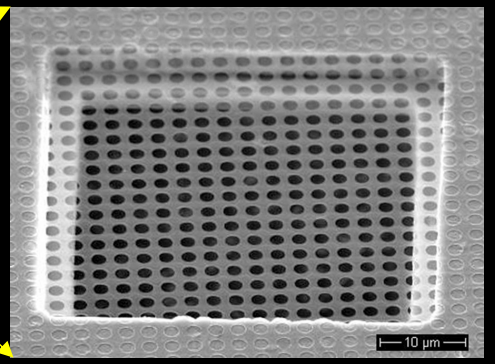


New virus

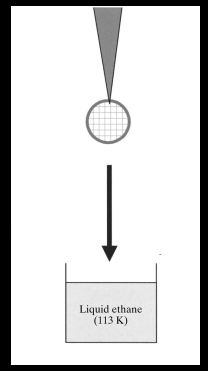
Cryo-electron tomography/subtomogram averaging
High-resolution study of protein structures



⌀: 3mm

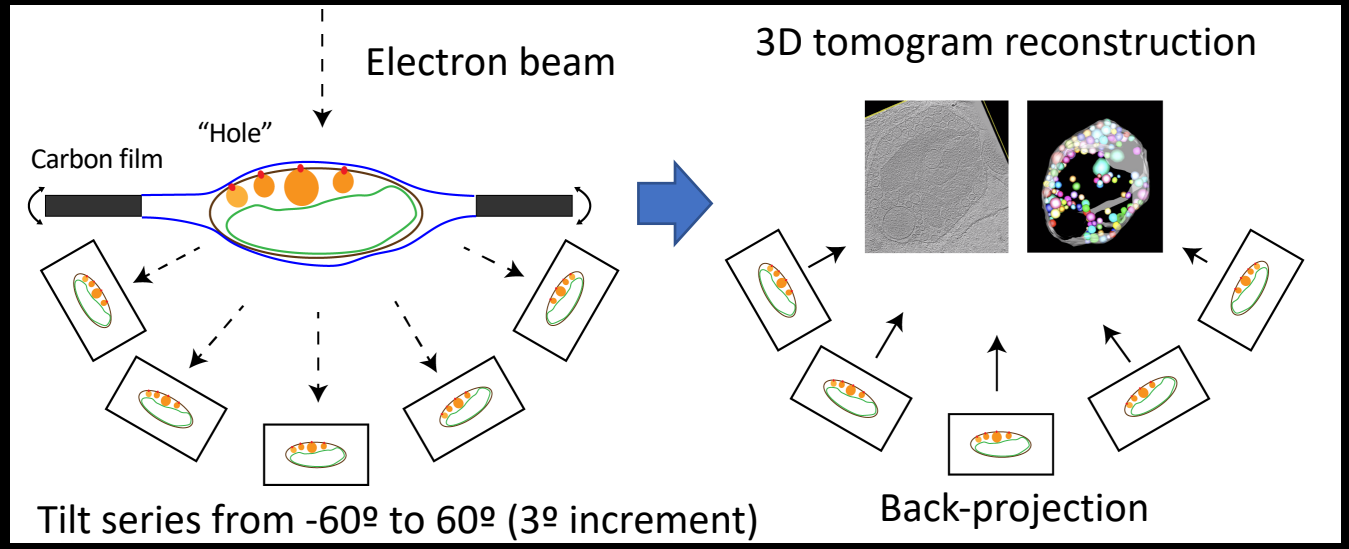


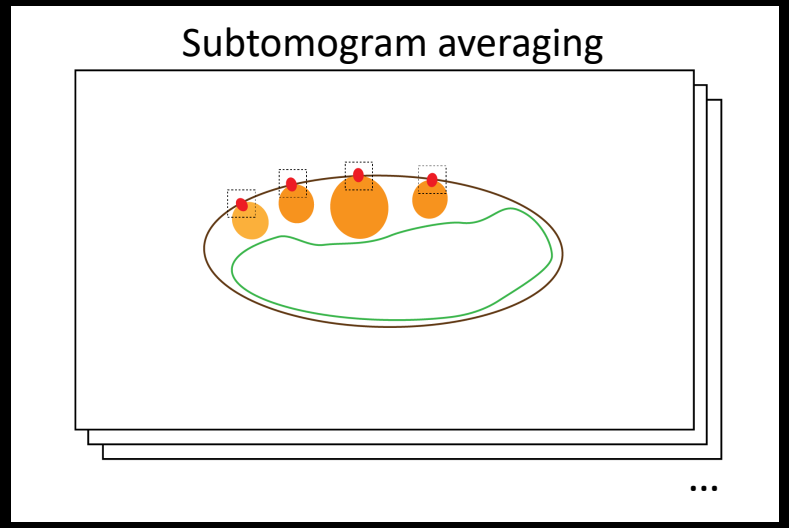
Hole size: 2µm
Hole center distance: 4µm

