American Museum of Natural History

Connecting Campus Users to Remote dHTC Resources Using OSG Access Points

Sajesh Singh (ssingh@amnh.org)

Manager, Scientific Computing

About the Museum

OUR MISSION

To discover, interpret, and disseminate—through scientific research and education—knowledge about human cultures, the natural world, and the universe.

About the Museum

- Founded in 1869.
- 28 interconnected buildings housing permanent and temporary exhibition halls, a library, and the Hayden Planetarium.
- Collections containing 33 million specimens.
- Occupies 2 million square feet
- 5 million annual visitors.
- Richard Gilder Center for Science, Education, and Innovation under construction (opening Dec 2023).





AMNH and the Open Science Grid

- AMNH piloted the OSG to provide additional computational resources to our researchers.
- After evaluating the OSG technology stack and the benefits to our researchers, AMNH decided to move forward with implementing OSG in production.
- OSG also helped AMNH meet the NSF CC* grant requirement to make available 20% of the funded computing resources to be shared with outside researchers.
- Engaged with the OSG User School to bring AMNH staff rapidly up to speed.

Open Science Grid

AMNH and OSG Access Points

- Even though AMNH was new to OSG we felt it necessary to immerse ourselves in their technologies.
- One of those technologies that we felt would be of use to our on-site researchers was the local Access Point

Open Science Grid

AMNH and OSG Access Points

- Built onsite Access Point.
  - Researchers can login to an OSG access point using local credentials.
  - Jobs can access input data locally without need to transfer to intermediary first.
  - Allows for output to be available at job completion without need to manually transfer data back.

Open Science Grid

AMNH and OSG Access Points

Authentication and Authorization

- Researchers can login to an OSG access point using local credentials
- Benefits:
  - No need to maintain separate login/SSH keys for remote server.
  - Organizational security polices can be applied such as 2FA (DUO).

Open Science Grid

AMNH and OSG Access Points

- Jobs can access input data locally without need to transfer to intermediary first.
  - No need to stage data via OSG Connect Access Points.
  - User can always be sure they are working off the latest set of data if using only one reference set.
  - Reduction of data sprawl.
  - Able to mount data generated from on-site HPC resources locally on the Access Point.

Open Science Grid

AMNH and OSG Access Points

- Allows for output to be available at job completion without need to manually transfer data back.
- Further post processing/analysis using on-site HPC resources for jobs that are not suited for the OSG.

Open Science Grid

AMNH and OSG Access Points

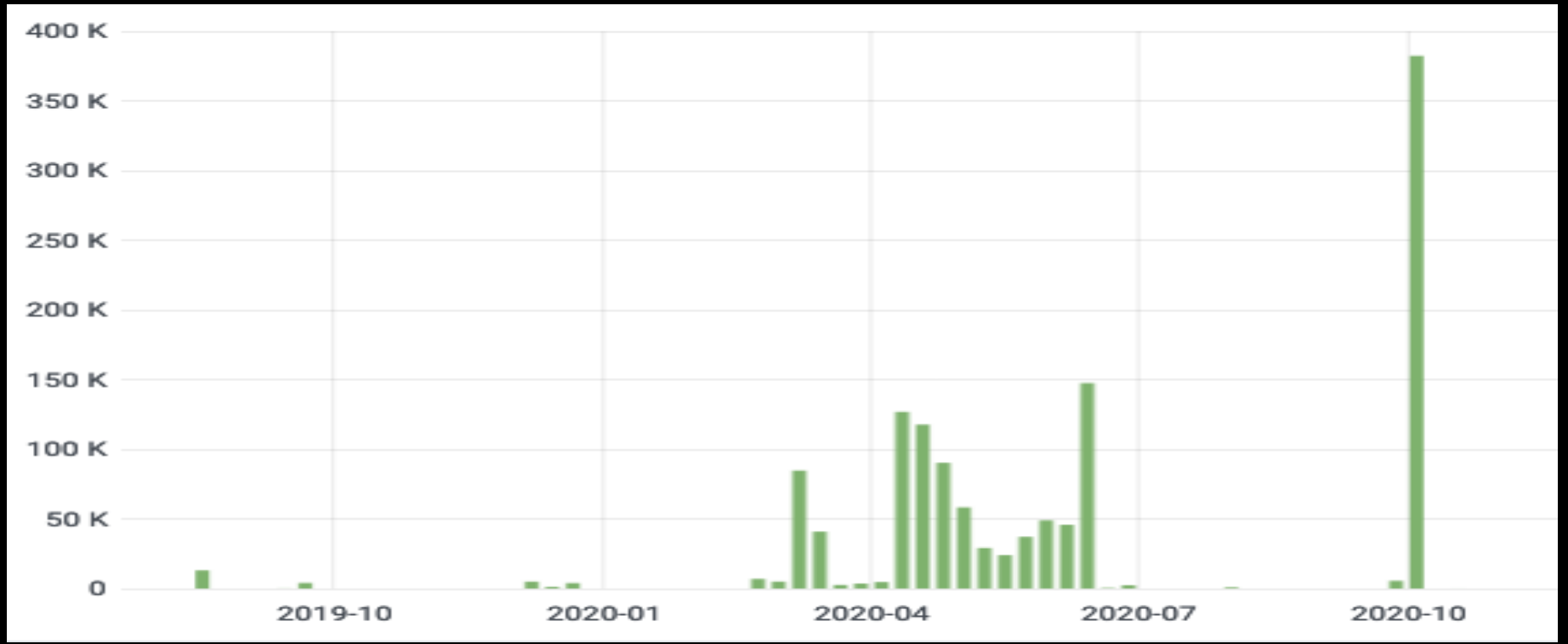
- Access Point server should be sized appropriately for your environment.

