

January 19, 2016

Apache Airavata <-> Science Gateways

IU Science Gateways Group

XSEDE

Extreme Science and Engineering
Discovery Environment

seg@iu.edu



Outline

- This is not a typical ECSS Symposium talk.
- Will not discuss much of science, more of a summary of process
 - transition from ESRT and ESCC projects into ESSGW.
- Goal is to provoke thinking to cultivate more gateways from ECSS projects.



Phasta Science Gateway

- Large Eddy Simulations of Turbulent Flow
 - PI: Mark Shephard, Cameron Smith, Rensselaer Polytechnic Institute
- ESCSS Consultants:
 - Lei Hang, David O'Neal, Carlos Rosales, Raminder Singh, Suresh Marru

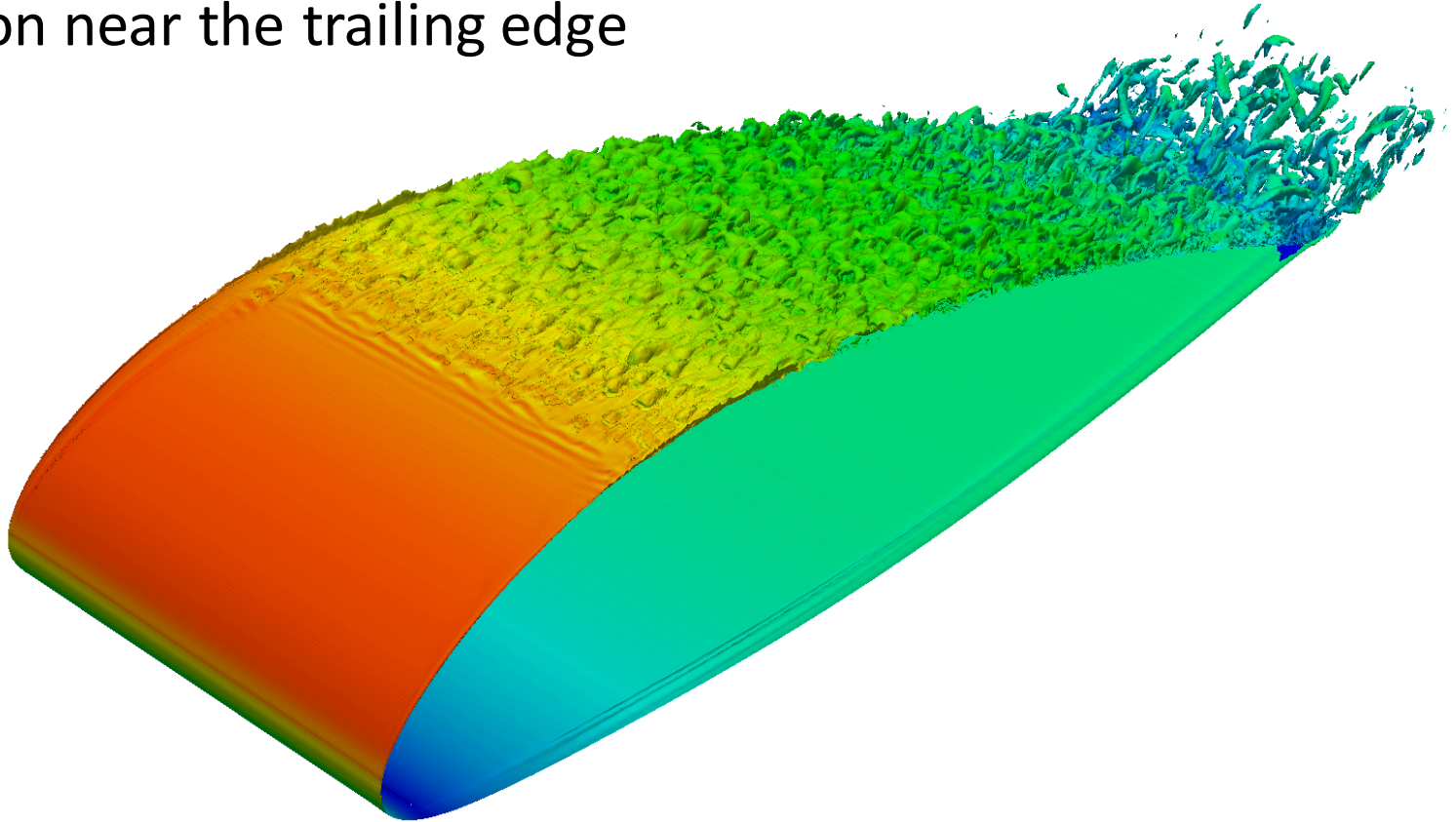


Science Use Case: PHASTA

- One of the challenges in studying turbulent flows is the multiscale nature of turbulence.
 - Direct Numerical Simulation (DNS) resolves turbulent flow structures up to dissipation scale, but is not practical with current HPC resources on engineering problems of interest.
- Large Eddy Simulation (LES) is a powerful tool which is capable of accurately resolving and modeling turbulence at a reasonable cost.
 - LES resolves those flow features that can be captured by the numerical discretization/grid while modeling their interactions with the unresolved (subgrid scale) flow features.
 - To account for inhomogeneous flows, Lagrangian-averaged dynamic Smagorinsky model is employed in stabilized finite element simulation.
- Example case: Flow over Aerospatiale-A airfoil
 - Static angle of attack 13° , Reynolds number 2,000,000
 - Flow transitions to a turbulent flow and exhibits marginal flow separation near the trailing edge



- Flow over Aerospatiale-A airfoil
 - Static angle of attack: 13°
 - Reynolds number: 2,000,000
 - Large Eddy Simulation (LES)
- Lagrangian-averaged dynamic Smagorinsky model is employed in stabilized finite element simulation
- Flow transitions to a turbulent flow and exhibits marginal flow separation near the trailing edge



PHP Gateway with Airavata

PGA is a science gateway built with the Airavata API. You can reference PGA as you integrate Airavata into your own gateway, or you can create your gateway on top of PGA by cloning it at the link below. PGA is known to work well in the Chrome, Firefox, and Internet Explorer browsers.

[See the code](#)

[XSEDE 2015 tutorial documentation](#)



SciGaP is a hosted service with a public API that science gateways can use to manage applications and workflows running on remote supercomputers, as well as other services. Gateway developers can thus concentrate their efforts on building their scientific communities and not worry about operations.