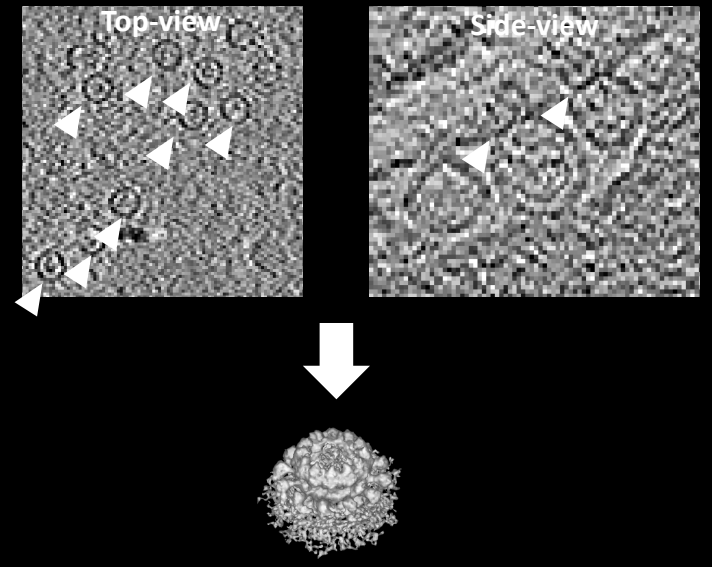


Plunge freeze vitrification

From: Electron Microscopy Sciences

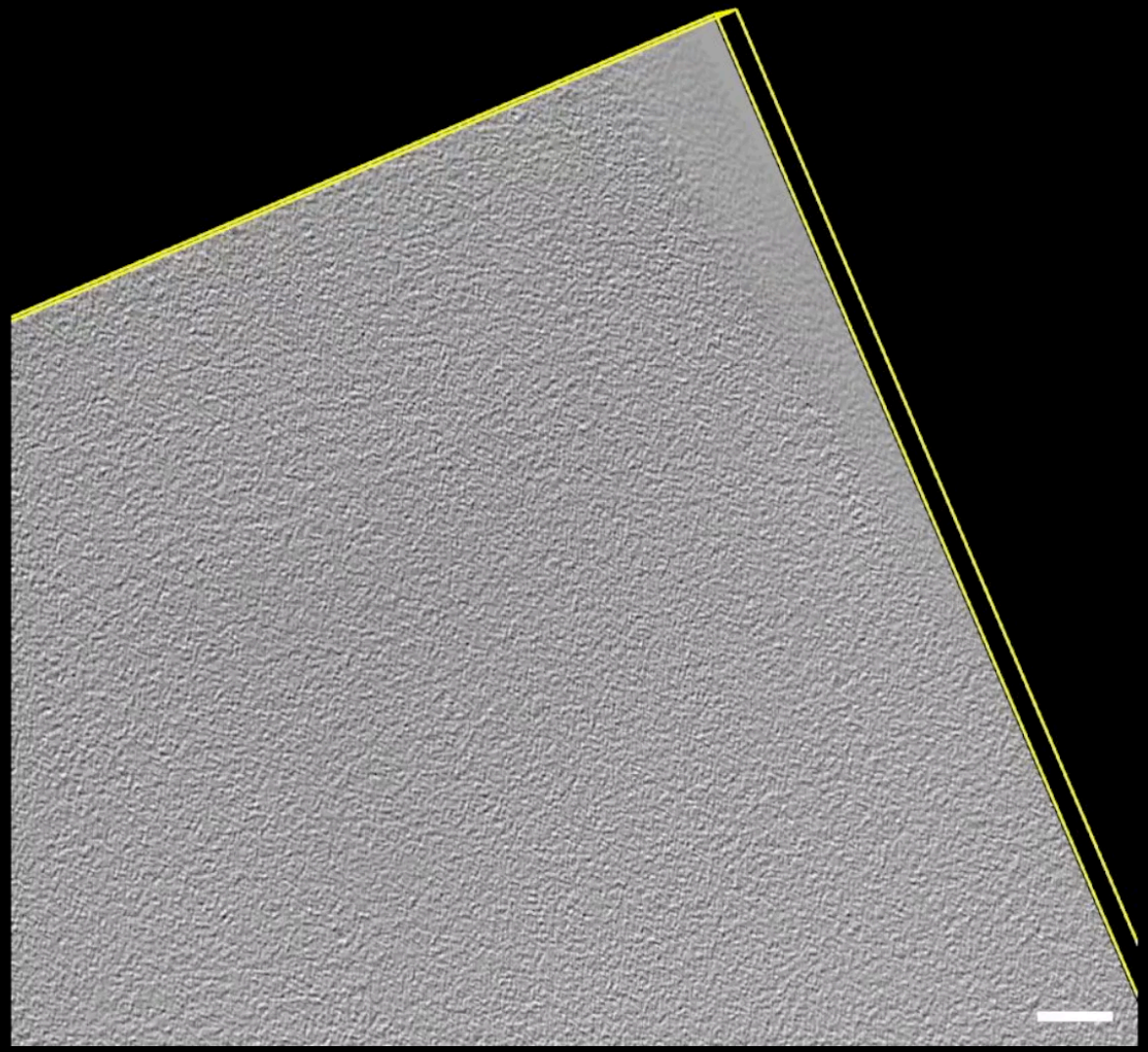






