

All-In-One Managed AWS Services For Biomedical Research

Vijay Nagarajan PhD, Director, MolBioCloud



DATA

Data discovery, Data cleaning
Data curation, Data management
Data integration, Data visualization



RESOURCES

Tools - Design and development
Database - Design and development
Pipelines - Building and testing



RESEARCH

Funded Collaborations
Private Contracts
Authored Collaborations



TRAINING

On-site, hands-on training
Cutting edge topics
Cost-effective, great content



INFRASTRUCTURE

Cloud setup
Cloud instance maintenance
Cloud instance managing



CONSULTATION

Biomedical informatics
Network medicine
Big-Data



AABI

American Academy for Biomedical Informatics

[AABI](#) [About](#) [Mission](#) [Services](#) [Current Trainings](#) [Login](#) [Member Benefits](#) [Contact](#)

4 Days Hands-on Training, January 10-13, 2017

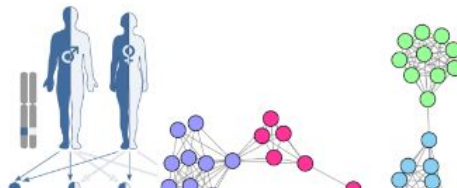
Training in Advanced Transcriptomics (RNASeq) Analysis

@ National Institutes of Health, Bethesda, MD 20892, USA. Earn 2.8 CEU's from FAES@NIH

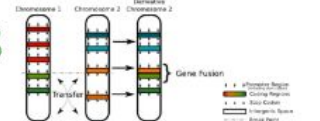


National Institutes
of Health

FAES@NIH
BIOTECHNOLOGY TRAINING WORKSHOPS



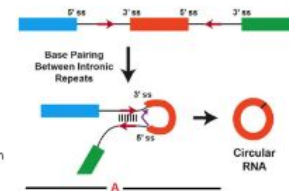
A. Chromosomal Translocation

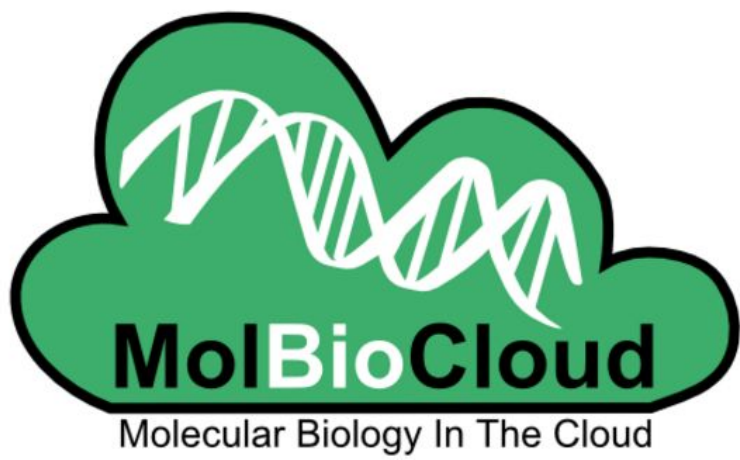


B. Interstitial Deletion



C. Chromosomal Inversion





We Make Cloud Computing For Molecular Biologists ...