Science Gateway Platform as a Service (SciGaP) provides application programmer interfaces (APIs) to hosted generic infrastructure services that can be used by domain science communities to create Science Gateways.

[Learn more](#)



powered by

Apache Airavata

Apache Airavata is a software framework which is dominantly used to build Web-based science gateways and assist to compose, manage, execute and monitor large scale applications and workflows on distributed computing resources such as local clusters, supercomputers, national grids, academic and commercial clouds. Airavata mainly supports long running applications and workflows on distributed computational resources.

[Learn more](#)

Apache Airavata manages multi-stepped workflow, multiple code versions.

PHASTA team uses Apache Airavata and the PGA to run simulations on TACC's Stampede.

Experiment Summary

Experiment Id	Phast-Exp-Stampede_4e07a9e1-31a2-4210-b3b6-d32bd48cf498
Name	Phast-Exp-Stampede
Description	Phasta exp running on Stampede
Project	November/12/2015
Application	Phasta_P
Compute resource	stampede.tacc.xsede.org
Experiment Status	COMPLETED
Job Status	COMPLETE
Creation time	2015-11-12, 12:43 PM - GMT-0500 (EST)
Last Modified Time	2015-11-12, 12:44 PM - GMT-0500 (EST)
Enable Auto Schedule	false
Wall time	30
CPU count	16
Node count	1
Queue	normal
Inputs	geom.xmt_txt geom.smd geom.sms solver.inp
Outputs	Phasta-Output-TAR : PHASTA_Output.tar.gz Phasta-Standard-Error : Phasta_P.stderr Phasta-Standard-Out : Phasta_P.stdout

Clone

Example Phasta Gateway Tasks

- Build a gateway for PHASTA application
- Enable auto-scheduling to run 120 hour job on stampede
 - Application checkpoints very well, so split that into multiple jobs with 48 hours each and run them in sequence.
- Users want to use their own version of Phasta
 - Enabled through gateway application abstractions

Vortex Shedding Simulation Gateway

- PI: Arne J. Pearlstein, UIUC
- ECSS Consultants:
 - Sudhakar Pamidighantam, Mark Vanmoer



Next generation XSEDE Resources

- Science Gateways and Cloud Research on Jetstream
 - Prof. Madhusudhan Govindaraju
 - State University of New York Binghamton



Chrome

File Edit View History Bookmarks People Window Help

Mesos

scigap-1r.jetstream-cloud.org:5050/#/

Mesos

	CPU	Mem
Total	36	88.0 GB
Used	12	12.0 GB
Offered	0	0 B
Idle	24	76.0 GB

Active Tasks

Find...