- AMNH Access Point Specs:
  - 1U Server
  - 24 Cores @ 3 GHz
  - 72 GB RAM
  - SSD drives for Condor spool directory
  - CentOS 7

Open Science Grid

AMERICAN MUSEUM OF NATURAL HISTORY

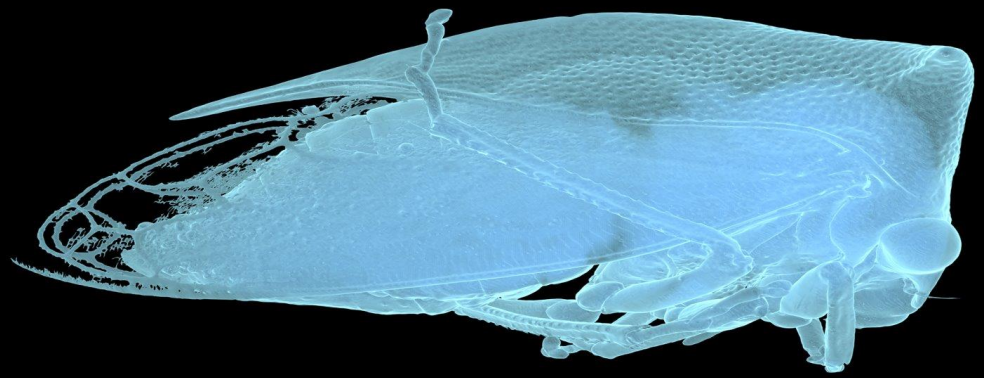
AMNH AS OSG CONSUMER



Hollister Herhold, PhD. Student, RGGS

This is an image of a Tree Hopper (Order Hemiptera, Family Membracidae).

- **Micro-CT scanned at the Microscopy and Imaging Facility (MIF) at AMNH, at a voxel size (resolution) of 8 microns.**
- **Entire specimen is less than 1cm in length, demonstrating the ability to recover extremely fine morphological details.**
- **Post-processed using 3D Slicer, an open-source project for volume image data and analysis. Lighting and texturing was done in Blender 2.80, a popular open-source animation and rendering package.**
- **Image was rendered using Blender on the Open Science Grid.**



Kaiya Provost, Research Associate & PhD. Candidate

- Using OSG to run population simulations to evaluate how species have evolved in southwestern deserts.
- Generated 24,000+ and counting simulations using population simulation software SLiM 3.
- Using OSG instead of on-site resources only this project went from a 6-year simulation project to a 6-week simulation project.
- Paper in pre-print: https://www.biorxiv.org/content/10.1101/2020.06.17.157842v1.article-metrics