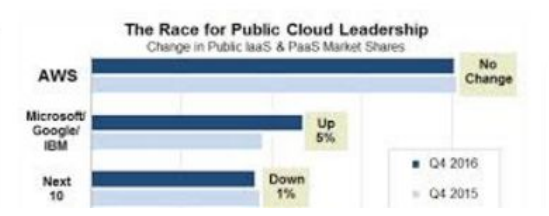
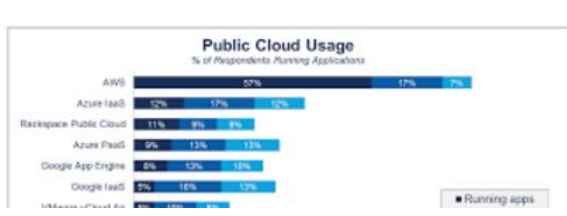
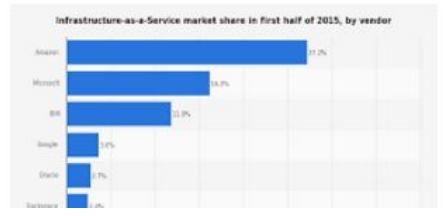
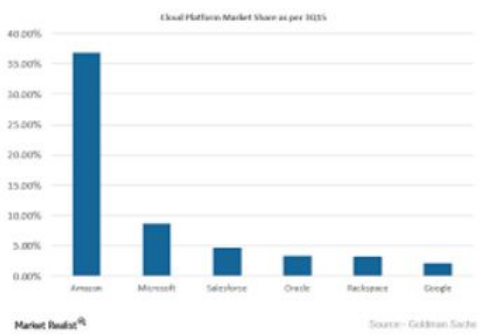
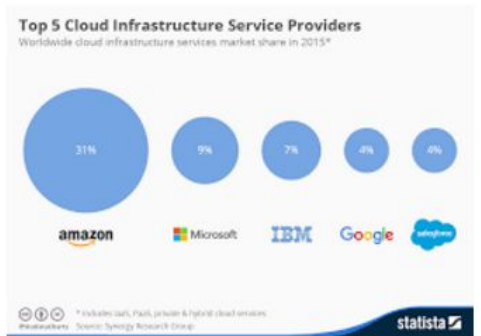
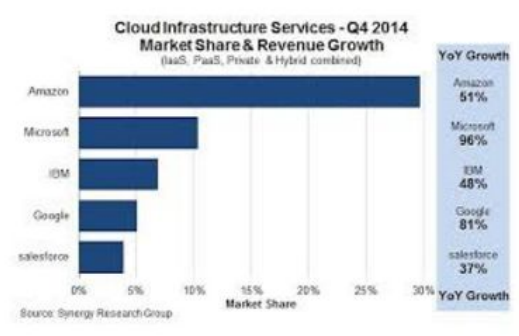
Fun
Easy

Secure
Compliant

Cheap
Powerful

Efficient
Productive

cloud provider mobile o software market pc market computer market social network market ups server os server operati



Why Cloud ?

- **Most secure**
 - securing data/system at several different levels
- **Most reliable**
 - protect data/system against natural/manmade disruptions
- **Most compliant**
 - industry standard compliance
- **Most cost effective**
 - pay-as-you-go, pay-only-for-what-you-use, cheap
- **Most flexible**
 - options for computing, storage, analysis
- **Most powerful**
 - scale from 1 to 100 cores in just one single click

Great things VS Not-So Great things



Compute

EC2
EC2 Container Service
Lightsail
Elastic Beanstalk
Lambda
Batch



Storage

S3
EFS
Glacier
Storage Gateway



Database

RDS
DynamoDB
ElastiCache
Amazon Redshift



Networking & Content Delivery

VPC
CloudFront
Direct Connect
Route 53



Migration

AWS Migration Hub
Application Discovery Service
Database Migration Service



Developer Tools

CodeStar
CodeCommit
CodeBuild
CodeDeploy
CodePipeline
X-Ray



Management Tools

CloudWatch
CloudFormation
CloudTrail
Config
OpsWorks
Service Catalog
Trusted Advisor
Managed Services



Security, Identity & Compliance

IAM
Inspector
Certificate Manager
Directory Service
WAF & Shield
Artifact
Amazon Macie
CloudHSM



Analytics

Athena
EMR
CloudSearch
Elasticsearch Service
Kinesis
Data Pipeline
QuickSight
AWS Glue



Artificial Intelligence

Lex
Amazon Polly
Rekognition
Machine Learning



Internet Of Things

AWS IoT
AWS Greengrass



Contact Center

Amazon Connect



Game Development

Amazon GameLift



Mobile Services

Mobile Hub
Cognito



Application Services

Step Functions
SWF
API Gateway
Elastic Transcoder



Messaging

Simple Queue Service
Simple Notification Service
Simple Email Service



Business Productivity

WorkDocs
WorkMail
Amazon Chime



Desktop & App Streaming

WorkSpaces
AppStream 2.0

Customer Data

Platform, Applications, Identity & Access Management

Operating System, Network & Firewall Configuration

Client-side Data Encryption & Data Integrity Authentication

Server-side Encryption (File System and/or Data)

Network Traffic Protection (Encryption / Integrity / Identity)

Customer



Responsible for security *'in'* the Cloud

Compute

Storage

Database

Networking

AWS



Responsible for security *'of'* the Cloud

AWS Global Infrastructure

Regions

Availability Zones

Edge Locations

We Make Cloud Computing For Molecular Biologists ...