ID	Name	State	Started	Host
Marathon				
scigap-1r.jetstream-cloud.org:8080/ui/#/apps/%2Fidv-cloud-withp...				
Suspend Scale Restart App Destroy App				
Tasks Configuration Debug				
Refresh				
ID	Status	Version	Updated	
idv-cloud-withpsw.3e38b7b1-873c-11e5-964a-0242711de25c	Started	2 days ago	11/9/2015, 6:47:33 PM	
scigap-1r.jetstream-cloud.org:31116				
idv-cloud-withpsw.b45354a3-87d2-11e5-964a-0242711de25c	Started	2 days ago	11/10/2015, 12:45:23 PM	
scigap-5r.jetstream-cloud.org:31131				

noVNC

scigap-1r.jetstream-cloud.org:31116

Connected (unencrypted) to: 199dedb13890:1

Send CtrlAltDel Shutdown Reboot Reset

Dashboard

File Edit Displays Data Tools Help

Unidata IDV - Globe View - One Pane

File Edit Displays Data Tools Help

View Projections

2015-11-20 12:00:00Z

15:00:53 GMT Latitude: 2.5 Longitude: 173.7 Altitude: 0.0 m

15:00:53 GMT

Workspace 1 11 Nov, Wed 15:00:53 idv@199dedb13890: ~ Dash



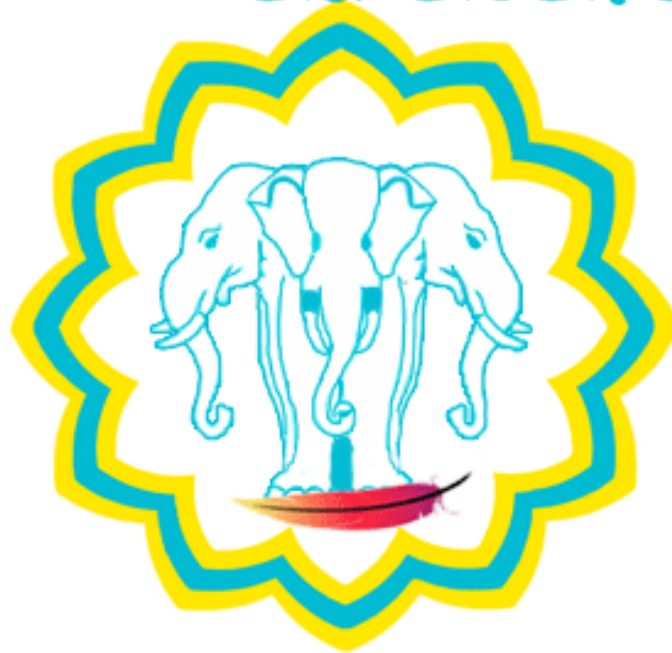
sgg@iu.edu

XSEDE

Preparing for Bridges

- Transcriptome Assembly and Evaluation, using Sequencing Quality Control (SEQC) Data
 - PI: Noushin Ghaffari, Texas A&M University
- Started on Backlight, brief work on Greenfield
- Waiting on Bridges

Apache *airavata*



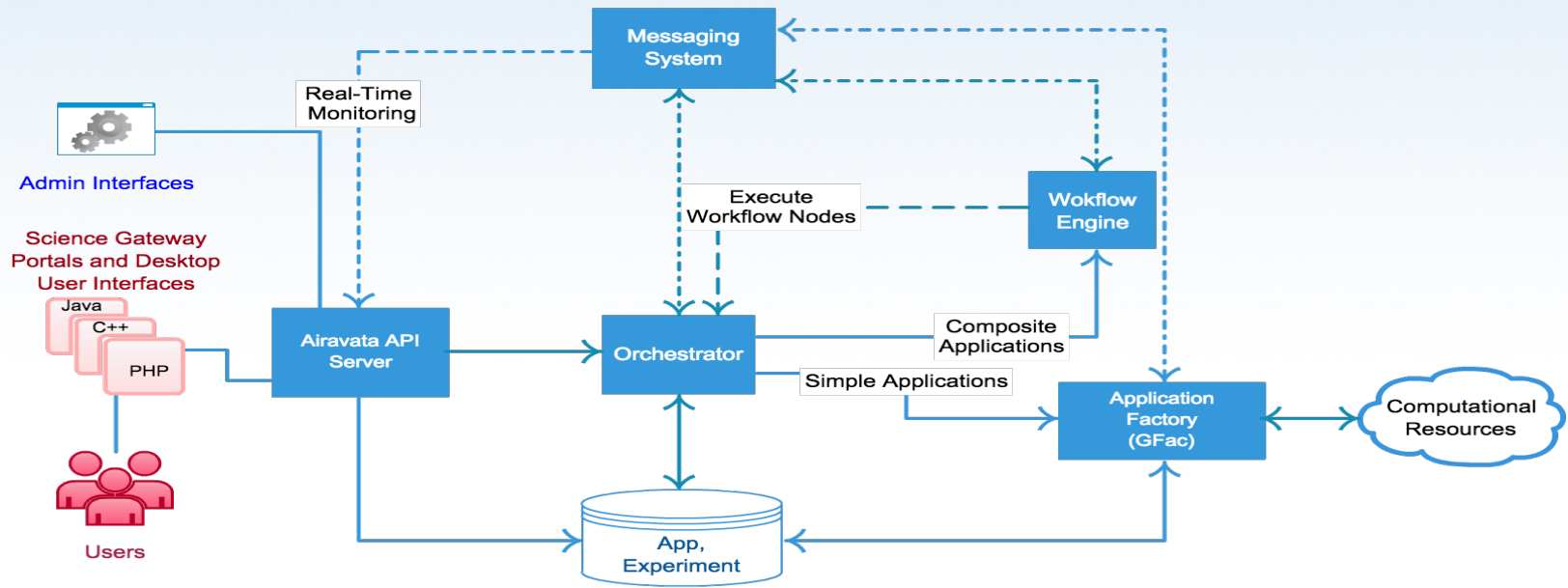
**ECSS Projects Leveraging from and
Contributing to Airavata**

