

Next Generation Sequencing Tutorial

Andreas Gisel
International Institute of Tropical Agriculture (IITA)
Ibadan, Nigeria

a.gisel@cgiar.org



Working Platform

GALAXY

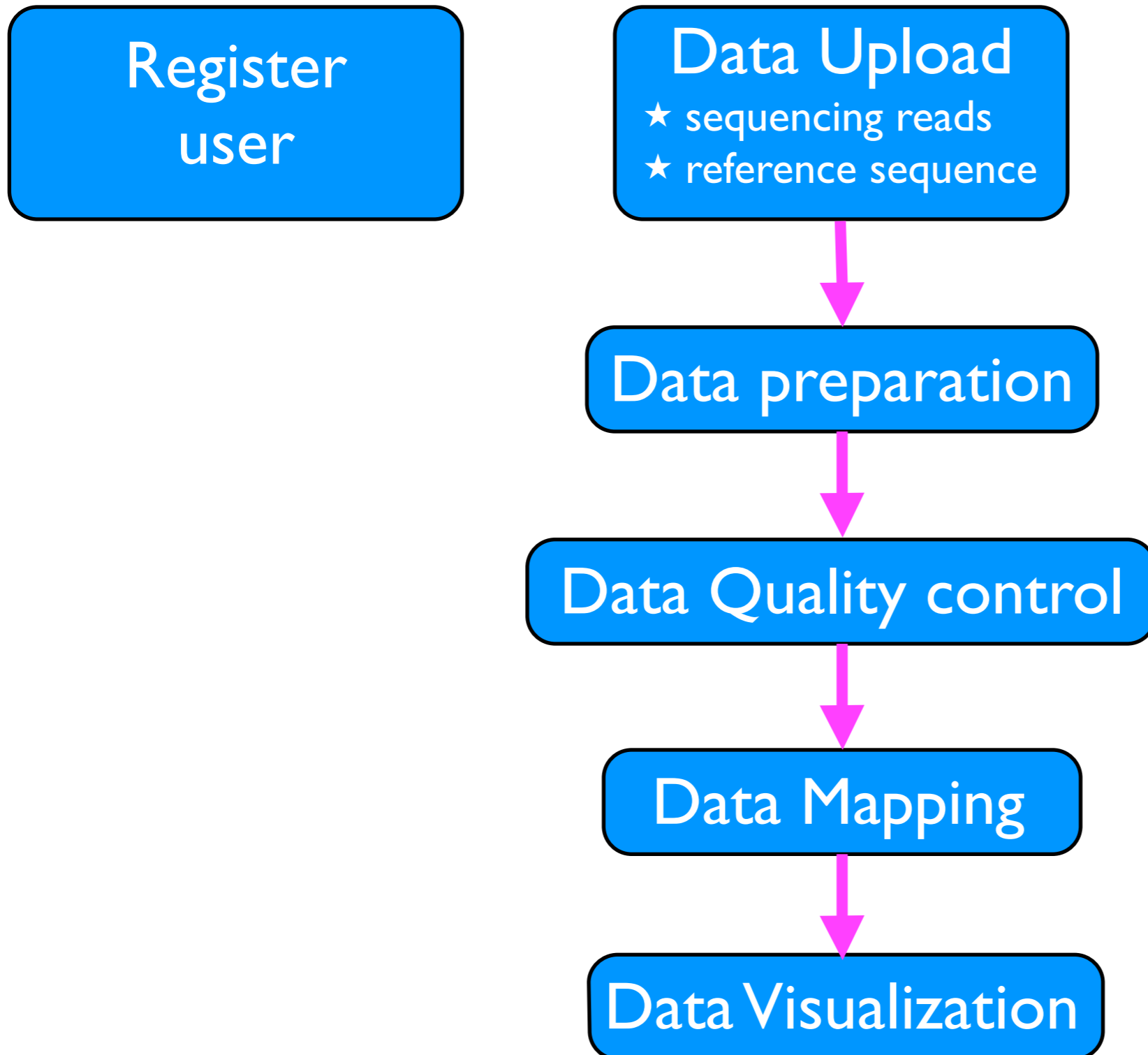
- <http://main.g2.bx.psu.edu/> (Penn Stat Uni)
- eBioKit

GALAXY

The image shows a screenshot of the Galaxy web interface running in a Firefox browser. The browser's address bar shows the URL <http://main.g2.bx.psu.edu/>. The page features a dark navigation bar with tabs for 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. On the left, a 'Tools' sidebar lists various bioinformatics tools such as 'Get Data', 'Send Data', 'ENCODE Tools', 'Lift-Over', 'Text Manipulation', 'Convert Formats', 'FASTA manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Extract Features', 'Fetch Sequences', 'Fetch Alignments', 'Get Genomic Scores', 'Operate on Genomic Intervals', 'Statistics', 'Graph/Display Data', 'Regional Variation', 'Multiple regression', 'Multivariate Analysis', 'Evolution', 'Metagenomic analyses', 'Human Genome Variation', and 'EMBOSS'. The main content area is titled 'Here is what's happening...' and features a large central box with the text 'Managing Data Store, Manage, and Share data with Libraries An in-depth tutorial'. Below this, a 'Live Quickies' section displays four interactive cards: 'Illumina mapping: Paired Ends', 'Basic fastQ manipulation', 'Advanced fastQ manipulation', and '454 Mapping: Single End'. On the right, a 'History' sidebar contains a message: 'Your history is empty. Click 'Get Data' on the left pane to start'. At the bottom of the page, there is a footer with the text: 'The Galaxy team is a part of BX at Penn State. This project is supported in part by NSF, NHGRI, The Huck Institutes of the Life Sciences, and The Institute for CyberScience at Penn State. Galaxy build: \$Rev: 5265aef9c73836f35'. The browser's status bar at the bottom shows the URL <http://bitbucket.org/galaxy/galaxy-central/wiki/DataLibraries/Tutorial/DataLibrariesSampleTracking>.

GALAXY

Workflow



GALAXY

Data upload

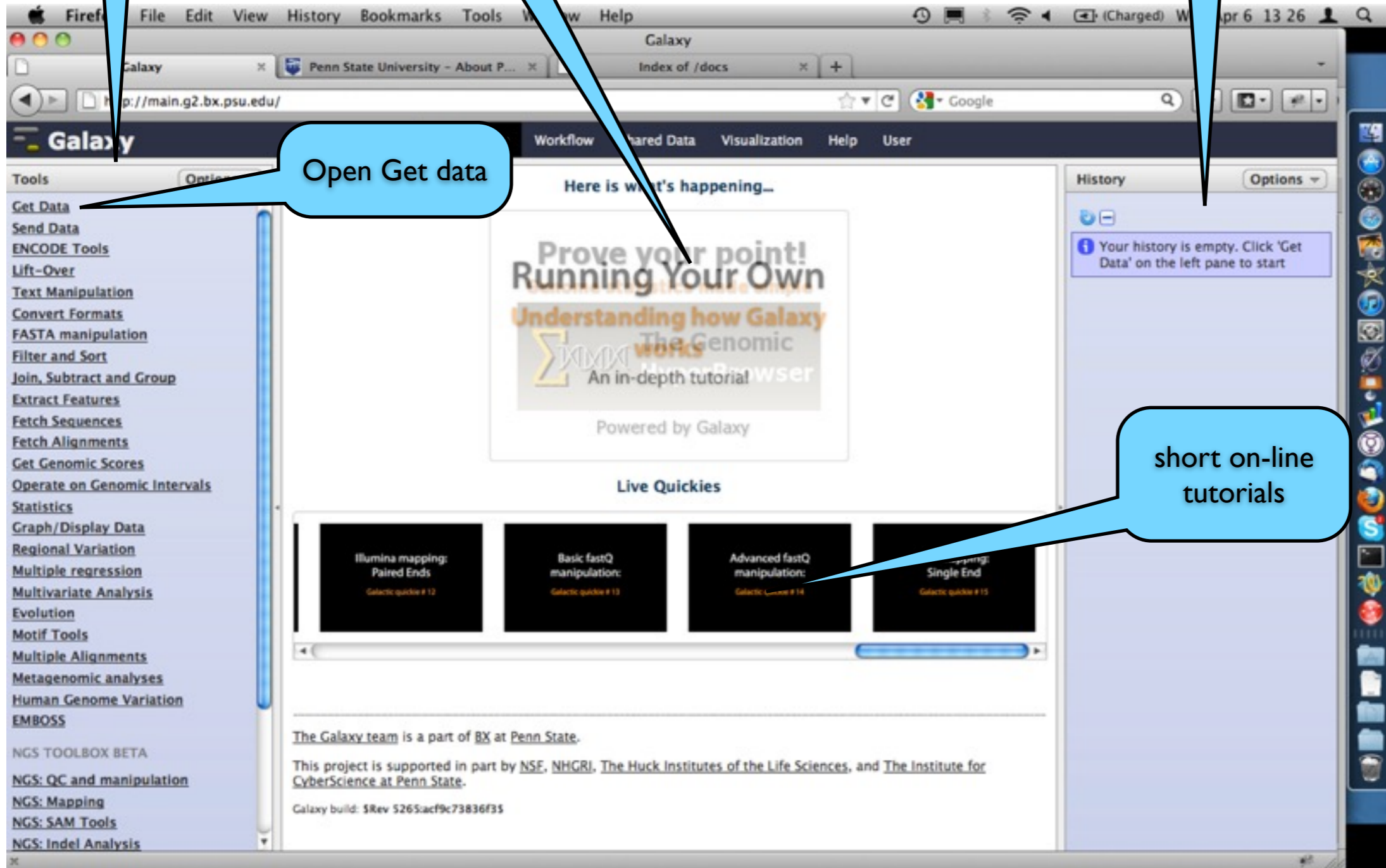
Tool Box

Display

Workbench

Open Get data

short on-line tutorials



Data Upload
★ sequencing reads
★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

GALAXY

Data upload

The screenshot shows the Galaxy web interface. At the top, there is a navigation bar with tabs for 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. The 'User' tab is active, showing a dropdown menu with 'Login' and 'Register' options. A blue callout box labeled 'Register' points to the 'Register' option. Below the navigation bar, there is a green status bar with a checkmark and the text 'Hello world! It's running...'. To the right, there is a 'History' panel showing 'Unnamed history' with a size of 2.9 MB. Below the status bar, there is a text area with the following text: 'Galaxy is an open, web-based platform for data intensive biomedical research. The Galaxy team is a part of BX at Penn State, and the Bioinformatics and Computer Science departments at Emory University. The Galaxy Project is supported in part by NHGRI, NSF, The Huck Institute of the Life Sciences, The Institute for CyberScience at Penn State, and Emory University.'