Fun

Easy

- Full Featured Linux Desktops
- No More Terminal Hassles
- Fully Configured
- Anywhere, Anytime, Anyhow
- We Set Up Everything For You
- Just One Click Connection

Secure

Compliant

- SSH Keys Based Login
- Host Based Access
- Encrypted Keys
- Separate Encrypted Data Storage
- FISMA, FedRAMP, etc.
- Access Monitors, Alerts

Cheap

Powerful

- Prices Starting From Pennies/Hour
- Unbeatable Price
- Discounts For Referrals
- 1 CPU To 128 CPU In A Click
- 1 GB RAM To 2 TB RAM In A Click
- Compute In CPU, GPU, Clusters

Efficient

Productive

- 100s Of OpenSource MolBio Software
- Fully Installed, Configured, Tested
- Worry Free Updates, Maintenance
- Free CookBook Style Manuals
- Latest Tools Availability
- No More Computing Limitations

Certifications and Compliance



 <p>amazon web services™</p>	<p>Certified</p> <p>Solutions Architect - Professional</p>
 <p>amazon web services™</p>	<p>Accredited Technology Professional</p>
 <p>amazon web services™</p>	<p>Accredited Business Professional</p>

One-On-One Training

Faculty, Students, Staff
Cloud, Biomedical Informatics

Graphics systems in R

I. Base Graphics

Base package `{graphics}` is the graphics system that was originally developed for S language at Bell Labs in the 1970s. It is loaded by default when R starts.

- `plot()` is the main, high-level function for creating graphics
- `par()` sets-up parameters which control the layout and look of the graphics

1. Set Environment

```
setwd("C:/Users/YO/Documents/R Graphics")  
load('helperFunctions.RData')
```

2. Set Plotting data

Generate random data points ("Variable" - R vector) from standard normal distribution (N=50)


```
set.seed(100); Variable <- rnorm(50)  
set.seed(100); names(Variable) <- randomNames(n=50, length=3)
```

3. Generic X-Y Plot

```
plot(Variable) ## type = 'p'; points
```








 **ngs**
MATE on 34.213.206.106



Login:


Password:


 manager@52.42.214.202


 MATE


 800x600



 Enabled 

 **New session**

 manager@35.162.237.101

 MATE

 800x600

 Enabled 





Computer



fortestmnt



samtools



gencode.v26.
annotation.gff3



installs



manager's
Home



htslib



temp



bbmap



IGV_2.3.80



temptest



Wastebasket



bcftools



macs2



tophat2



Documentation



rnaseq



unix



faesdata



Sample Data



variant



Applications Places Sy

- Accessories >
- Bioinformatics >
- Education >
- Games >
- Graphics >
- Internet >
- Office >
- Other >
- Programming >
- Sound & Video >
- System Tools >
- Universal Access >
- Wine >

Documentation

faesdata

- QTLCart >
- Acacia
- Act
- Arb
- Archaeopteryx
- Artemis
- Clustal X
- Cn3D
- cytoscape
- Dendroscope
- entrez
- FastQC
- Galaxy
- Jalview
- Jemboss
- jMOTU
- JProfilegrid
- mesquite
- Mr Bayes
- njplot
- oligoarray
- ofast