“Apache” Means “Open”



- Join the Airavata developer mailing list, get involved, submit patches, contribute.

Airavata: Multi-Tenant Gateway Middleware



- External clients interact with Airavata API (based on Apache Thrift).
- Internally, components interact with each other through Component Programming Interfaces (thrift based CPIs).

Airavata Overview

- Airavata is a general purpose distributed system software framework build on micro-service and component based architecture principles.
- Airavata provides capabilities to compose, manage, execute and monitor large scale applications and workflows on distributed computing resources.
- Airavata supports executions on local clusters, national grids, academic and commercial clouds.
- Airavata is inherently multi-tenanted.



XSEDE

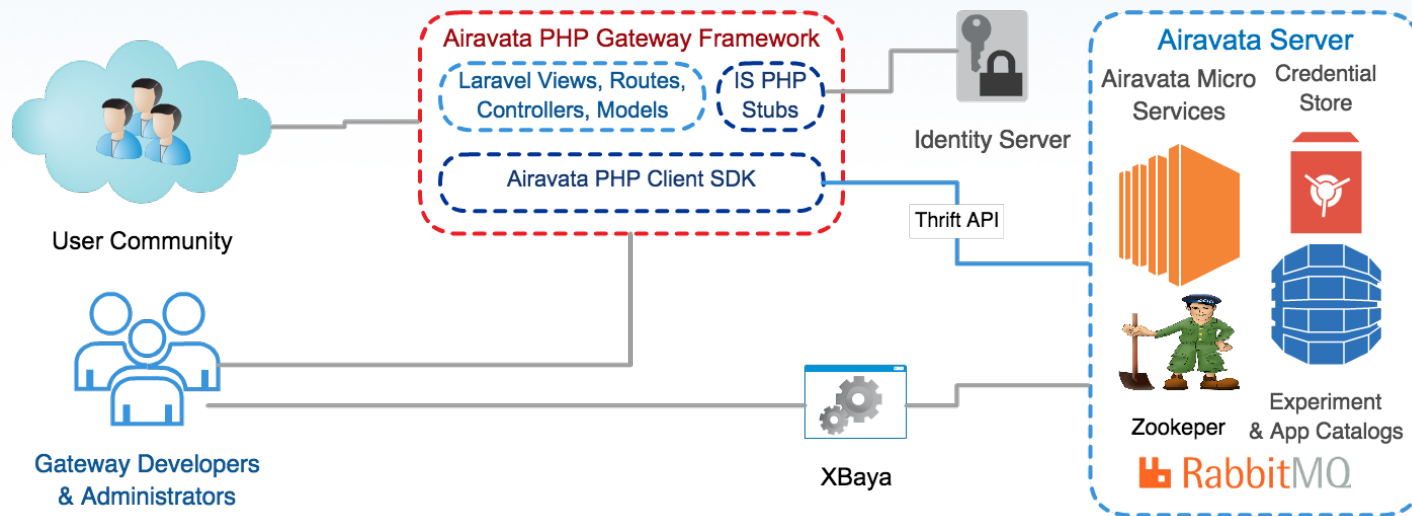
Airavata as a Science Gateway Middleware