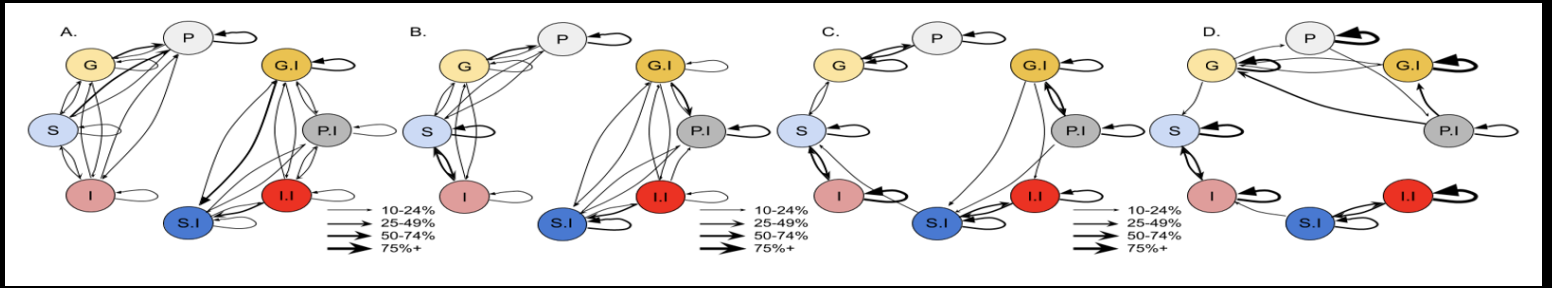
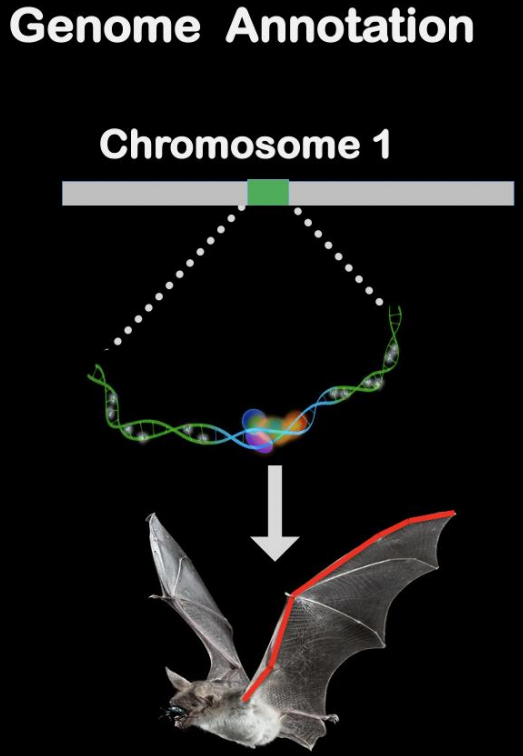


Figure 6: Overall model classifications are easily confused at young ages but improve with older ages. Arrows begin at the true model and end at the assigned model. Line thickness of arrows indicates the percentage of time those assignments are made. Assignments under 10% are omitted for clarity. Demographics include single panmictic populations ("P", grey), isolation-with-migration or gene flow ("G", yellow), secondary contact ("S", blue), and isolation ("I", red). Suffix of ".I" after demography indicates that IBD is present. A) 6,000 generations. B) 21,000 generations. C) 120,000 generations. D) 1,000,000 generations. Models with IBD are on the right, models without IBD are on the left.