Averaged 3D crown

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

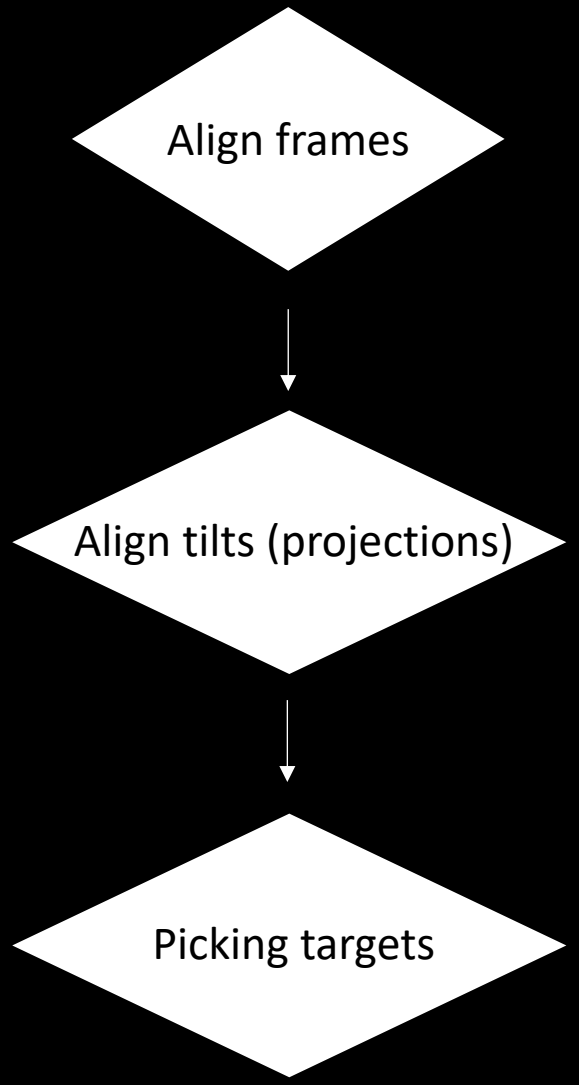
3. Using HTC in study RNA viral replication machinery