- Airavata is dominantly used to build science gateways.
- Airavata supports secured communications to HPC resources and empowers gateway operators to administer and monitor long running executions.
- A reference PHP based gateway is provided to illustrate the Airavata capabilities and can be used to customize science-centric gateways



XSEDE

Reference Gateway <-> Airavata Services



SciGaP – Powered by Airavata

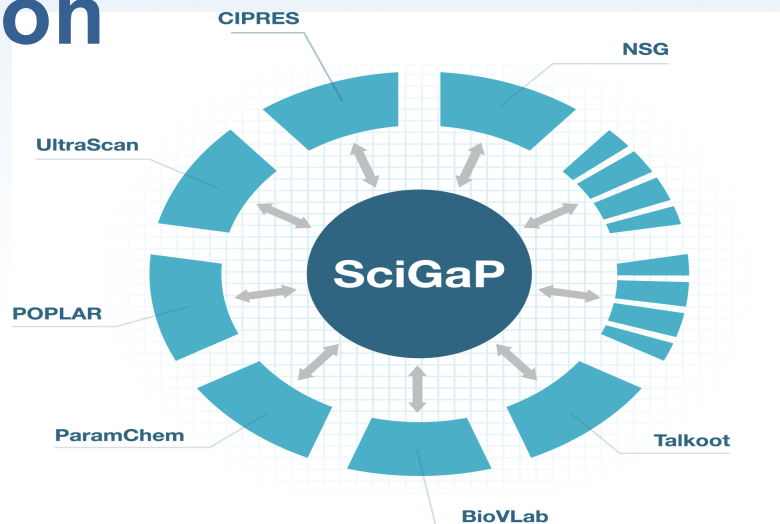
- Science Gateway Platform as a Service (SciGaP) provides application programmer interfaces (APIs) to hosted generic infrastructure services that can be used by domain science communities to create Science Gateways.
- SciGaP hosted service platform is powered by Apache Airavata.
- SciGaP helps gateway developers to concentrate their efforts on building their scientific communities and not worry about operations.



XSEDE

SciGaP Key Mission

Scale number of gateways without having to scale FTE's needed to support them



powered by

Apache Airavata



XSEDE

Challenges for Gateways

- Providing a rich user experience
- Defining an API for the application server
- Defining the right sub-components for the application server.
- Implementing the components, wiring them together correctly.
- Supporting multiple gateway tenants
- Fault tolerance for components
- State management
- Continuous delivery
- Security management
- Supporting full scientific exploratory cycle



XSEDE

Colaboration between SNU bioinformatic institute and IU Airavata group

- SNU Bioinformatic institute developed BioVLAB-MMIA-NGS system to solve biological problem on miRNA mRNA interaction
- Main infrastructure of BioVLAB-MMIA-NGS based on Apache Airavata developed by IU
- SNU takes role of developed, maintain the system and handling users' requests, and in colaboration with IU Airavata group, main functionalities of Apache Airavata infrastructure is improving with considering the requests.



XSEDE

miRNA mRNA Integrated Analysis (MMIA) and system novelty

- miRNA controls mRNA expression by targeting / sequence pairing
- Biological meaning of MMIA
 - MiRNAs have been known to play critical roles in gene regulation
 - miRNA targeting causes differential expression of genes, and aberrant expressed genes have disease association
 - Integrated analysis between miRNAs and mRNA has potential to reveal biological relationships between miRNA expression and their targeting mRNA expression changes, and eventually their biological meaning
- System novelty
 - Handling RNA-seq / microarray data
 - Integrated analysis of miRNA and targeted mRNA
 - Integrated with computational resources (Amazon Cloud, Super computer)
 - Support additional research by providing JAWS based application such as IGV, Cytoscape



XSEDE

BioVLab- MMIA-NGS Result

- miRNA mRNA targeting result as well as various statistical result will be delivered to user
- Network analysis
- Alignment navigation
- Gene set analysis

miRNA mRNA Integrated Analysis Result

Integrated analysis result

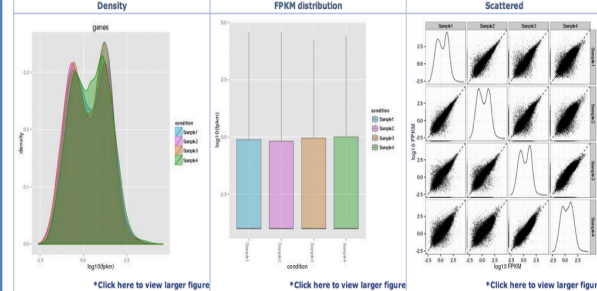
Sample1 vs. Sample2 (Top 100 sorted by pvalue), Download whole table