amtools

gencode.v26.
annotation.gff3

FASTQ
cp ne
/usr/

installs

temp

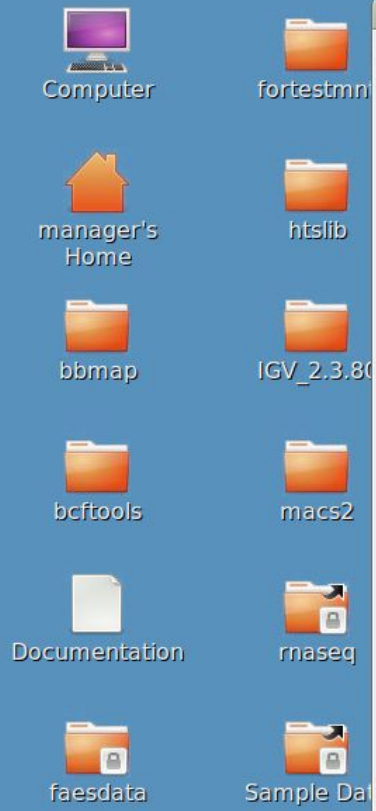
Wastebasket

emptest

tophat2

unix

variant



FastQC

File Help



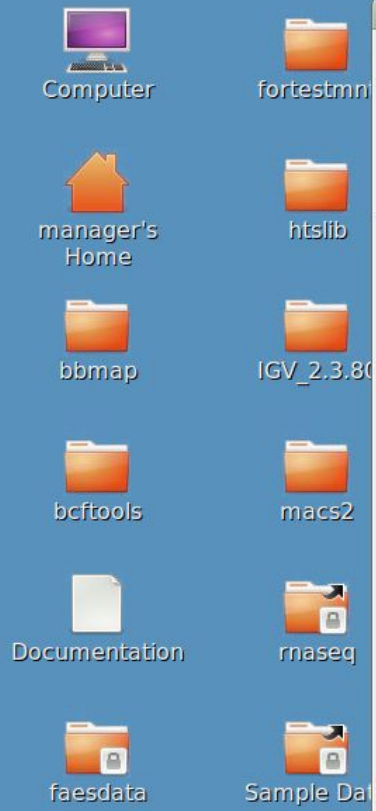
FastQC High Throughput Sequence QC Report

Version: 0.11.5

www.bioinformatics.babraham.ac.uk/projects/

© Simon Andrews, Pierre Lindenbaum, Brian Howard, Phil Ewels 2011-15,
Picard BAM/SAM reader © The Broad Institute, 2013
BZip decompression © Matthew J. Francis, 2011
Base64 encoding © Robert Harder, 2012
Java HDF5 reader © ETH, CISD and SIS, 2007-14