Create account

Email address:

Password:

Confirm password:

Public name:

Your public name is an identifier that will be used to generate addresses for information you share publicly. Public names must be at least four characters in length and contain only lower-case letters, numbers, and the '-' character.

Register user

Data Upload
★ sequencing reads
★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

GALAXY

Data upload

Firefox File Edit View History Bookmarks Tools Window Help

Galaxy

Galaxy

http://main.g2.bx.psu.edu/

Galaxy

Analyze Data Workflow Shared Data Visualization Help User

Tools Options

Get Data

- Upload File from your computer
- UCSC Main table browser
- UCSC Archaea table browser
- BX_main browser
- BioMart Central server
- GrameneMart Central server
- Flymine server
- modENCODE fly server
- modENCODE modMine server
- Ratmine server
- YeastMine server
- modENCODE worm server
- Wormbase server
- EuPathDB server
- EncodeDB at NHGRI
- EpiGRAPH server

Send Data

ENCODE Tools

Lift-Over

Text Manipulation

Convert Formats

FASTA manipulation

Filter and Sort

Join, Subtract and Group

Extract Features

Fetch Sequences

choose Upload file

Galaxy 2011 Community Conference

25-26 May Lunteren, The Netherlands

Register now!

Live Quickies

- Illumina mapping: Paired Ends
- Basic fastQ manipulation:
- Advanced fastQ manipulation:
- 454 Mapping: Single End

History Options

Your history is empty. Click 'Get Data' on the left pane to start

Data Upload

- ★ sequencing reads
- ★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

GALAXY

Data upload

The screenshot shows the Galaxy web interface with the 'Upload File' tool selected. The 'File Format' dropdown is set to 'Auto-detect'. The 'File' field contains the path '/Users/andreas/Downloads' and the filename 'mac-l8_noad.fastq'. The 'Execute' button is visible at the bottom of the tool panel.

Data Upload
★ sequencing reads
★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

GALAXY

Data upload

Upload in queue

The screenshot shows the Galaxy web interface in a Firefox browser window. The address bar displays `http://main.g2.bx.psu.edu/`. The main content area features a large purple information box with the following text:

i Your upload has been queued. History entries that are still uploading will be blue, and turn green upon completion.

Please do not use your browser's "stop" or "reload" buttons until the upload is complete, or it may be interrupted.

You may safely continue to use Galaxy while the upload is in progress. Using "stop" and "reload" on pages other than Galaxy is also safe.

The left sidebar contains a 'Tools' menu with categories like 'Get Data', 'Send Data', and 'ENCODE Tools'. The right sidebar shows a 'History' panel with a single entry: `_ 22: mac-18_noad.fastq`.

Data Upload

★ sequencing reads
★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

GALAXY

Data upload

Upload in process

Data Upload

★ sequencing reads
★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

The screenshot shows the Galaxy web interface in a Firefox browser window. The address bar displays `http://main.g2.bx.psu.edu/`. The main navigation bar includes **Analyze Data**, **Workflow**, **Shared Data**, **Visualization**, **Help**, and **User**. On the left, a **Tools** sidebar lists various data sources and manipulation tools. The central workspace features a blue information box with the following text:

i Your upload has been queued. History entries that are still uploading will be blue, and turn green upon completion.

Please do not use your browser's "stop" or "reload" buttons until the upload is complete, or it may be interrupted.

You may safely continue to use Galaxy while the upload is in progress. Using "stop" and "reload" on pages other than Galaxy is also safe.

On the right, the **History** panel shows a single entry: `22: mac-18_noad.fastq`, which is highlighted in yellow. A blue callout bubble with the text "Upload in process" points to this entry.

GALAXY

Data upload

Upload done

Data Upload

★ sequencing reads
★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

The screenshot shows the Galaxy web interface in a Firefox browser window. The address bar displays `http://main.g2.bx.psu.edu/`. The main content area features a blue information box with the following text:

i Your upload has been queued. History entries that are still uploading will be blue, and turn green upon completion.

Please do not use your browser's "stop" or "reload" buttons until the upload is complete, or it may be interrupted.

You may safely continue to use Galaxy while the upload is in progress. Using "stop" and "reload" on pages other than Galaxy is also safe.

The History panel on the right shows a single entry: `22: mac-18_noad.fastq`, which is highlighted in green. A blue callout bubble with the text "Upload done" points to this entry.

The left sidebar contains a "Tools" menu with various categories:

- Get Data
 - Upload File from your computer
 - UCSC Main table browser
 - UCSC Archaea table browser
 - BX main browser
 - BioMart Central server
 - GrameneMart Central server
 - Flymine server
 - modENCODE fly server
 - modENCODE modMine server
 - Ratmine server
 - YeastMine server
 - modENCODE worm server
 - Wormbase server
 - EuPathDB server
 - EncodeDB at NHGRI
 - EpiGRAPH server
- Send Data
- ENCODE Tools
- Lift-Over
- Text Manipulation
- Convert Formats
- FASTA manipulation
- Filter and Sort
- Join, Subtract and Group
- Extract Features
- Fetch Sequences

GALAXY

Data upload

Click title for summary

Firefox File Edit View History Bookmarks Tools Window Help

Galaxy

http://main.g2.bx.psu.edu/

Galaxy Analyze Data Workflow Shared Data Visualization Help User

Tools Options

Get Data

- Upload File from your computer
- UCSC Main table browser
- UCSC Archaea table browser
- BX_main browser
- BioMart Central server
- GrameneMart Central server
- Flymine server
- modENCODE fly server
- modENCODE modMine server
- Ratmine server
- YeastMine server
- modENCODE worm server
- Wormbase server
- EuPathDB server
- EncodeDB at NHGRI
- EpiGRAPH server

Send Data

ENCODE Tools

Lift-Over

Text Manipulation

Convert Formats

FASTA manipulation

Filter and Sort

Join, Subtract and Group

Extract Features

javascript:void(0);

History Options

22: mac-18_noad.fastq 2.7 Mb format: fastq, database: ? Info: uploaded fastq file

```
@HWI-EAS210R_0008:6:107:16795:4771#NORON
GAGAAAAGACGACGGTAGACAGA
+HWI-EAS210R_0008:6:107:16795:4771#NORON
edfc`fel`fdeedboeYedBeb
@HWI-EAS210R_0008:6:107:16797:18323#NORC
ATGTGTTTTAAATGCAGTTGTAGT
```

Data Upload

★ sequencing reads
★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

GALAXY

Data upload

Click Eye for display

Click Cross to delete data

Data Upload
★ sequencing reads
★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

The screenshot shows the Galaxy web interface in a Firefox browser. The main content area displays a FASTQ dataset with a warning: "This dataset is large and only the first megabyte is shown below." The dataset name is "22: mac-18_noad.fastq". The interface includes a "Tools" sidebar on the left with categories like "Get Data", "Send Data", and "ENCODE Tools". A "History" panel on the right shows the current dataset. A notification at the bottom indicates "1 new message from Antonio Carrieri".

GALAXY

Data upload

Uploading
2 Illumina read files
mac-18_noad.fastq
mac18red.fastq

2 ref seq
PLMV_PC-C40.fasta
peach_chloro.fa

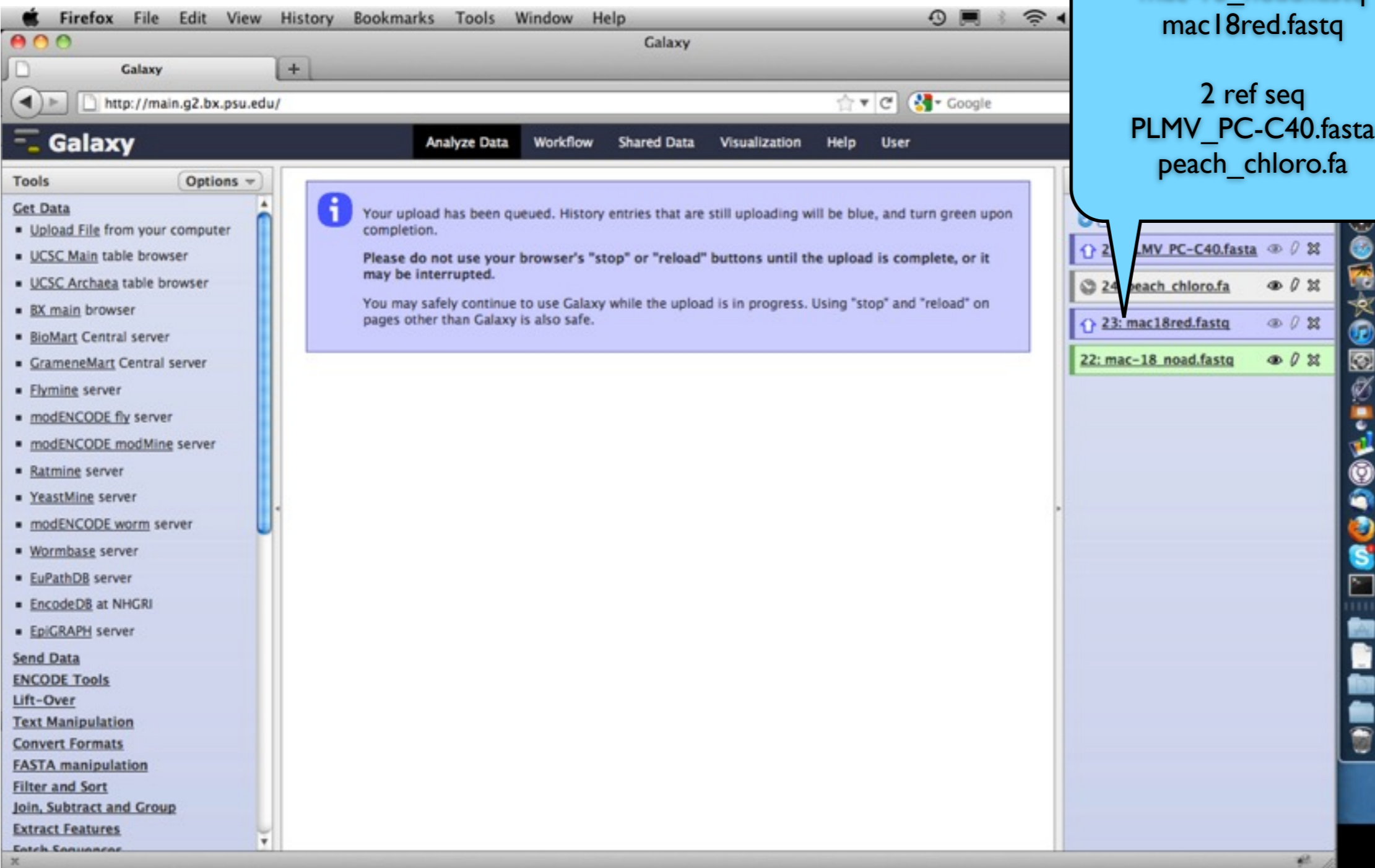
Data Upload
★ sequencing reads
★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization



GALAXY

Data upload

2 fastq files
2 fasta files

Data Upload

★ sequencing reads
★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

Firefox File Edit View History Bookmarks Tools Window Help

Galaxy

http://main.g2.bx.psu.edu/

Galaxy Analyze Data Workflow Shared Data Visualization Help User

Tools Options

Get Data

- Upload File from your computer
- UCSC Main table browser
- UCSC Archaea table browser
- BX main browser
- BioMart Central server
- GrameneMart Central server
- Flymine server
- modENCODE fly server
- modENCODE modMine server
- Ratmine server
- YeastMine server
- modENCODE worm server
- Wormbase server
- EuPathDB server
- EncodeDB at NHGRI
- EpiGRAPH server

Send Data

ENCODE Tools

Lift-Over

Text Manipulation

Convert Formats

FASTA manipulation

Filter and Sort

Join, Subtract and Group

Extract Features

Fetch Sequences

>gi | 32139935 | PC-C40
CTCAAAGTTTCGTCGCATCTCAGCAACTCATCAGTGGGCTTAGCCAGACTTTTGAGAGATTAAGACC
TCTCAGCCCCCTCCACCTTGGGGTGCCTATCCGAGCAC TGCAGTTTCGGTAGAAAGCCTAAGCACCTCC
AAGGAGGTAAAGTGGGACTTTTCCTTCGGGAACCAAGCGGTTGGTCCGAGGGGGGTGTGATCCAGGTAC
CGCCGTAGAAACTGGATTACGACGTCTACCCGGGATTCAAACCCGGTCCCTCCAGAAGTGAATTCGGA
AGAAGAGTCTGTGCTAAGCACACTGATGAGTCTCTGAAATGAGACGAAACTCTTAAAGAGCTTTTGTCC
TCAAAAGTTTCGTCGCATCTCAGCAACTCATCAGTGGGCTTAGCCAGACTTTTGAGAGATTAAGACC

History Options

25: PLMV_PC-C40.fasta 0 0

1 sequences
format: fasta, database: ?
Info: uploaded fasta file

>gi | 32139935 | PC-C40
CTCAAAGTTTCGTCGCATCTCAGCAACTCATCAGTGGGCTTAGCCAGACTTTTGAGAGATTAAGACC
TCTCAGCCCCCTCCACCTTGGGGTGCCTATCCGAGCAC TGCAGTTTCGGTAGAAAGCCTAAGCACCTCC
AAGGAGGTAAAGTGGGACTTTTCCTTCGGGAACCAAGCGGTTGGTCCGAGGGGGGTGTGATCCAGGTAC
CGCCGTAGAAACTGGATTACGACGTCTACCCGGGATTCAAACCCGGTCCCTCCAGAAGTGAATTCGGA
AGAAGAGTCTGTGCTAAGCACACTGATGAGTCTCTGAAATGAGACGAAACTCTTAAAGAGCTTTTGTCC
TCAAAAGTTTCGTCGCATCTCAGCAACTCATCAGTGGGCTTAGCCAGACTTTTGAGAGATTAAGACC

24: peach_chloro.fa 0 0

1 sequences
format: fasta, database: ?
Info: uploaded fasta file

>gi | 309321413 | gb | BQ336405.1 | Prunus pe
TGGGCGAACGACGGGAATGAAACCCGGCATGGTGGAT
ATCCGGCCCTTATACATTACAAATATTACACCATTT
ATAAAGTGAACCTTTATATTTTAAATTAATTTGTA
CAATATAGTAAAGTTAAGTAGTAAATAAAAAATACT
TATAACAGAAATTTATTTGCTCCTTACTTTCAG

23: mac18red.fastq 0 0

22: mac-18_noad.fastq 0 0

GALAXY

Data upload

2 fastq files
2 fasta files

Data Upload

★ sequencing reads
★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

Firefox File Edit View History Bookmarks Tools Window Help

Galaxy

http://main.g2.bx.psu.edu/

Galaxy Analyze Data Workflow Shared Data Visualization Help User

Tools Options

Get Data

- Upload File from your computer
- UCSC Main table browser
- UCSC Archaea table browser
- BX main browser
- BioMart Central server
- GrameneMart Central server
- Flymine server
- modENCODE fly server
- modENCODE modMine server
- Ratmine server
- YeastMine server
- modENCODE worm server
- Wormbase server
- EuPathDB server
- EncodeDB at NHGRI
- EpiGRAPH server

Send Data

ENCODE Tools

Lift-Over

Text Manipulation

Convert Formats

FASTA manipulation

Filter and Sort

Join, Subtract and Group

Extract Features

Fetch Sequences

>gi|32139935|PC-C40
CTCAAAGTTTCGTCGCATCTCAGCAACTCATCAGTGGGCTTAGCCAGACTTTTGAGAGATTAAGACC
TCTCAGCCCCCTCCACCTTGGGGTGCCTATCCGAGCACGCACTTTCGGTAGAAAGCCTAAGCACCTCC
AAGGAGGTAAAGTGGGACTTTTCCTTCGGGAACCAAGCGGTTGGTCCGAGGGGGGTGTGATCCAGGTAC
CGCCGTAGAAACTGGATTACGACGCTACCCGGGATTCAAACCCGGTCCCTCCAGAAGTGAATTCGGA
AGAAGAGTCTGTGCTAAGCACACTGATGAGTCTCTGAAATGAGACGAAACTCTTAAAGAGCTTTTGTCC
TCAAAAGTTTCGTCGCATCTCAGCAACTCATCAGTGGGCTTAGCCAGACTTTTGAGAGATTAAGACC

History Options

25: PLMV_PC-C40.fasta 0 0

1 sequences
format: fasta, database: ?
Info: uploaded fasta file

>gi|32139935|PC-C40
CTCAAAGTTTCGTCGCATCTCAGCAACTCATCAGTGGGCTTAGCCAGACTTTTGAGAGATTAAGACC
TCTCAGCCCCCTCCACCTTGGGGTGCCTATCCGAGCAC
AAGGAGGTAAAGTGGGACTTTTCCTTCGGGAACCAAGCGGTTGGTCCGAGGGGGGTGTGATCCAGGTAC
CGCCGTAGAAACTGGATTACGACGCTACCCGGGATTCAAACCCGGTCCCTCCAGAAGTGAATTCGGA
AGAAGAGTCTGTGCTAAGCACACTGATGAGTCTCTGAAATGAGACGAAACTCTTAAAGAGCTTTTGTCC

24: peach_chloro.fa 0 0

1 sequences
format: fasta, database: ?
Info: uploaded fasta file

>gi|309321413|gb|BQ336405.1| Prunus pe
TGGGCGAACGACGGGAATGAAACCCGGCATGGTGGAT
ATCCGGCCCTTATACATTACAAATATTACACCATTT
ATAAAGTGAACCTTTATATTTTAAATTAATTTGTA
CAATATAGTAAAGTTAAGTAGTAAATAAAAAATACT
TATAACAGAAATTTATTTGCTCCTTACTTTCAAC

23: mac18red.fastq 0 0

22: mac-18_noad.fastq 0 0

GALAXY

Data upload

Data Libraries

Data Upload

- ★ sequencing reads
- ★ reference

The image displays three overlapping screenshots of the Galaxy web interface, illustrating the data upload process. The top screenshot shows the main Galaxy dashboard with a navigation bar (Analyze Data, Workflow, Shared Data, Visualization, Help) and a status bar (Using 8.3 MB). A green message box says "Hello world! It's running...". The "Data Libraries" menu item is highlighted. The middle screenshot shows the "Data Library 'NGS data'" page. The bottom screenshot shows the "Data Library 'NGS data'" page with a table of uploaded datasets and a list of tips for downloading.

Name	Message	Data type	Date uploaded	File size
<input type="checkbox"/> mac-18_noad.fastq	None	fastq	Tue Dec 2 09:35:38 2014 (UTC)	2.7 MB

For selected datasets:

TIP: You can download individual library datasets by selecting "Download this dataset" from the context menu (triangle) next to each dataset's name.

TIP: Several compression options are available for downloading multiple library datasets simultaneously:

- gzip: Recommended for fast network connections
- bzip2: Recommended for slower network connections (smaller size but takes longer to compress)
- zip: Not recommended but is provided as an option for those who cannot open the above formats

GALAXY

Data preparation

The screenshot shows the Galaxy web interface in a Firefox browser window. The main content area displays a command prompt with the command `>gi | 32139935 | PC-C40` and a block of DNA sequence data. The left sidebar contains a 'Tools' menu with categories like 'Operate on Genomic Intervals', 'Statistics', 'Graph/Display Data', 'Regional Variation', 'Multiple regression', 'Multivariate Analysis', 'Evolution', 'Motif Tools', 'Multiple Alignments', 'Metagenomic analyses', 'Human Genome Variation', and 'EMBOSS'. Under the 'NGS TOOLBOX BETA' section, the 'NGS: QC and manipulation' category is expanded, showing tools such as 'FASTQ Groomer', 'FASTQ splitter', 'FASTQ joiner', 'FASTQ Summary Statistics', 'ROCHE-454 DATA', 'Build base quality distribution', 'Select high quality segments', and 'Combine FASTA and QUAL into FASTQ'. A 'History' panel on the right shows a list of workflow steps: '25: PLMV_PC-C40.fasta', '24: peach_chloro.fa', '23: mac18red.fastq', and '22: mac-18_noad.fastq'. Two callout boxes are present: one pointing to the 'FASTQ Groomer' tool with the text 'open NGS tools', and another pointing to the 'FASTQ splitter' tool with the text 'Change score format from Illumina to Sanger'.

Data Upload
★ sequencing reads
★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

GALAXY

Data preparation

Choose data to work on

Data Upload

- ★ sequencing reads
- ★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

Reads are
Illumina data
Score version 1.3

The screenshot shows the Galaxy web interface with the FASTQ Groomer tool selected. The tool configuration is as follows:

- File to groom: 23: mac18red.fastq
- Input FASTQ quality scores type: Illumina 1.3+
- Advanced Options: Hide Advanced Options

The "What it does" section explains the tool's purpose and provides detailed instructions on handling quality scores and format conversions. A "Quality Score Comparison" section includes a visual representation of quality score scales for Sanger, Illumina 1.3, and Solexa formats.