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

One sample
- Parameter 1
- Parameter 2
- Parameter 3
- ⋮

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

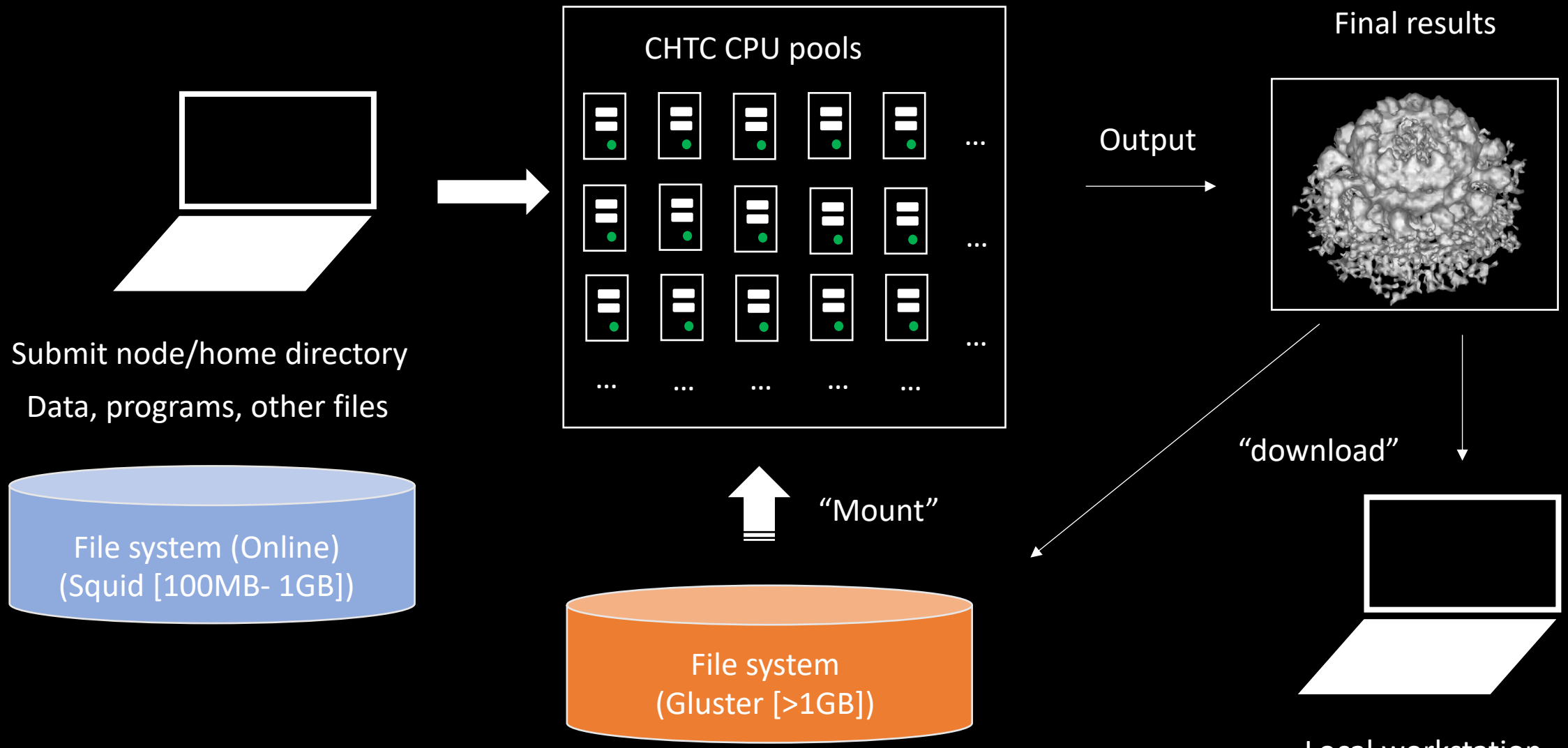
Outlines

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



Queue jobs from a list with parameters

List of movies to align independently:

List of 3503

```
 GNU nano 2.3.1                          File: listname.txt

projectDir124
projectDir125
projectDir126
projectDir127
projectDir129
projectDir130
projectDir131
projectDir400
projectDir401
projectDir404
projectDir409
projectDir411
projectDir412
projectDir414
projectDir415
projectDir416
projectDir418
projectDir419
projectDir420
projectDir421
projectDir422
projectDir423
projectDir424
projectDir468
projectDir485
projectDir505
projectDir506
projectDir514
projectDir516
projectDir526

^G Get Help    ^O WriteOut    ^R Read File    ^Y Prev Page
^X Exit        ^J Justify     ^W Where Is     ^V Next Page
```

Executable script:
Using $1, $2, $3 … to call different field in a list file as an input

```
#!/bin/bash

# copy tar file from /mnt/gluster into the working directory
cp /mnt/gluster/hzhan3/Micrographs/$1 ./

# untar unblur from cistem
tar -xzf cistem.tar.gz

# add path to the environment
mkdir home
export HOME=$(pwd)/home
export PATH=$(pwd)/cistem:$PATH

# unblur
unblur << foo
$1
$1_output.mrc
1.77
2
yes
300
1.0
0.0
yes
2.0
80.0
150
1
1
1
20
yes
yes
1
0
no
foo

# to transfer big outout to proper file system
tar -czvf $1_output.tar.gz $1_output.mrc
mv $1_output.tar.gz /mnt/gluster/hzhan3/relion_test/Micrographs
rm $1 $1_output.mrc

## END
```