Use File > Open to select the sequence file you want to check



FastQC

File Help

- Open... Ctrl-O
- Save report... Ctrl-S
- Close Ctrl-W
- Close All
- Exit



FastQC High Throughput Sequence QC Report

Version: 0.11.5

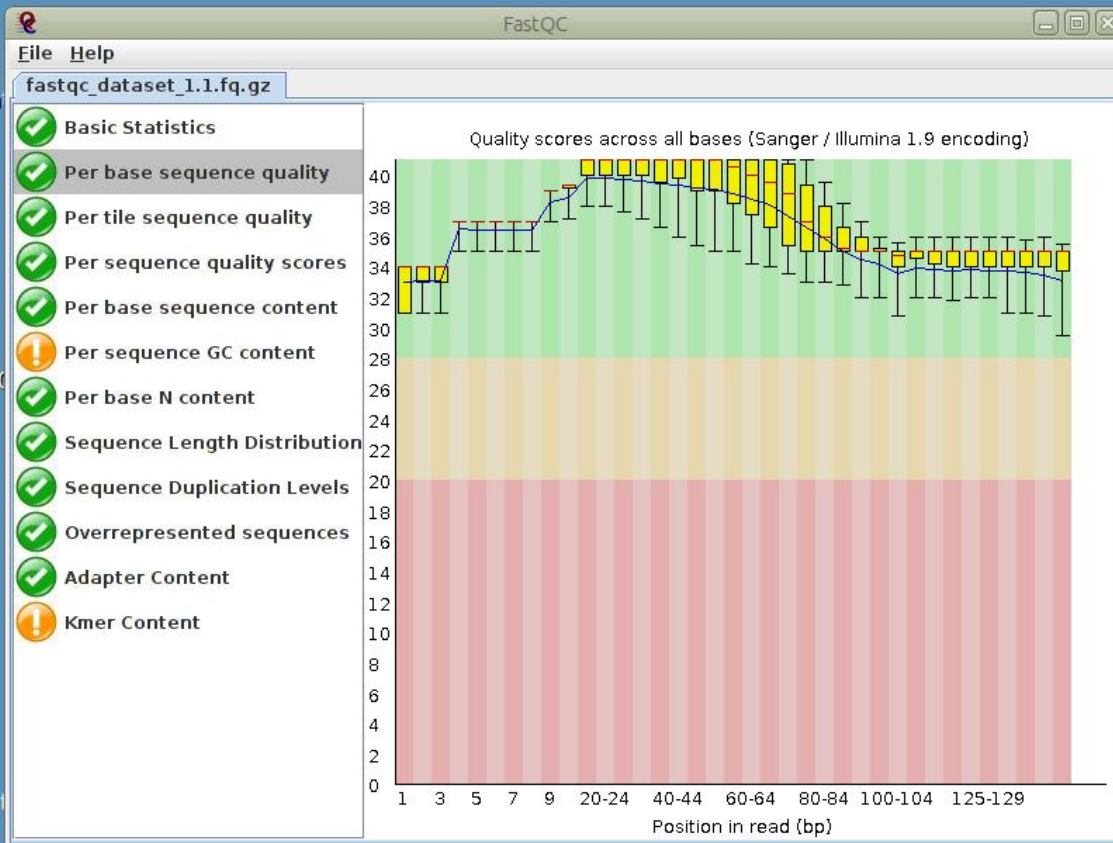
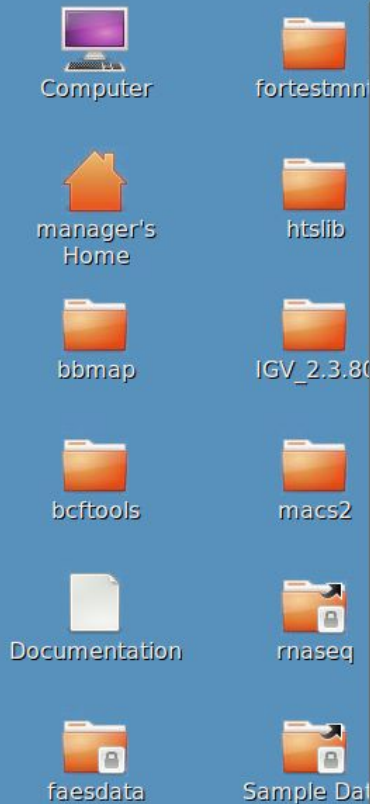
www.bioinformatics.babraham.ac.uk/projects/