GALAXY

Data preparation

The screenshot shows the Galaxy web interface in a Firefox browser window. The address bar shows the URL <http://main.g2.bx.psu.edu/>. The navigation menu includes 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. On the left, there is a 'Tools' sidebar with various categories like 'Operate on Genomic Intervals', 'Statistics', 'Graph/Display Data', etc. The main content area features a green notification box with a checkmark icon, stating: 'The following job has been successfully added to the queue: 27: FASTQ Groomer on data 22'. Below this, it provides instructions on how to check the status of jobs in the 'History' pane. The 'History' pane on the right lists several jobs, including '27: FASTQ Groomer on data 22', '26: FASTQ Groomer on data 23', '25: PLMV_PC-C40.fasta', '24: peach_chloro.fa', '23: mac18red.fastq', and '22: mac-18_noad.fastq'. A blue callout box with a pointer highlights the notification text.

Score conversions in queue

Data Upload

- ★ sequencing reads
- ★ reference

Data preparation

Data Quality control

Data Mapping

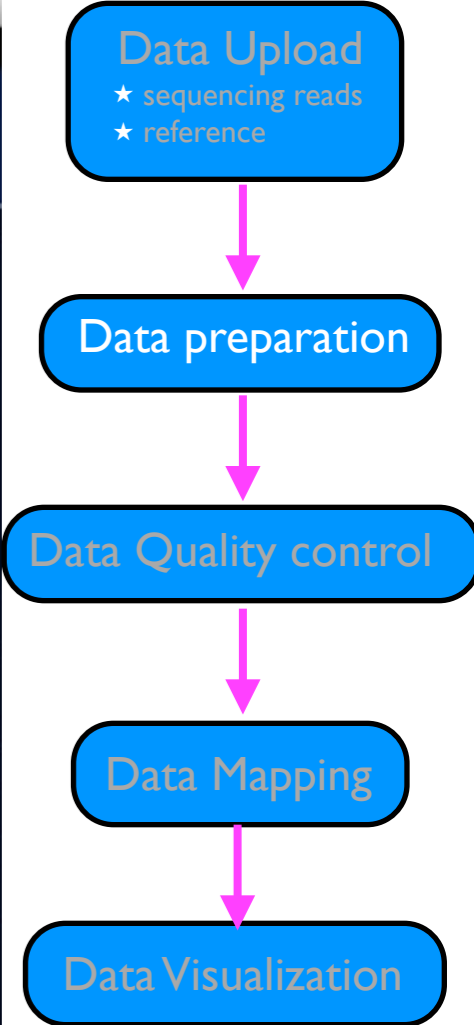
Data Visualization

GALAXY

Data preparation

The screenshot shows the Galaxy web interface in a Firefox browser window. The main content area displays a green notification box with a checkmark icon, stating: "The following job has been successfully added to the queue: 27: FASTQ Groomer on data 22". Below this, it provides instructions: "You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered." The History pane on the right shows a list of jobs, with job 27 selected and expanded to show details: "27: FASTQ Groomer on data 22", "2.7 Mb", "format: fastqsanger, database: ? Info: Groomed 20000 sanger reads into sanger reads. Based upon quality and sequence, the input data is valid for: solexa, sanger, illumina. Input ASCII range: 'B'(66) - 'Y'(102) Input decimal range: 33 - 69". Below the details, a scrollable list of FASTQ reads is visible, including headers like "@HWI-EAS2103_0008:6:107:16795:4771#NNGS" and sequence lines like "GAGAAAAGACGACCGGTAGAACGGA". The left sidebar contains a "Tools" menu with categories like "Operate on Genomic Intervals", "Statistics", "Regional Variation", "Multiple regression", "Multivariate Analysis", "Evolution", "Motif Tools", "Multiple Alignments", "Metagenomic analyses", "Human Genome Variation", "EMBOSS", "NGS TOOLBOX BETA", and "NGS: QC and manipulation". Under "NGS: QC and manipulation", several tools are listed, including "FASTQ Groomer", "FASTQ splitter", "FASTQ joiner", "FASTQ Summary Statistics", "ROCHE-454 DATA", "Build base quality distribution", "Select high quality segments", and "Combine FASTA and QUAL into FASTQ".

Scores converted
see all info



GALAXY

Data preparation

Edit attributes

Change file name

Data Upload

- ★ sequencing reads
- ★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

The screenshot shows the Galaxy web interface. The main content area is divided into two panels: 'Edit Attributes' and 'Change data type'. The 'Edit Attributes' panel has a 'Name' field with the value 'mac-18_noad_sa.fastq', an 'Info' field with 'Groomed 20000 sanger reads into sanger reads.Ba:', and a 'Database/Build' dropdown menu. The 'Change data type' panel has a 'New Type' dropdown menu set to 'fastqsanger'. The 'History' panel on the right shows a list of jobs, with the top one being '27: FASTQ Groomer on data 22'. The 'Tools' sidebar on the left lists various tools for genomic analysis.

GALAXY

Data preparation

The screenshot shows the Galaxy web interface in a Firefox browser window. The main content area displays a workflow step with a green checkmark and the message "Attributes updated". The history panel on the right shows a list of files, with the selected file "27: mac-18_noad_sa.fastq" (2.7 Mb) highlighted. A callout box points to this file name with the text "File name changed".

Tools available on the left sidebar include:

- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Regional Variation
- Multiple regression
- Multivariate Analysis
- Evolution
- Motif Tools
- Multiple Alignments
- Metagenomic analyses
- Human Genome Variation
- EMBOSS
- NGS TOOLBOX BETA
- NGS: QC and manipulation
 - ILLUMINA DATA
 - FASTQ Groomer convert between various FASTQ quality formats
 - FASTQ splitter on joined paired end reads
 - FASTQ joiner on paired end reads
 - FASTQ Summary Statistics by column
 - ROCHE-454 DATA
 - Build base quality distribution
 - Select high quality segments
 - Combine FASTA and QUAL into FASTQ

Data Upload

- ★ sequencing reads
- ★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

GALAXY

Data quality control

Choose data set

The screenshot shows the Galaxy web interface in a Firefox browser window. The main panel displays the 'Compute quality statistics' tool configuration. The 'Library to analyse' dropdown is set to '27: mac-18_noad_sa.fastq'. Below the configuration, there is a description of the tool's function and a list of output fields. A blue callout bubble points to the 'Compute quality statistics' tool in the left-hand 'Tools' sidebar.

Tools

- Filter FASTQ reads by quality score and length
- FASTQ Trimmer by column
- FASTQ Quality Trimmer by sliding window
- FASTQ Masker by quality score
- Manipulate FASTQ reads on various attributes
- FASTQ to FASTA converter
- FASTQ to Tabular converter
- Tabular to FASTQ converter
- FASTX-TOOLKIT FOR FASTQ DATA
- Quality format converter (ASCII-Numeric)
- Compute quality statistics**
- Draw quality score boxplot
- Draw nucleotides distribution chart
- FASTQ to FASTA converter
- Filter by quality
- Remove sequencing artifacts
- Barcode Splitter
- Clip adapter sequences
- Collapse sequences
- Rename sequences

Compute quality statistics

Library to analyse: 27: mac-18_noad_sa.fastq

Execute

What it does

Creates quality statistics report for the given Solexa/FASTQ library.

TIP: This statistics report can be used as input for Quality Score and Nucleotides Distribution tools.

The output file will contain the following fields:

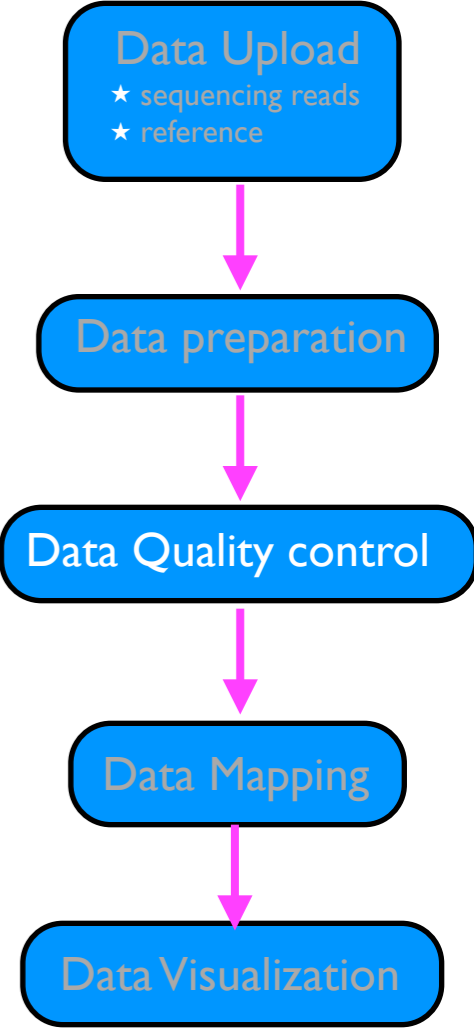
- column = column number (1 to 36 for a 36-cycles read Solexa file)
- count = number of bases found in this column
- min = Lowest quality
- max = Highest quality
- sum = Sum of quality
- mean = Mean quality
- Q1 = 1st quartile of quality
- IQR = Inter-Quartile Range
- IW = 'Left-Whisker' value (not boxplotting)
- rW = 'Right-Whisker' value (not boxplotting)
- A_Count = Count of 'A' nucleotides found in this column.
- C_Count = Count of 'C' nucleotides found in this column.
- G_Count = Count of 'G' nucleotides found in this column.
- T_Count = Count of 'T' nucleotides found in this column.
- N_Count = Count of 'N' nucleotides found in this column.

For example:

```
1 6362991 -4 40 250734117 39.41 40 40 40 0 40 40 1396976 1329101 678730 2958184 0
2 6362991 -5 40 250531036 39.37 40 40 40 0 40 40 1786786 1055766 1738025 1782414 0
3 6362991 -5 40 248722469 39.09 40 40 40 0 40 40 2296384 984875 1443989 1637743 0
4 6362991 -4 40 248214827 39.01 40 40 40 0 40 40 2536861 1167423 1248968 1409739 0
36 6362991 -5 40 117158566 18.41 7 15 30 23 -5 40 4074444 1402980 63287 822035 245
```

This tool is based on FASTX-toolkit by Assaf Gordon.