Ariadna Morales - Gerstner Postdoctoral Scholar

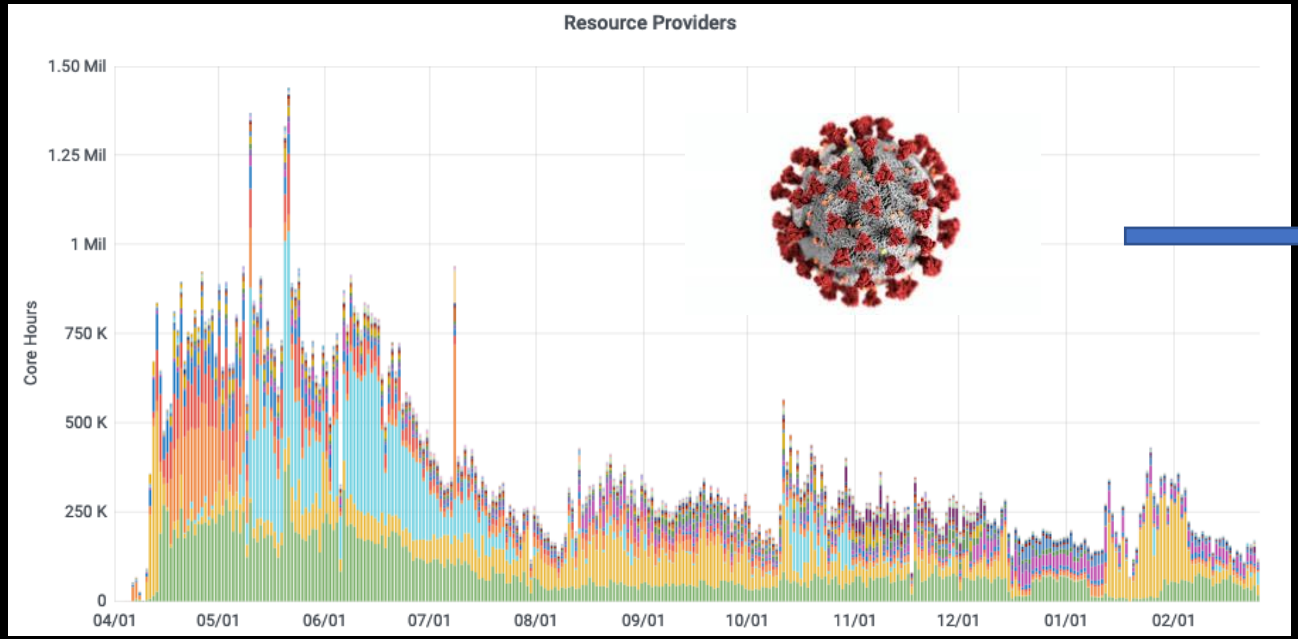


- Analysis of the genome data of of the genus Myotis
- Research at the AMNH is setting the genomic basis of a new model system in mammals (bats) to study the mechanisms of repeated evolution of traits.
- Used OSG to run 80K jobs that took 4 weeks to complete.
  - Using only onsite resources would have taken 4 months or more.
- Results generated by this research are relevant to understand immune system of bats and resistance to diseases that may impact humans.
  - Especially important considering COVID-19.

AMNH AS OSG PROVIDER (GENERAL)

~4.17 Million core hours provided to the OSG via three AMNH clusters (ARES, HEL, MENDEL)

AMNH AS OSG PROVIDER (COVID-19)



	total
Brookhaven National Laboratory	28.14 Mil
Fermi National Accelerator Laboratory	22.80 Mil
Syracuse University	19.09 Mil
University of Chicago	11.55 Mil
Massachusetts Institute of Technology	8.68 Mil
University of Nebraska	5.67 Mil
University of Michigan	4.72 Mil
University of Texas Arlington	3.55 Mil
American Museum of Natural History	2.80 Mil
University of California San Diego	2.45 Mil
University of Wisconsin	2.12 Mil
Vanderbilt University	1.84 Mil
Rutgers University	1.56 Mil
Universidade Estadual Paulista	1.21 Mil
University of Connecticut	1.17 Mil
New Mexico State University	816.88 K
University of Utah	810.36 K
University of Colorado	652.17 K
University of Mississippi	618.28 K
University of Southern California	520.41 K
Clarkson University	451.73 K
University of Wisconsin Milwaukee	393.53 K
Universidade do Estado do Rio de Janeiro	314.86 K
University of Notre Dame	286.45 K
CancerComputer	214.48 K

OSG resources provided via three AMNH clusters (ARES, HEL, MENDEL)

Thank You

Questions?