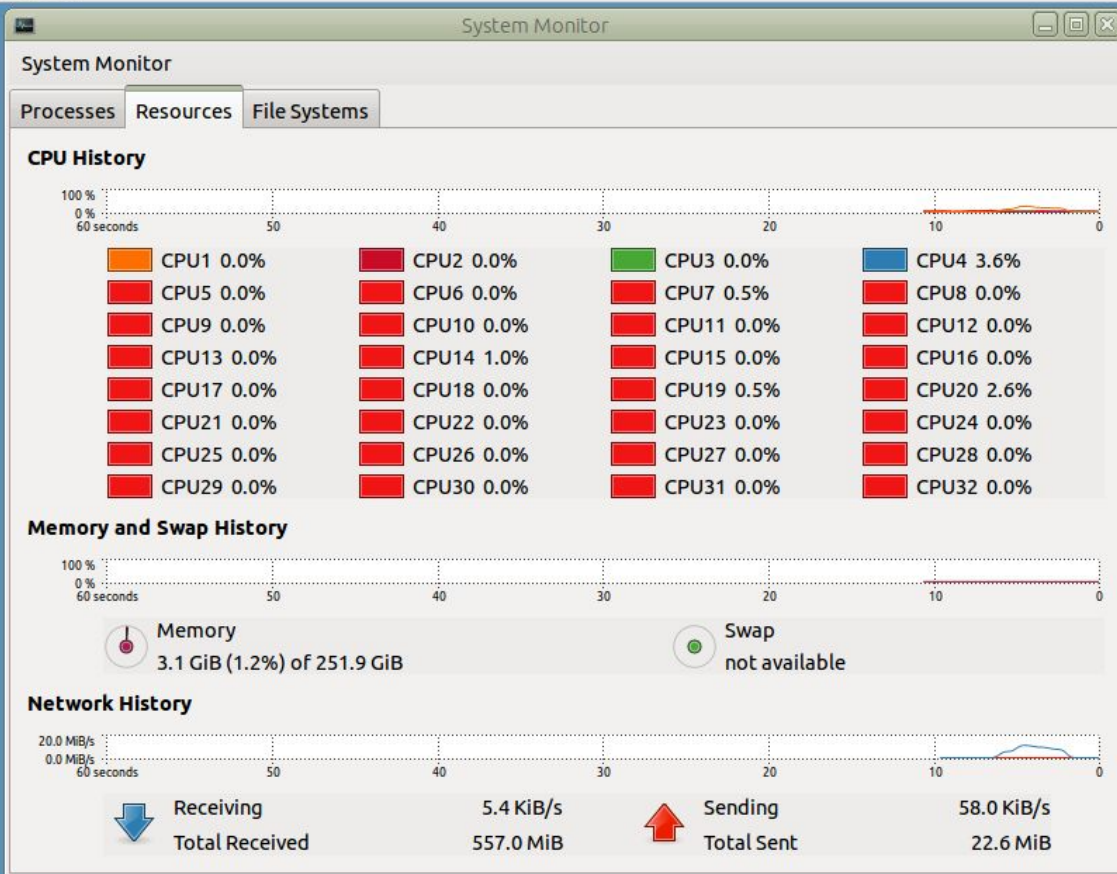
© Simon Andrews, Pierre Lindenbaum, Brian Howard, Phil Ewels 2011-15,
Picard BAM/SAM reader © The Broad Institute, 2013
BZip decompression © Matthew J. Francis, 2011
Base64 encoding © Robert Harder, 2012
Java HDF5 reader © ETH, CISD and SIS, 2007-14

Use File > Open to select the sequence file you want to check

The desktop environment features several icons and folders on a blue background. On the left side, there are icons for 'Computer', 'manager's Home', 'Documentation', and 'faesdata'. On the right side, there are folders named 'fortestmri', 'htslib', 'IGV_2.3.80', 'macs2', and 'Sample Data'. The system tray at the bottom left shows the 'FastQC' application icon.

The 'FastQC' application window is open, displaying a file browser interface. The title bar reads 'FastQC' and the menu bar contains 'File' and 'Help'. An 'Open' dialog box is overlaid on the main window. The dialog box has a title bar with a close button and the text 'Open'. The 'Look In:' field is set to 'input'. The file list shows several folders: 'align', 'annotation', 'dataset_2', 'dataset_3', 'germline', 'somatic', and 'sv'. Two files are selected: 'fastqc_dataset_1.1.fq.gz' and 'fastqc_dataset_1.2.fq.gz'. The 'File Name:' field contains 'fastqc_dataset_1.1.fq.gz' and the 'Files of Type:' dropdown is set to 'Sequence Files'. 'Open' and 'Cancel' buttons are at the bottom right of the dialog box.





Tools

search tools

- [Get Data](#)
- [Send Data](#)
- [Lift-Over](#)
- [Text Manipulation](#)
- [Filter and Sort](#)
- [Join, Subtract and Group](#)
- [Convert Formats](#)
- [Extract Features](#)
- [Fetch Sequences](#)
- [Fetch Alignments](#)
- [Get Genomic Scores](#)
- [Statistics](#)