Compute quality statistics



GALAXY

Data quality control

nucleotide
position

Statistics

The screenshot shows the Galaxy web interface with a table of quality control statistics. The table has 14 columns: column, count, min, max, sum, mean, Q1, med, Q3, IQR, 1W, and zW. The rows represent different data points, with the last row (34) showing a mean of 24.02 and a Q1 of 22.

column	count	min	max	sum	mean	Q1	med	Q3	IQR	1W	zW
1	20000	2	38	717629	35.88	36	38	38	2	33	38
2	20000	2	38	720412	36.02	36	38	38	2	33	38
3	20000	2	38	718804	35.94	35	38	38	3	31	38
4	20000	2	38	718577	35.93	35	38	38	3	31	38
5	20000	2	38	717356	35.87	35	38	38	3	31	38
6	20000	2	38	718108	35.91	35	38	38	3	31	38
7	20000	2	38	717694	35.88	35	38	38	3	31	38
8	20000	2	38	715419	35.77	35	38	38	3	31	38
9	20000	2	38	715443	35.77	35	38	38	3	31	38
10	20000	2	38	715854	35.79	35	38	38	3	31	38
11	20000	2	38	713220	35.66	35	37	38	3	31	38
12	20000	2	38	713319	35.67	35	37	38	3	31	38
13	20000	2	38	709686	35.48	35	37	38	3	31	38
14	20000	2	38	709703	35.49	35	37	38	3	31	38
15	20000	2	38	707768	35.39	35	37	38	3	31	38
16	20000	2	38	705802	35.29	35	37	38	3	31	38
17	20000	2	38	702605	35.13	35	37	38	3	31	38
18	20000	2	38	699346	34.97	34	37	38	4	28	38
19	20000	2	38	696119	34.81	34	37	38	4	28	38
20	20000	2	38	694476	34.72	34	37	38	4	28	38
21	20000	2	38	690926	34.55	34	36	38	4	28	38
22	20000	2	38	686380	34.32	34	36	38	4	28	38
23	20000	2	38	685258	34.26	34	36	38	4	28	38
24	20000	2	38	680263	34.01	33	36	38	5	26	38
25	20000	2	38	674994	33.75	33	36	38	5	26	38
26	20000	2	38	680357	34.02	34	36	38	4	28	38
27	20000	2	38	676636	33.83	34	36	38	4	28	38
28	20000	2	38	673539	33.68	34	36	38	4	28	38
29	20000	2	38	668262	33.41	33	36	38	5	26	38
30	20000	2	38	662717	33.14	33	36	38	5	26	38
31	20000	2	38	651528	32.58	33	36	37	4	27	38
32	20000	2	34	461891	23.09	20	26	30	10	5	34
33	20000	2	35	457093	22.85	20	26	29	9	7	35
34	20000	2	34	480489	24.02	22	28	30	8	10	34

Data Upload

- ★ sequencing reads
- ★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

GALAXY

Data quality control

The screenshot shows the Galaxy web interface in a Firefox browser window. The main content area displays a table with the following columns: med, Q3, IQR, lW, rW, A_Count, C_Count, G_Count, T_Count, N_Count, and Max_count. The table contains 30 rows of data. On the left, there is a 'Tools' panel with a list of available tools, including 'Filter FASTQ reads by quality score and length', 'FASTQ Trimmer by column', 'FASTQ Quality Trimmer by sliding window', 'FASTQ Masker by quality score', 'Manipulate FASTQ reads on various attributes', 'FASTQ to FASTA converter', 'FASTQ to Tabular converter', 'Tabular to FASTQ converter', 'FASTX-TOOLKIT FOR FASTQ DATA', 'Quality format converter (ASCII-Numeric)', 'Compute quality statistics', 'Draw quality score boxplot', 'Draw nucleotides distribution chart', 'FASTQ to FASTA converter', 'Filter by quality', 'Remove sequencing artifacts', 'Barcode Splitter', 'Clip adapter sequences', 'Collapse sequences', and 'Rename sequences'. On the right, there is a 'History' panel showing a list of recent jobs, including '29: Compute quality statistics on data 26', '28: Compute quality statistics on data 27', '27: mac-18_noad_sa.fastq', '26: mac18red_sa.fastq', '25: PLMV_PC-C40.fasta', '24: peach_chloro.fa', '23: mac18red.fastq', and '22: mac-18_noad.fastq'.

med	Q3	IQR	lW	rW	A_Count	C_Count	G_Count	T_Count	N_Count	Max_count
38	38	2	33	38	8299	3607	2413	5681	0	20000
38	38	2	33	38	4714	3533	5219	6534	0	20000
38	38	3	31	38	5237	3348	5244	6171	0	20000
38	38	3	31	38	5358	3500	4905	6237	0	20000
38	38	3	31	38	6746	3634	4697	4923	0	20000
38	38	3	31	38	5982	4141	4856	5021	0	20000
38	38	3	31	38	5063	3986	4513	6438	0	20000
38	38	3	31	38	6451	3508	4165	5876	0	20000
38	38	3	31	38	6035	3745	4817	5403	0	20000
38	38	3	31	38	5120	3704	5776	5400	0	20000
37	38	3	31	38	5836	3858	4619	5687	0	20000
37	38	3	31	38	5337	4625	4961	5077	0	20000
37	38	3	31	38	6323	3226	5155	5296	0	20000
37	38	3	31	38	5766	3161	4836	6237	0	20000
37	38	3	31	38	5787	4380	5500	4333	0	20000
37	38	3	31	38	6309	3609	5203	4879	0	20000
37	38	3	31	38	6167	3318	4483	6032	0	20000
37	38	4	28	38	5193	5304	4556	4947	0	20000
37	38	4	28	38	4871	3868	5341	5920	0	20000
37	38	4	28	38	4934	3348	6631	5087	0	20000
36	38	4	28	38	4579	5208	6097	4104	12	20000
36	38	4	28	38	8100	3761	3736	4384	19	20000
36	38	4	28	38	5684	2837	3766	7671	42	20000
36	38	5	26	38	3687	7429	2908	5921	55	20000
36	38	5	26	38	7595	4241	1062	7039	63	20000
36	38	4	28	38	1441	6921	1207	10367	64	20000
36	38	4	28	38	1185	10052	5004	3684	75	20000
36	38	4	28	38	1465	2902	3525	12030	78	20000
36	38	5	26	38	5259	7578	2468	4608	87	20000
36	38	5	26	38	3668	1596	7628	7018	90	20000
36	37	4	27	38	2641	1735	5237	10300	87	20000
26	30	10	5	34	7447	5797	3363	3305	88	20000
26	29	9	7	35	1300	7995	2608	8013	84	20000
28	30	8	10	34	947	5291	11403	2277	82	20000

Data Upload

- ★ sequencing reads
- ★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

GALAXY

Data quality control

Statistics data for box plot

Data Upload

- ★ sequencing reads
- ★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

Tools

- Filter FASTQ reads by quality score and length
- FASTQ Trimmer by column
- FASTQ Quality Trimmer by sliding window
- FASTQ Masker by quality score
- Manipulate FASTQ reads on various attributes
- FASTQ to FASTA converter
- FASTQ to Tabular converter
- Tabular to FASTQ converter
- FASTX-TOOLKIT FOR FASTQ DATA
- Quality format converter (ASCII-Numeric)
- Compute quality statistics
- Draw quality score boxplot
- Draw nucleotides distribution chart
- FASTQ to FASTA converter
- Filter by quality
- Remove sequencing artifacts
- Barcode Splitter
- Clip adapter sequences
- Collapse sequences
- Rename sequences

Draw quality score boxplot

Statistics report file:
29: Compute quality s... on data 26
output of 'FASTQ Statistics' tool

Execute

What it does
Creates a boxplot graph for the quality scores in the I...

TIP: Use the FASTQ Statistics tool to generate the statistics needed for this tool.

Output Examples

- Black horizontal lines represent medians
- Rectangular boxes show the Inter-quartile Range (IQR) (top value is Q3, bottom value is Q1)
- Whiskers show outlier at max. 1.5*IQR

An example quality library (median quality is 40 for almost all 36 cycles):

Quality Score

Quality Score

Quality score box plot

GALAXY

Data quality control

The screenshot shows the Galaxy web interface in a Firefox browser window. The browser address bar shows the URL `http://main.g2.bx.psu.edu/`. The Galaxy navigation bar includes links for **Analyze Data**, **Workflow**, **Shared Data**, **Visualization**, **Help**, and **User**. On the left, a **Tools** panel lists various bioinformatics tools, including **Draw quality score boxplot**. The main workspace contains a green notification box with a checkmark icon, stating: "The following job has been successfully added to the queue: 31: Draw quality score boxplot on data 28. You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered." To the right, a **History** pane lists recent jobs, with job 31 at the top: "31: Draw quality score boxplot on data 28" (15.9 Kb, format: png, database: 2). A blue callout box with the text "Box plot for download" points to the download icon in the history entry for job 31.

Data Upload

- ★ sequencing reads
- ★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

GALAXY

Data quality control

The screenshot shows the Galaxy web interface. The main panel displays a boxplot titled "Quality Score for Compute quality statistics on data 26". The x-axis represents quality scores from 1 to 35, and the y-axis represents quality scores from 0 to 35. The boxplot shows a distribution of quality scores across 35 positions. A callout bubble points to the plot with the text "data set with different # nt".

The History panel on the right shows a list of tools and their outputs:

- 31: Draw quality score boxplot on data 28
- 30: Draw quality score boxplot on data 29
- 29: Compute quality statistics on data 26
35 lines
format: txt, database: ?

column	count	min	max	sum
1	20000	2	38	717
2	20000	2	38	720
3	20000	2	38	718
4	20000	2	38	718
5	20000	2	38	717
- 28: Compute quality statistics on data 27
25 lines
format: txt, database: ?

column	count	min	max	sum
1	20000	33	69	132
2	20000	33	69	133
3	20000	33	69	133
4	20000	33	69	133
5	20000	33	69	132