Perform ”alignment” more efficiently

Each movie alignment takes ~ 2-5min on one cpu of a standalone workstation ~5hr to 13hr

Using multiple CPUs within one job

PID	USER	PR	NI	VIRT	RES	SHR	S	%CPU	%MEM	TIME+	COMMAND
5673	hzhan3	20	0	300552	299852	772	R	100.0	0.2	122:48.81	CTFcorrect.exe
5684	hzhan3	20	0	277044	270124	772	R	100.0	0.2	122:46.93	CTFcorrect.exe
12492	hzhan3	20	0	204608	204384	772	R	100.0	0.2	108:26.46	CTFcorrect.exe
30561	hzhan3	20	0	143204	142952	772	R	100.0	0.1	47:35.06	CTFcorrect.exe
34735	hzhan3	20	0	313996	313804	772	R	98.4	0.2	256:54.91	CTFcorrect.exe
34748	hzhan3	20	0	285200	285016	772	R	96.8	0.2	256:37.02	CTFcorrect.exe
12504	hzhan3	20	0	198876	195472	772	R	95.5	0.1	102:42.80	CTFcorrect.exe
30545	hzhan3	20	0	146076	145820	772	R	91.9	0.1	42:16.78	CTFcorrect.exe
19552	hzhan3	20	0	233404	231096	772	R	88.4	0.2	106:44.20	CTFcorrect.exe
19592	hzhan3	20	0	213240	213024	772	R	70.3	0.2	135:52.54	CTFcorrect.exe
19540	hzhan3	20	0	246836	246208	772	R	68.4	0.2	105:11.63	CTFcorrect.exe
19565	hzhan3	20	0	225836	223696	772	R	54.8	0.2	105:21.76	CTFcorrect.exe
16394	hzhan3	20	0	60524	4940	3644	R	0.6	0.0	0:00.11	top

Cryo-tomography/subtomogram averaging using HTC

Align frames

2hrs on HTC

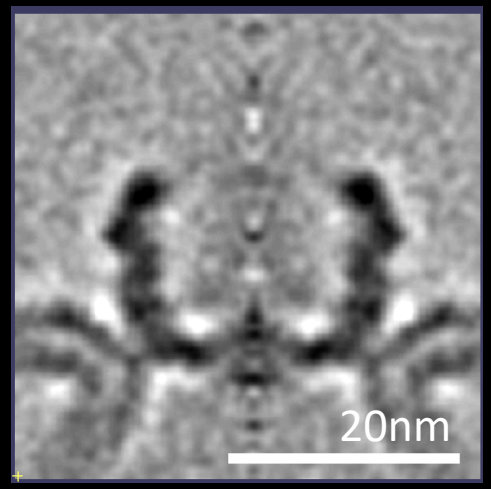
Align tilts (projections)