DEmiRNAs	miRNA	Sample1_count	miRNA	Sample2_count	log2FoldChange	pvalue	Targeted_DEGs	locus	Sample1_FPKM	Sample2_FPKM	log2(fold_change)
hsa-miR-503	3964	264			-1.224342636923520	0.145912638717349	ADAMTS5	21:28261694-2833832	0	0.0065608	inf
hsa-miR-503	3964	264			-1.224342636923520	0.145912638717349	ANLN	7:36363629-36493400	127.566	367.74	1.52745
hsa-miR-503	3964	264			-1.224342636923520	0.145912638717349	CTorf13	7:196264899-19649824	0.521784	2.43434	2.22201
hsa-miR-503	3964	264			-1.224342636923520	0.145912638717349	CCND1	11:69455854-69490184	60.208	567.837	3.23745
hsa-miR-	3964	264			-1.224342636923520	0.145912638717349	CDY77	19:43857810-	0.107406	0.557471	2.37592

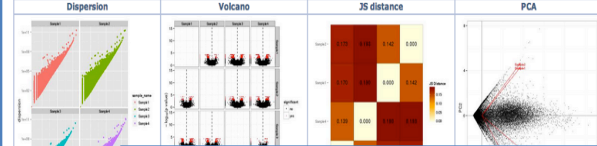
Sample1 vs. Sample3 (Top 100 sorted by pvalue), Download whole table

DEmiRNAs	miRNA	Sample1_count	miRNA	Sample3_count	log2FoldChange	pvalue	Targeted_DEGs	locus	Sample1_FPKM	Sample3_FPKM	log2(fold_change)
hsa-miR-153	8	502			1.03288468584119	0.295088038991734	ACTBL2	5:5660986-56829251	0.057789	0.0196801	-1.55406
hsa-miR-153											
hsa-miR-153											
hsa-miR-											

Expression (FPKM) analysis

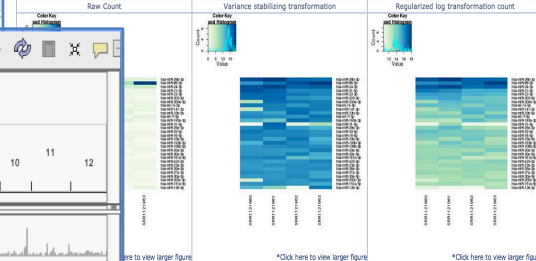


Dispersion, Volcano plot, 3S distance, and PCA

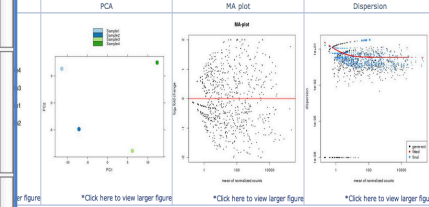


miRNA Analysis Result

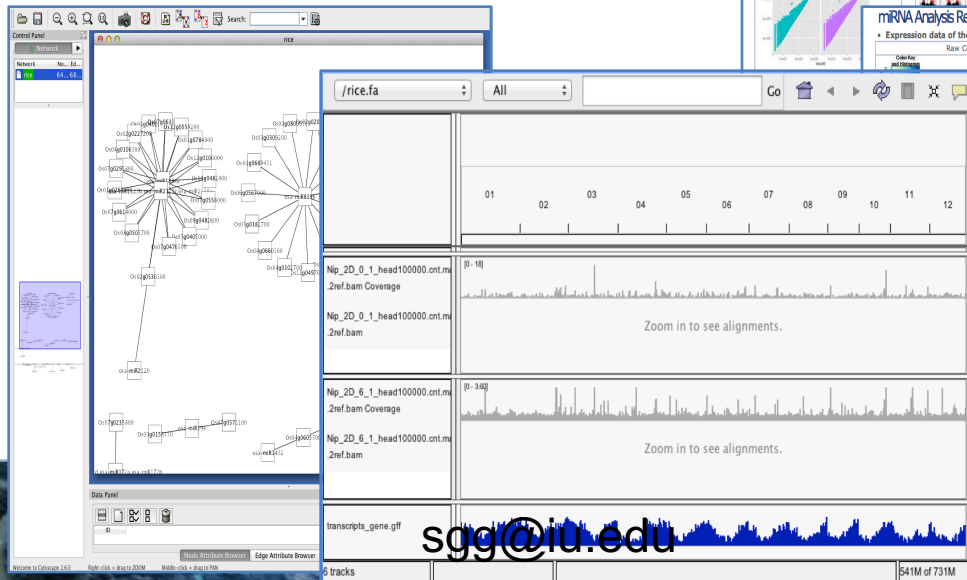
Expression data of the 30 most highly expressed miRNAs