Data Upload

- ★ sequencing reads
- ★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

GALAXY

Data mapping

NGS Mapping

Use our own reference

Choose a fasta file (reference)

keep default

choose size reduced reads

Bowtie for Illumina

keep default

Data Upload

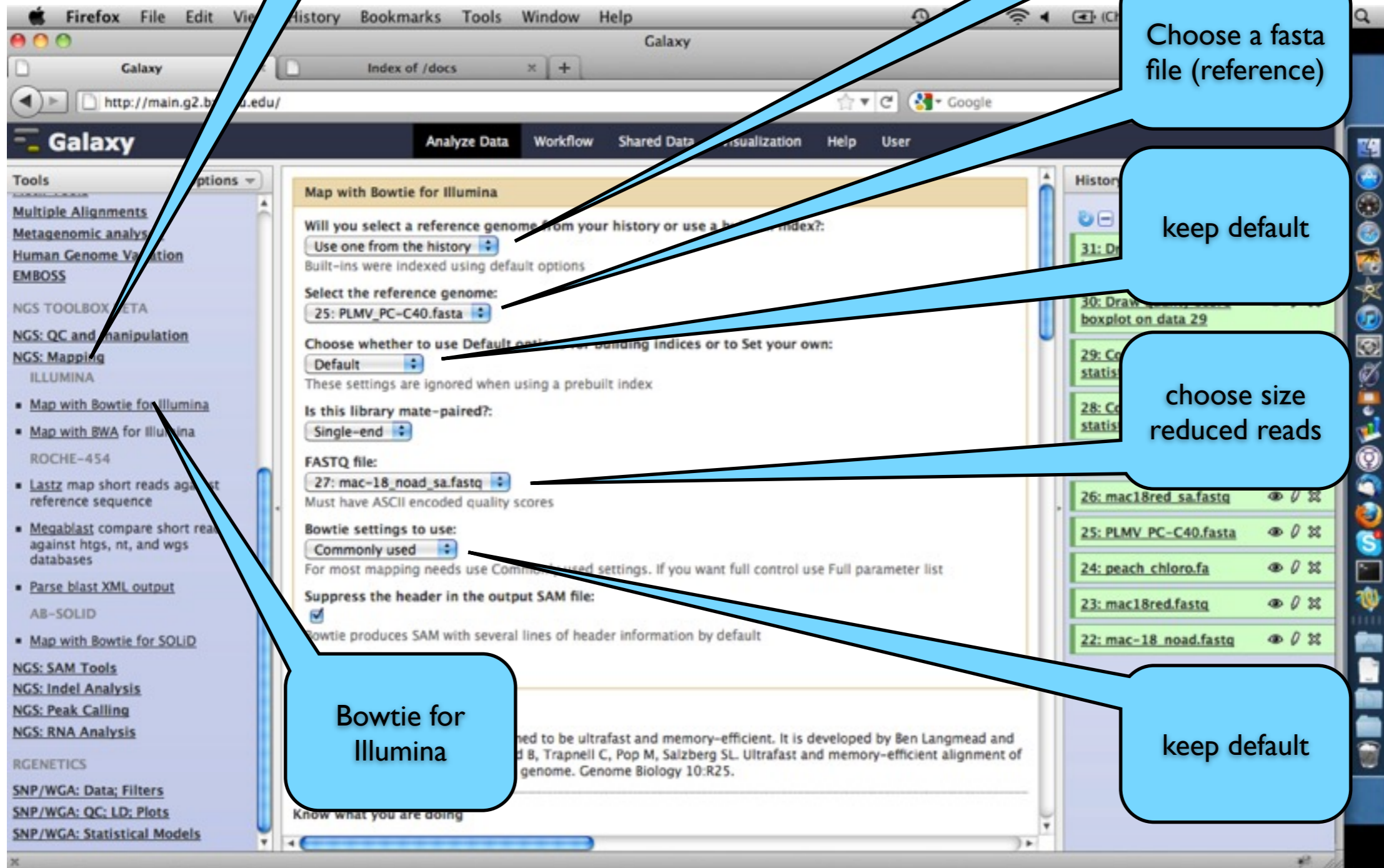
★ sequencing reads
★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization



GALAXY

Data mapping

Mapping file in SAM format

NGS Mapping

Data Upload
★ sequencing reads
★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

Firefox File Edit View History Bookmarks Window Help Galaxy

http://main.g2.bx.psu.edu/

Galaxy Analyze Data Workflow Shared Data Visualization Help User

Tools Options

- Multiple Alignments
- Metagenomic analyses
- Human Genome Variation
- EMBOSS
- NGS TOOLBOX BETA
- NGS: QC and manipulation
- NGS: Mapping
 - ILLUMINA
 - Map with Bowtie for Illumina
 - Map with BWA for Illumina
 - ROCHE-454
 - Lastz map short reads against reference sequence
 - Megablast compare short reads against htgs, nt, and wgs databases
 - Parse blast XML output
 - AB-SOLID
 - Map with Bowtie for SOLID
- NGS: SAM Tools
- NGS: Indel Analysis
- NGS: Peak Calling
- NGS: RNA Analysis
- RGENTICS
- SNP/WGA: Data; Filters
- SNP/WGA: QC; LD; Plots
- Statistical Models

```
HWIEAS210R_0008:6:107:16809:1172#NNGNNN/1_24 4 * 0 0 * * * 0 0 TTOTGC
HWIEAS210R_0008:6:107:16808:1178#NNGNNN/1_24 4 * 0 0 * * * 0 0 ATCTTC
HWIEAS210R_0008:6:107:16811:516#NNGNNN/1_24 4 * 0 0 * * * 0 0 ACGATT
HWIEAS210R_0008:6:107:16810:516#NNGNNN/1_24 4 * 0 0 * * * 0 0 GTCAGG
HWIEAS210R_0008:6:107:16812:117#NNGNNN/1_24 4 * 0 0 * * * 0 0 AATCTT
HWIEAS210R_0008:6:107:16812:111#NNGNNN/1_24 4 * 0 0 * * * 0 0 ATGAGC
HWIEAS210R_0008:6:107:16813:213#NNGNNN/1_24 4 * 0 0 * * * 0 0 CCTTAC
HWIEAS210R_0008:6:107:16814:1318#NNGNNN/1_24 4 * 0 0 * * * 0 0 CGGGAG
HWIEAS210R_0008:6:107:16817:1107#NNGNNN/1_24 4 * 0 0 * * * 0 0 ATTCTT
HWIEAS210R_0008:6:107:16817:1172#NNGNNN/1_24 4 * 0 0 * * * 0 0 ATCAGC
HWIEAS210R_0008:6:107:16816:503#NNGNNN/1_24 4 * 0 0 * * * 0 0 ACAAGA
HWIEAS210R_0008:6:107:16816:492#NNGNNN/1_24 4 * 0 0 * * * 0 0 AGAGAT
HWIEAS210R_0008:6:107:16820:1486#NNGNNN/1_24 4 * 0 0 * * * 0 0 ATACAA
HWIEAS210R_0008:6:107:16820:14146#NNGNNN/1_24 4 * 0 0 * * * 0 0 AGGGTT
HWIEAS210R_0008:6:107:16819:16807#NNGNNN/1_24 4 * 0 0 * * * 0 0 AGAGAA
HWIEAS210R_0008:6:107:16819:13954#NNGNNN/1_24 4 * 0 0 * * * 0 0 ATAGAA
HWIEAS210R_0008:6:107:16820:9052#NNGNNN/1_24 4 * 0 0 * * * 0 0 GATTTT
HWIEAS210R_0008:6:107:16820:5386#NNGNNN/1_24 4 * 0 0 * * * 0 0 TCTATG
HWIEAS210R_0008:6:107:16822:7760#NNGNNN/1_24 4 * 0 0 * * * 0 0 ATCAAA
HWIEAS210R_0008:6:107:16820:13229#NNGNNN/1_24 4 * 0 0 * * * 0 0 TCTTTA
HWIEAS210R_0008:6:107:16823:2007#NNGNNN/1_24 4 * 0 0 * * * 0 0 ATATAA
HWIEAS210R_0008:6:107:16824:7599#NNGNNN/1_24 4 * 0 0 * * * 0 0 TCGGAC
HWIEAS210R_0008:6:107:16826:19079#NNGNNN/1_24 4 * 0 0 * * * 0 0 ACACTG
HWIEAS210R_0008:6:107:16825:8418#NNGNNN/1_24 4 * 0 0 * * * 0 0 ATTAAG
HWIEAS210R_0008:6:107:16824:13596#NNGNNN/1_24 4 * 0 0 * * * 0 0 AGATCA
HWIEAS210R_0008:6:107:16826:4226#NNGNNN/1_24 4 * 0 0 * * * 0 0 TAACGT
HWIEAS210R_0008:6:107:16826:19389#NNGNNN/1_24 4 * 0 0 * * * 0 0 ATGGGT
HWIEAS210R_0008:6:107:16827:14916#NNGNNN/1_24 4 * 0 0 * * * 0 0 ATATCG
HWIEAS210R_0008:6:107:16830:13358#NNGNNN/1_24 4 * 0 0 * * * 0 0 ATGCGG
HWIEAS210R_0008:6:107:16830:11968#NNGNNN/1_24 4 * 0 0 * * * 0 0 ATTTTC
HWIEAS210R_0008:6:107:16831:2431#NNGNNN/1_24 4 * 0 0 * * * 0 0 AGTTGG
HWIEAS210R_0008:6:107:16833:9093#NNGNNN/1_24 4 * 0 0 * * * 0 0 ATGGGG
HWIEAS210R_0008:6:107:16834:3612#NNGNNN/1_24 4 * 0 0 * * * 0 0 TAAAGG
HWIEAS210R_0008:6:107:16836:4041#NNGNNN/1_24 4 * 0 0 * * * 0 0 TGGTCC
HWIEAS210R_0008:6:107:16841:14875#NNGNNN/1_24 4 * 0 0 * * * 0 0 ATGGAG
HWIEAS210R_0008:6:107:16843:7350#NNGNNN/1_24 4 * 0 0 * * * 0 0 ACTTAA
HWIEAS210R_0008:6:107:16837:3173#NNGNNN/1_24 4 * 0 0 * * * 0 0 ATTTCA
HWIEAS210R_0008:6:107:16843:16008#NNGNNN/1_24 4 * 0 0 * * * 0 0 GAGTTA
HWIEAS210R_0008:6:107:16843:6852#NNGNNN/1_24 4 * 0 0 * * * 0 0 AGAGAT
HWIEAS210R_0008:6:107:16838:4931#NNGNNN/1_24 4 * 0 0 * * * 0 0 GAATAG
HWIEAS210R_0008:6:107:16844:8805#NNGNNN/1_24 4 * 0 0 * * * 0 0 GCTAAG
HWIEAS210R_0008:6:107:16845:13101#NNGNNN/1_24 4 * 0 0 * * * 0 0 CGATGT
HWIEAS210R_0008:6:107:16844:6242#NNGNNN/1_24 4 * 0 0 * * * 0 0 TTCAAT
HWIEAS210R_0008:6:107:16845:2048#NNGNNN/1_24 4 * 0 0 * * * 0 0 TGCATT
HWIEAS210R_0008:6:107:16847:1394#NNGNNN/1_24 4 * 0 0 * * * 0 0 ACGGCA
HWIEAS210R_0008:6:107:16846:10697#NNGNNN/1_24 4 * 0 0 * * * 0 0 ACGGCC
HWIEAS210R_0008:6:107:16846:16819#NNGNNN/1_24 4 * 0 0 * * * 0 0 ATGGAC
HWIEAS210R_0008:6:107:16849:13025#NNGNNN/1_24 4 * 0 0 * * * 0 0 CCGTGG
HWIEAS210R_0008:6:107:16851:4128#NNGNNN/1_24 4 * 0 0 * * * 0 0 ATGGTA
HWIEAS210R_0008:6:107:16849:4960#NNGNNN/1_24 0 gi|309321413|gb|HQ336405.1| 87969 255 24M
HWIEAS210R_0008:6:107:16852:16291#NNGNNN/1_24 4 * 0 0 * * * 0 0 ATACGA
HWIEAS210R_0008:6:107:16853:3441#NNGNNN/1_24 4 * 0 0 * * * 0 0 CCTTTG
HWIEAS210R_0008:6:107:16851:16384#NNGNNN/1_24 4 * 0 0 * * * 0 0 CCAATA
HWIEAS210R_0008:6:107:16853:4264#NNGNNN/1_24 4 * 0 0 * * * 0 0 AGTATT
HWIEAS210R_0008:6:107:16853:18475#NNGNNN/1_24 4 * 0 0 * * * 0 0 AATGGC
HWIEAS210R_0008:6:107:16853:9868#NNGNNN/1_24 4 * 0 0 * * * 0 0 TCCAGT
```