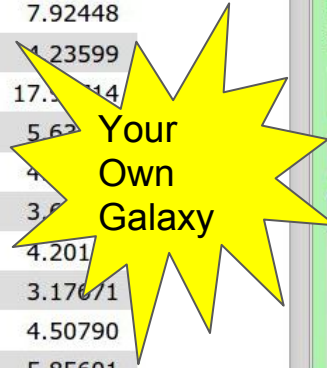
chr1	838163	838164	FoxA1_peak_1	3.93313
chr1	842725	842726	FoxA1_peak_2	3.72141
chr1	848792	848793	FoxA1_peak_3	3.86568
chr1	858478	858479	FoxA1_peak_4	4.79018
chr1	930317	930318	FoxA1_peak_5	3.17671
chr1	930871	930872	FoxA1_peak_6	6.04711
chr1	999057	999058	FoxA1_peak_7	7.92448
chr1	1040561	1040562	FoxA1_peak_8	4.23599
chr1	1050176	1050177	FoxA1_peak_9	17.5014
chr1	1121837	1121838	FoxA1_peak_10	5.62
chr1	1135741	1135742	FoxA1_peak_11	4.201
chr1	1183103	1183104	FoxA1_peak_12	3.86568
chr1	1239597	1239598	FoxA1_peak_13	4.201
chr1	1443405	1443406	FoxA1_peak_14	3.17671
chr1	1527422	1527423	FoxA1_peak_15	4.50790
chr1	1674837	1674838	FoxA1_peak_16	5.85601
chr1	1684463	1684464	FoxA1_peak_17	11.05453

History

Unnamed history
44.3 KB

1: 1000summits.bed
1,000 regions
format: interval, database: ?
uploaded interval file

1.Chrom	2.Start	3.End	4
chr1	838163	838164	FoxA1_peak_1
chr1	842725	842726	FoxA1_peak_2
chr1	848792	848793	FoxA1_peak_3
chr1	858478	858479	FoxA1_peak_4



- Administration**
 - View data tables registry
 - View tool lineage
 - Reload a tool's configuration
 - Profile memory usage
 - Manage jobs
 - Review tool migration stages
- Tool sheds**
 - Search and browse tool sheds
- Form Definitions**
 - Manage form definitions
- Sample Tracking**
 - Manage sequencers and external services

Name	Description	Valid repositories
Assembly	Tools for working with assemblies	81
ChIP-seq	Tools for analyzing and manipulating ChIP-seq data.	44
Combinatorial Selections	Tools for combinatorial selection	7
Computational chemistry	Tools for use in computational chemistry	29
Constructive Solid Geometry	Tools for constructing and analyzing 3-dimensional shapes and their properties	11
Convert Formats	Tools for converting data formats	76
Data Export	Tools for exporting data to various destinations	1
Data Managers	Utilities for Managing Galaxy's built-in data cache	36
Data Source	Tools for retrieving data from external data sources	53
Entomology	Tools that involve insect studies	

```
manager@bl8vbox[~]  
File Edit View Search Terminal Help  
manager@bl8vbox[manager] nproc [ 7:23PM]  
64  
manager@bl8vbox[manager] [ 7:24PM]
```

Freedom...
all the way to
the shell

faesdata Sample Data variant



Enter search term...

Control Panel

Network Style Select

Network	Nodes	Edges
<ul style="list-style-type: none"> <ul style="list-style-type: none"> BIOGRID-ORGANISM-Escherichia_coli-3 BIOGRID-ORGANISM-Escherichi 	136(0)	126(0)