4hr with GPU for 31 tomograms

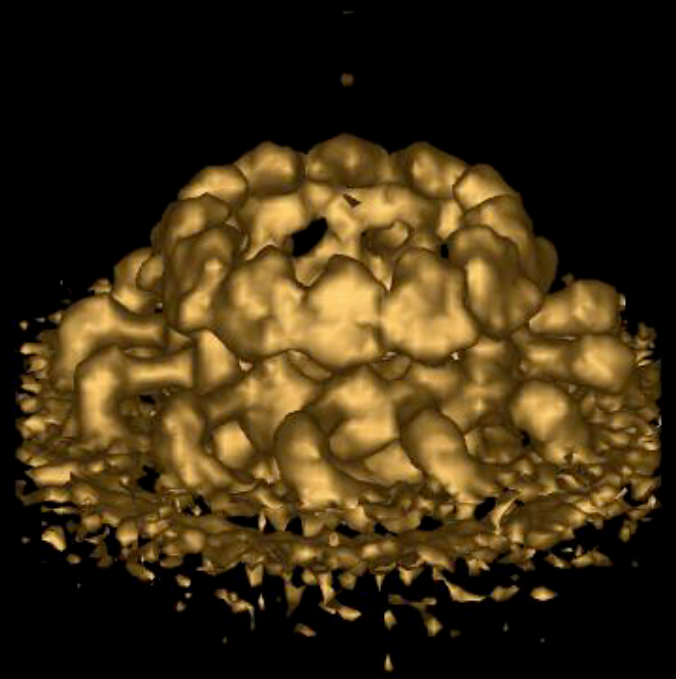
Picking targets

HPC system

Workstation



4 days

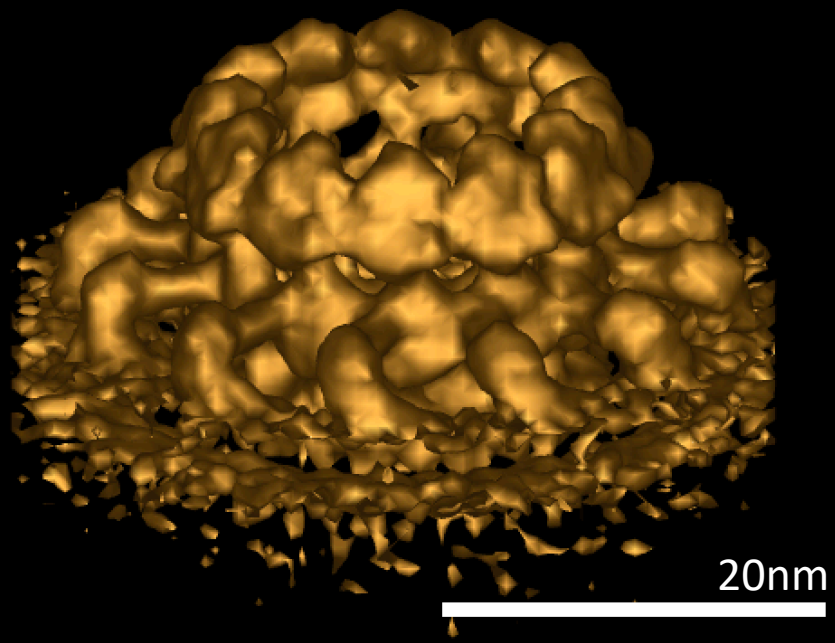


Future direction …

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

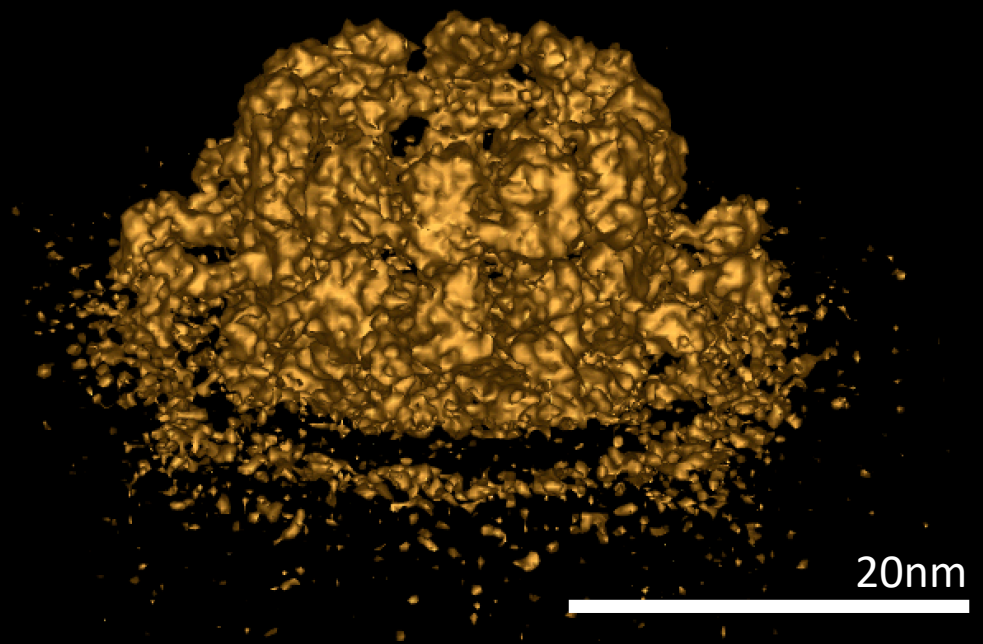
High resolution crown structure: More data, better pre-processing

18Å



Forward

12Å



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