HWIEAS210R_0008:6:107:16849:4960#NNGNNN/1_24 0 gi|309321413|gb|HQ336405.1| 87969 255 24M * 0 0 CAATAAGAATGCTAGTTCTTACTG ``Y``\L\]\Z__ _b^]]bb`` XA:i:0 MD:Z:24 NM:i:0

History Options

37: Map with Bowtie for Illumina on data 27 and data 24: mapped reads ~19,000 lines format: sam, database Info: Settings: Output files: "/tmp/1542209.cyber...psu.edu /tmp/K0ll/tmp_wpEw...ebwt" Line rate: 6 (line is 64 bytes) Lines per side: 1 (side is 64 bytes) Offset rate: 5 (one in 3) FTable chars: 10 Strings: unpacked Max bucket size: default

36: Map with Bowtie for Illumina on data 27 and data 25: mapped reads

GALAXY

Data mapping

Firefox File Edit View History Bookmarks Tools Window Help

Galaxy

Index of /docs

http://main.g2.bx.psu.edu/

Galaxy

Analyze Data Workflow Shared Data Visualization Help User

Tools Options

- Multiple Alignments
- Metagenomic analyses
- Human Genome Variation
- EMBOSS
- NGS TOOLBOX BETA
- NGS: QC and manipulation
- NGS: Mapping
 - ILLUMINA
 - Map with Bowtie for Illumina
 - Map with BWA for Illumina
 - ROCHE-454
 - Lastz map short reads against reference sequence
 - Megablast compare short reads against htgs, nt, and wgs databases
 - Parse blast XML output
 - AB-SOLID
 - Map with Bowtie for SOLID
- NGS: SAM Tools
- NGS: Indel Analysis
- NGS: Peak Calling
- NGS: RNA Analysis
- RGENETICS
- SNP/WGA: Data; Filters
- SNP/WGA: QC; LD; Plots
- Statistical Models

```
HWI-EAS210X_0008:6:107:16809:12072#SNHG2SN/1_24 4 * 0 0 * * * * * TTOTGC
HWI-EAS210X_0008:6:107:16808:14978#SNHG2SN/1_24 4 * 0 0 * * * * * ATCTTC
HWI-EAS210X_0008:6:107:16811:5776#SNHG2SN/1_24 4 * 0 0 * * * * * ACGATT
HWI-EAS210X_0008:6:107:16810:9594#SNHG2SN/1_24 4 * 0 0 * * * * * GTCAGG
HWI-EAS210X_0008:6:107:16812:7917#SNHG2SN/1_24 4 * 0 0 * * * * * AATCTT
HWI-EAS210X_0008:6:107:16812:17411#SNHG2SN/1_24 4 * 0 0 * * * * * ATGAGC
HWI-EAS210X_0008:6:107:16813:2983#SNHG2SN/1_24 4 * 0 0 * * * * * CCTTAC
HWI-EAS210X_0008:6:107:16814:3998#SNHG2SN/1_24 4 * 0 0 * * * * * CGGGAG
HWI-EAS210X_0008:6:107:16817:19807#SNHG2SN/1_24 4 * 0 0 * * * * * ATTCTC
HWI-EAS210X_0008:6:107:16817:17172#SNHG2SN/1_24 4 * 0 0 * * * * * ATCAAG
HWI-EAS210X_0008:6:107:16816:5693#SNHG2SN/1_24 4 * 0 0 * * * * * ACAAGA
HWI-EAS210X_0008:6:107:16816:4492#SNHG2SN/1_24 4 * 0 0 * * * * * AGAGAT
HWI-EAS210X_0008:6:107:16820:1486#SNHG2SN/1_24 4 * 0 0 * * * * * ATACAA
HWI-EAS210X_0008:6:107:16820:14146#SNHG2SN/1_24 4 * 0 0 * * * * * AGGGTT
HWI-EAS210X_0008:6:107:16819:16807#SNHG2SN/1_24 4 * 0 0 * * * * * AGAGAA
HWI-EAS210X_0008:6:107:16819:13954#SNHG2SN/1_24 4 * 0 0 * * * * * ATAGAA
HWI-EAS210X_0008:6:107:16820:9052#SNHG2SN/1_24 4 * 0 0 * * * * * GATTTT
HWI-EAS210X_0008:6:107:16820:5386#SNHG2SN/1_24 4 * 0 0 * * * * * TCTATG
HWI-EAS210X_0008:6:107:16822:7760#SNHG2SN/1_24 4 * 0 0 * * * * * ATCAAA
HWI-EAS210X_0008:6:107:16820:13229#SNHG2SN/1_24 4 * 0 0 * * * * * TCTTTA
HWI-EAS210X_0008:6:107:16823:2007#SNHG2SN/1_24 4 * 0 0 * * * * * ATATAA
HWI-EAS210X_0008:6:107:16824:7599#SNHG2SN/1_24 4 * 0 0 * * * * * TCGGAC
HWI-EAS210X_0008:6:107:16826:19079#SNHG2SN/1_24 4 * 0 0 * * * * * ACACTG
HWI-EAS210X_0008:6:107:16825:8418#SNHG2SN/1_24 4 * 0 0 * * * * * ATTAAG
HWI-EAS210X_0008:6:107:16824:13596#SNHG2SN/1_24 4 * 0 0 * * * * * AGATCA
HWI-EAS210X_0008:6:107:16826:4226#SNHG2SN/1_24 4 * 0 0 * * * * * TAACGT
HWI-EAS210X_0008:6:107:16826:19389#SNHG2SN/1_24 4 * 0 0 * * * * * ATGGGT
HWI-EAS210X_0008:6:107:16827:14916#SNHG2SN/1_24 4 * 0 0 * * * * * ATATCG
HWI-EAS210X_0008:6:107:16830:13558#SNHG2SN/1_24 4 * 0 0 * * * * * ATGCGG
HWI-EAS210X_0008:6:107:16830:11968#SNHG2SN/1_24 4 * 0 0 * * * * * ATTTTC
HWI-EAS210X_0008:6:107:16831:2431#SNHG2SN/1_24 4 * 0 0 * * * * * AGTTGG
HWI-EAS210X_0008:6:107:16833:9093#SNHG2SN/1_24 4 * 0 0 * * * * * ATGGGG
HWI-EAS210X_0008:6:107:16834:3612#SNHG2SN/1_24 4 * 0 0 * * * * * TAAAAA
HWI-EAS210X_0008:6:107:16836:4041#SNHG2SN/1_24 4 * 0 0 * * * * * ATGTCG
HWI-EAS210X_0008:6:107:16841:14875#SNHG2SN/1_24 4 * 0 0 * * * * * ATGGAG
HWI-EAS210X_0008:6:107:16843:7350#SNHG2SN/1_24 4 * 0 0 * * * * * ACTTAA
HWI-EAS210X_0008:6:107:16837:3173#SNHG2SN/1_24 4 * 0 0 * * * * * ATTTCA
HWI-EAS210X_0008:6:107:16843:16008#SNHG2SN/1_24 4 * 0 0 * * * * * GAGTTA
HWI-EAS210X_0008:6:107:16843:6852#SNHG2SN/1_24 4 * 0 0 * * * * * AGAGAT
HWI-EAS210X_0008:6:107:16838:4931#SNHG2SN/1_24 4 * 0 0 * * * * * GAATAG
HWI-EAS210X_0008:6:107:16844:8805#SNHG2SN/1_24 4 * 0 0 * * * * * GCTAAA
HWI-EAS210X_0008:6:107:16845:13101#SNHG2SN/1_24 4 * 0 0 * * * * * CGATGT
HWI-EAS210X_0008:6:107:16844:6242#SNHG2SN/1_24 4 * 0 0 * * * * * TTCAGT
HWI-EAS210X_0008:6:107:16845:2048#SNHG2SN/1_24 4 * 0 0 * * * * * TGCAAT
HWI-EAS210X_0008:6:107:16847:1394#SNHG2SN/1_24 4 * 0 0 * * * * * ACGGCA
HWI-EAS210X_0008:6:107:16846:10697#SNHG2SN/1_24 4 * 0 0 * * * * * ACGGCC
HWI-EAS210X_0008:6:107:16846:16819#SNHG2SN/1_24 4 * 0 0 * * * * * ATGGAC
HWI-EAS210X_0008:6:107:16849:13025#SNHG2SN/1_24 4 * 0 0 * * * * * CCGTCG
HWI-EAS210X_0008:6:107:16851:4128#SNHG2SN/1_24 4 * 0 0 * * * * * ATGGTA
HWI-EAS210X_0008:6:107:16849:4960#SNHG2SN/1_24 4 * 0 0 * * * * * ATACGA
HWI-EAS210X_0008:6:107:16852:16291#SNHG2SN/1_24 4 * 0 0 * * * * * ATGTTT
HWI-EAS210X_0008:6:107:16853:3441#SNHG2SN/1_24 4 * 0 0 * * * * * CCAATT
HWI-EAS210X_0008:6:107:16851:16584#SNHG2SN/1_24 4 * 0 0 * * * * * AGTATT
HWI-EAS210X_0008:6:107:16853:4264#SNHG2SN/1_24 4 * 0 0 * * * * * AATGGG
HWI-EAS210X_0008:6:107:16853:18475#SNHG2SN/1_24 4 * 0 0 * * * * * TCCAGT
HWI-EAS210X_0008:6:107:16853:9888#SNHG2SN/1_24 4 * 0 0 * * * * * ACGTTG
HWI-EAS210X_0008:6:107:16854:2336#SNHG2SN/1_24 4 * 0 0 * * * * * ATTGAC
HWI-EAS210X_0008:6:107:16855:19893#SNHG2SN/1_24 4 * 0 0 * * * * * ATCAAA
HWI-EAS210X_0008:6:107:16856:12384#SNHG2SN/1_24 4 * 0 0 * * * * * CGAGAA
HWI-EAS210X_0008:6:107:16854:11696#SNHG2SN/1_24 4 * 0 0 * * * * * GACACT
HWI-EAS210X_0008:6:107:16859:19375#SNHG2SN/1_24 4 * 0 0 * * * * * ACGAAA
HWI-EAS210X_0008:6:107:16858:2186#SNHG2SN/1_24 4 * 0 0 * * * * * TTCTGG
HWI-EAS210X_0008:6:107:16840:20959#SNHG2SN/1_24 4 * 0 0 * * * * * TTTTCC
```

History Options

37: Map with Bowtie for Illumina on data 27 and data 24: mapped reads

~19,000 lines

format: sam, database: ?