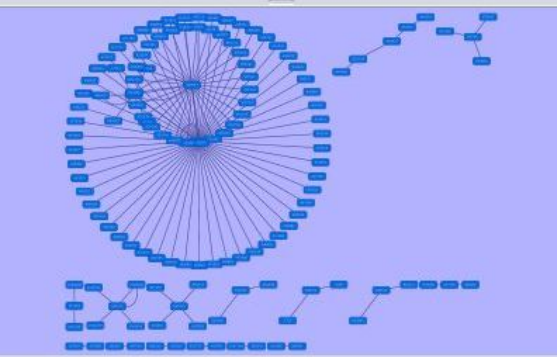
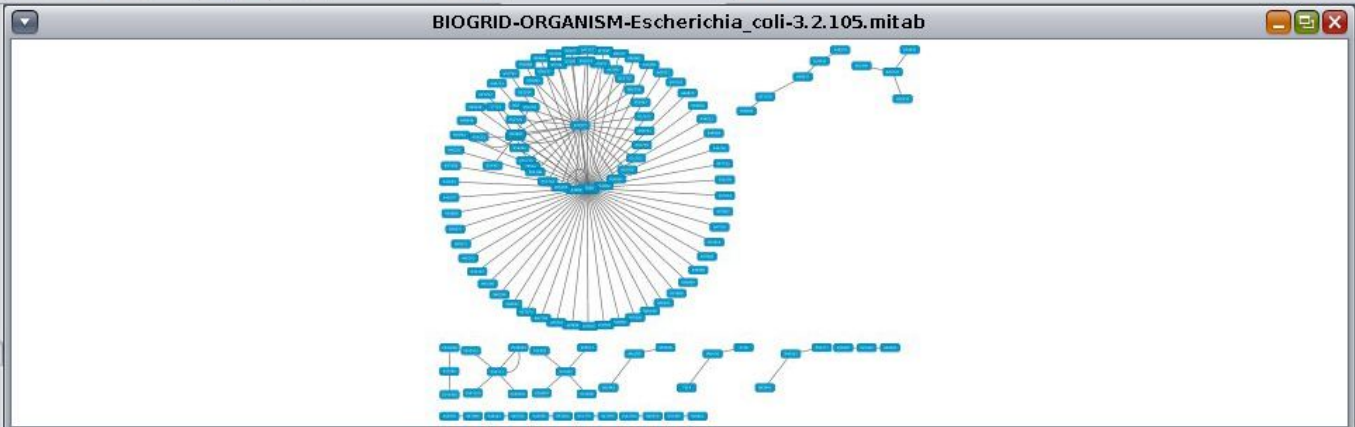


Table Panel

Settings, View, Print, Delete, Refresh, $f(x)$

share...	name	PSIMI-...	PSIMI-...	PSIMI-...	PSIMI-...	entrez...	psi-mi
852495	852495		[85249...	[559292]	[taxid]	852495	"MI:046...
948461	948461		[94846...	[511145]	[taxid]	948461	
2541704	2541704		[25417...	[4896]	[taxid]	2541704	"MI:046...
948559	948559		[94855...	[511145]	[taxid]	948559	

Node Table Edge Table Network Table

Memory: 0K

R demos x R demos x

Explore some properties of R objects and is.FOO() functions. Not for newbies!

recursion Using recursion for adaptive integration

scoping An illustration of lexical scoping.

Demos in package 'graphics':

Hershey Tables of the characters in the Hershey vector fonts

Japanese Tables of the Japanese characters in the Hershey vector fonts

graphics A show of some of R's graphics capabilities

image The image-like graphics builtins of R

persp Extended persp() examples

plotmath Examples of the use of mathematics

Console

```
> ## A little color wheel. This code just plots equally spaced hues in
> ## a pie chart. If you have a cheap SVGA monitor (like me) you will
> ## probably find that numerically equispaced does not mean visually
> ## equispaced. On my display at home, these colors tend to cluster at
> ## the RGB primaries. On the other hand on the SGI Indy at work the
> ## effect is near perfect.
>
> par(bg = "gray")
>
> pie(rep(1,24), col = rainbow(24), radius = 0.9)
Hit <Return> to see next plot: |
```

Environment History

Global Environment

Values

opar	List of 1
x	num [1:50] -0.515 -1.045 0.426 -1.563 -0.67 ...



Not Just Another Vendor

Cloud Complete Services

- Design
- Build
- Analyze
- Visualize
- Archive
- Publish

Not Just “omics”

- Image Analysis
- Data Science
- Algorithm Development
- Evolution
- Epidemiological
- 3D Structure
- Networks
- Prediction Models

MolBioCloud Guarantee

- 1 hr initial response
- Unbeatable Price
- Unbeatable Customer Service - Dedicated Project Manager
- Free Setup
- Free Monthly/Weekly/Daily Data Backups (rotated monthly, storage cost)
- Free Annual Bioinformatics Software Updates
- Free New Open Source software Installs at Request
- Virus-protected, Secure, Compliant, Logged, Archived system

Let Us Help You Do Amazing Research

Thanks

contact@molbiocloud.com

Example Compute Cost Analysis (1 Year)

Base system running 24x7 \$3504

M4.2xlarge (8,32)

Research system (4 Months) \$9216

M4.16xlarge (64,256)

Standard support \$588

Total \$13308