MA and dispersion



RNA Expression Table (All other miRNAs expression table are available : here)



sgg@u.edu



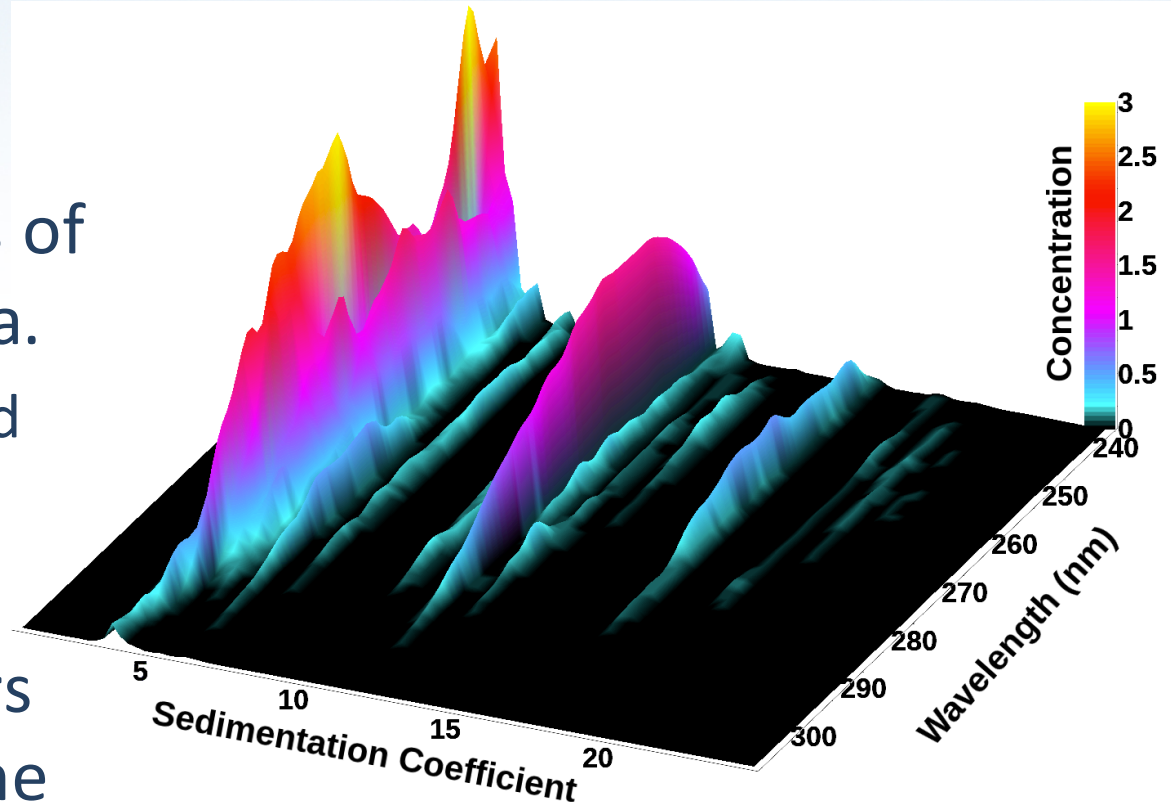
DE

Ultrascan Science Gateway

- Analytical Ultracentrifugation Data Analysis
 - Borries Demeler
 - University of Texas Health Science Center, San Antonio

Revolution in Ultracentrifugation

- New multi-wavelength instruments producing 100's of times more data.
 - Unprecedented accuracy
 - New science
- Supercomputers needed to do the analysis.



Cauma3d Test Database (uslims3_cauma3d)

Home

UltraScan III

UltraScan II

LIMS

SOMO

Wiki

Welcome to the TeraGrid Science Gateway for UltraScan!

This website offers access to the UltraScan Laboratory Information Management System (USLIMS), a [TeraGrid Science Gateway](#) supported by an allocation through a TeraGrid community account. This system provides web and database support for users of the [UltraScan software](#). You can use this portal to access data associated with your sedimentation experiments, and share your data with collaborators. Authorized users can also use this site to model analytical ultracentrifugation experiments with UltraScan's high-performance analysis modules by submitting analysis jobs to computing clusters available at the [University of Texas Health Science Center](#) and [TeraGrid sites at the Texas Advanced Computing Center](#) and at the [community account](#) (see below for [funding director](#)).

DISCLAIMER:

We do not take any responsibility for is your responsibility to always make provided via the [UltraScan mailing list](#) assume all risks involved with placing information placed on this server will a make arrangements for such a service

Funding for this facility is provided through

- Department of Biochemistry, U
- User fees collected from collabor
- San Antonio Life Science Institu
- The National Science Foundation (Demeler)
- The National Institutes of Health

When publishing, please credit our fac

Calculations were performed on the Bioinformatics Core Facility at the U Science Center at San Antonio and the Texas Advanced Computing Center at #MCB070038 (to Borries Demeler)."

Please enter the link to each manuscr

Before logging in, if you have not don will make it easier to use the secure p

Borries Demeler, Ph.D.
Associate Professor
UltraScan Project Director

2DSA Analysis

Initialize 2DSA Parameters - demo1_veloc_rs.RA.2.A.260.auc; Edit profile: 1308301540; Dataset 1 of 1

S-Value Resolution

1	S-Value Minimum
10	S-Value Maximum
60	S-Value Resolution (total grid points)

f/f0 Resolution

1	f/f0 Minimum
4	f/f0 Maximum
60	f/f0 Resolution (total grid points)

Uniform Grid Repetitions Setup

6 Uniform Grid Repetitions

Monte Carlo Iterations

Value: 1 Minimum: 1 Maximum: 100

Fit Time Invariant Noise

- ☐ On
☒ Off

Show Advanced Options

Edit Profiles Change Experiment

Dataset control:

- Current dataset number: 1
- Run Name: demo1_veloc_rs.RA.2.A.260.auc
- Number of datasets: 1

Select Cluster

Cluster	Status	Queue Name	running/queued
<input checked="" type="radio"/> stampede	*	normal	* / *
<input type="radio"/> lonestar	unknown	normal	0 / 0
<input type="radio"/> trestles	*	normal	* / *
<input type="radio"/> juropa	*	default	* / *
<input type="radio"/> alamo	unknown	default	0 / 0
<input type="radio"/> bcf	up	default	0 / 0

Submit

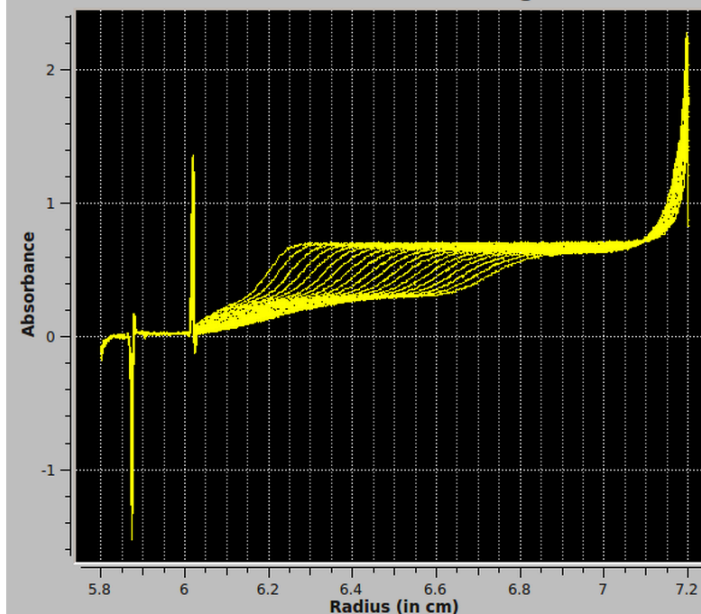
The UltraScan science gateway enables experimental scientists to their analyze data on supercomputers.

UltraScan uses Apache Airavata for managing analyses on HPCs across the world.

Run ID: demo1_veloc_rs
Cell 2, Channel A, Wavelength 260
Edited Dataset:

Raw Scan Data (PNG Plot)
Filename: cnvt.2A260.raw.png

Radial Absorbance Data
Run ID: demo1_veloc_rs
Cell: 2 Channel: A Wavelength: 260



Launch analysis and monitor through a browser

Desktop analysis tools are integrated with the Web portal.

World wide deployments (not just users)

- University of Texas Health Science Center at San Antonio
 - Uses Local cluster (Alamo), Texas (Lonestar), XSEDE (Comet, Stampede, Gordon, Quarry)
- Julich Supercomputing Center, Germany
- Indian Institute of Sciences, Bangalore, India
- La Trobe University, Australia

More gateway Talks:

- gateways@xsede.org Mailing List
- Send email to majordomo@xsede.org
 - with "subscribe gateways" in the body of the message