Info: Settings:

Output files:

/tmp/1542209.cyberstar.psu.edu/tmpvK0ll/tmp_wpEwm..ebwt

Line rate: 6 (line is 64 bytes)

Lines per side: 1 (side is 64 bytes)

Offset rate: 5 (one in 32)

FTable chars: 10

Strings: unpacked

Max bucket size: default

1 - QNAME

HWI-EAS210X_0008:6:107:16803:4320#SNHG2SN/1_24

HWI-EAS210X_0008:6:107:16800:1162#SNHG2SN/1_24

HWI-EAS210X_0008:6:107:16801:6560#SNHG2SN/1_24

HWI-EAS210X_0008:6:107:16797:18323#SNHG2SN/1_24

HWI-EAS210X_0008:6:107:16799:5109#SNHG2SN/1_24

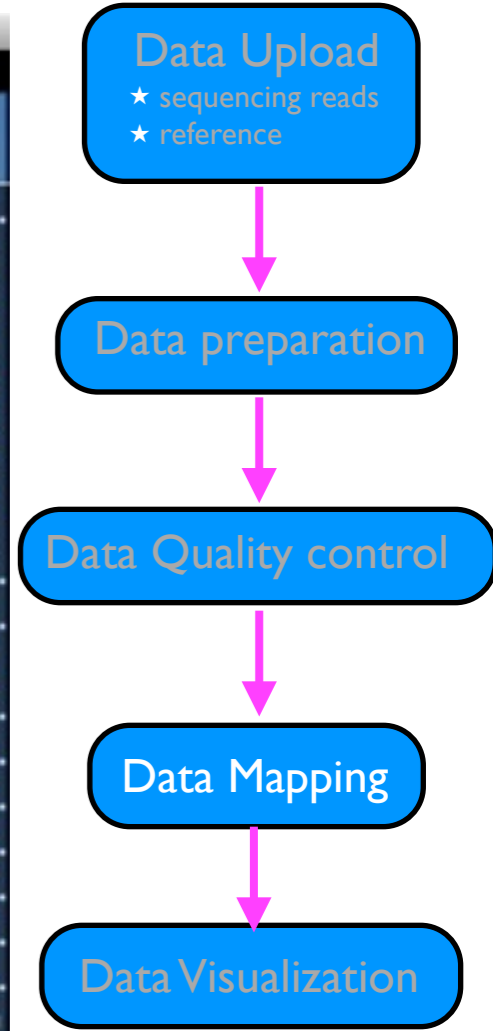
HWI-EAS210X_0008:6:107:16795:4771#SNHG2SN/1_24

36: Map with Bowtie for Illumina on data 27 and data 25: mapped reads

31: Draw quality score boxplot on data 28

30: Draw quality score boxplot on data 29

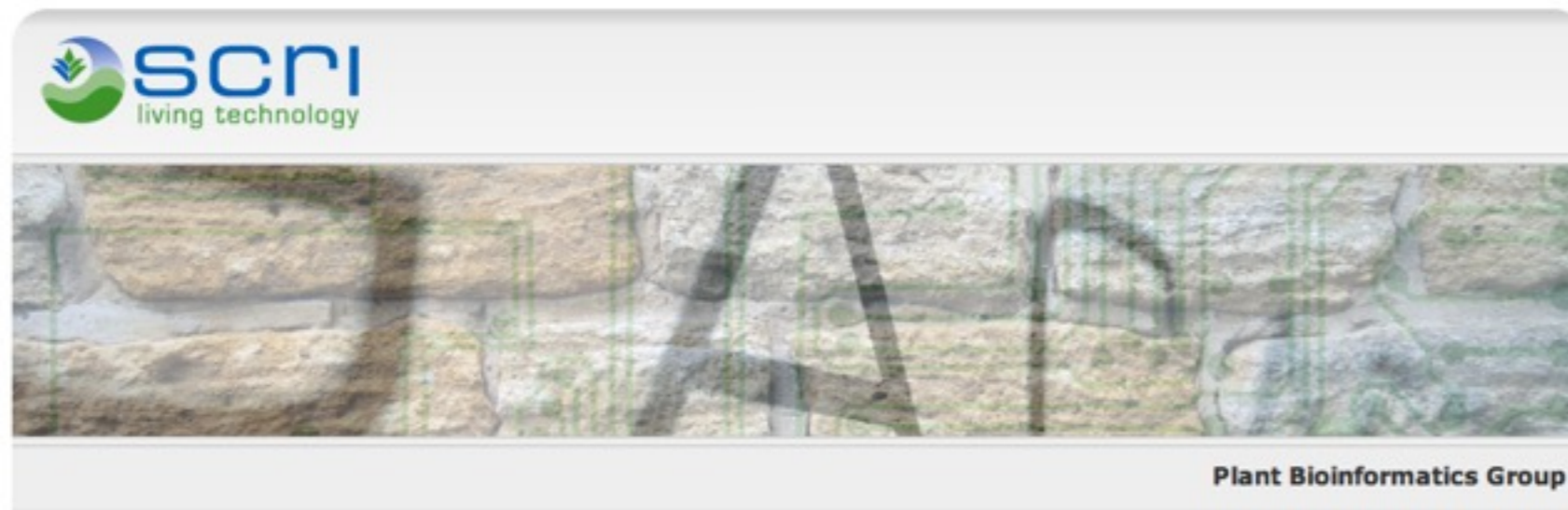
Export SAM files for external visualization



NOT GALAXY

Data visualization

<http://bioinf.scri.ac.uk/tablet/>



Tablet

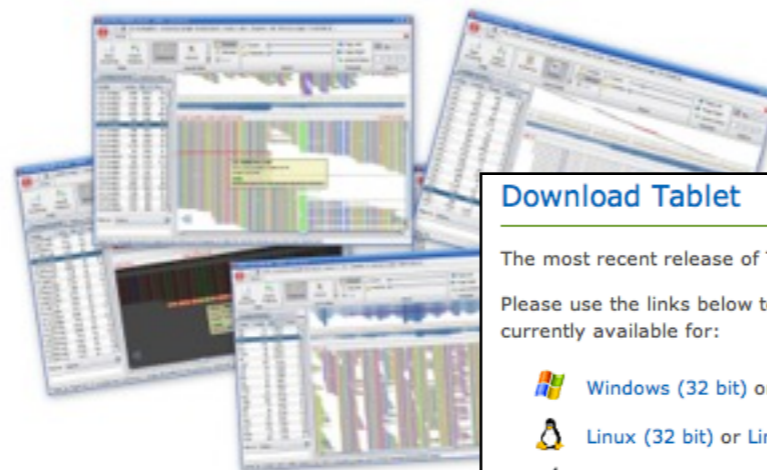
- Tablet Homepage
- Download Tablet
- Screenshots
- Tablet FAQ
- Sample Data
- Assembly Conversion
- Papers and Presentations
- Privacy Policy
- Tablet World Map
- Online Help

Our Software

- CurlyWhirly
- Flapjack
- OPTIRas
- Strudel
- Tablet (**new**)
- TetraploidMap
- TOPALI

Tablet - Next Generation Sequence Assembly Visualization

Tablet is a lightweight, high-performance graphical viewer for next generation sequence assemblies and alignments.



Download Tablet

The most recent release of Tablet is **1.11.02.18** (18th February 2011).

Please use the links below to download the Tablet installer most suitable for your operating system. Tablet is currently available for:

-  Windows (32 bit) or Windows (64 bit)
-  Linux (32 bit) or Linux (64 bit)
-  Apple Mac OS X (Java 6 required)
-  Solaris (Sparc)

Data Upload

- ★ sequencing reads
- ★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

Tablet

Data visualization

The screenshot shows the Tablet web application interface. At the top, the title bar reads "Tablet - 1.11.03.14" and the memory usage is "46.86 MB (3)". The interface includes a navigation menu with "Home" and "Advanced" tabs. Below this is a toolbar with sections for "Data" (Open Assembly, Import Features), "Layout Style" (Nucleotide, Direction, Read Type, Classic), "Adjust" (Zoom, Variants), "Navigate", and "Overlays". The main content area displays a "Welcome to Tablet" message and a "Getting started" section. A callout box labeled "Start Visualization" points to the "Import an assembly into Tablet" link. The "Getting started" section lists several quick-start options and a "Learning more" section with a list of topics. At the bottom, there is a "Citing Tablet" section and a "Tablet Tip" at the very bottom.

Memory usage: 46.86 MB (3)

Home Advanced

Open Assembly Import Features

Data

Nucleotide Direction Read Type Classic

Layout Style

Zoom: Variants:

Adjust

Navigate Overlays

Welcome to Tablet

Tablet 1.11.03.14 - © 2009-2011, Plant Bioinformatics Group, SCRI. [Send feedback](#) [Follow us](#)

Getting started [Learning more](#)

Import an assembly into Tablet

Quickly open a previously accessed assembly:

- Galaxy13-[Map_with_Bowtie_for_Illumina_on_data_12_and_data_11_mapped_rea
- Galaxy13-[Map_with_Bowtie_for_Illumina_on_data_12_and_data_11_mapped_rea
- mac-17_chloro_clean.sam ~ peach_chloro.fa
- mac-17_chloro.sam ~ peach_chloro.fa

Learning more

- Tablet quick start
- View the Tablet interface
- Opening assembly files
- Visualizing contigs and their reads
- Working with the overview panels
- Changing Tablet settings
- What's new in this version of Tablet
- Tips and shortcuts

Click to rate Tablet: ★★★★★

Citing Tablet

Please click here for information on how to cite Tablet if you use it in your work

Tablet Tip: Position data is often supplemented with U (unpadded position) and CV (read coverage at that position) values

Data Upload

- ★ sequencing reads
- ★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

Tablet

Data visualization

Tablet - 1.11.03.14

Memory usage: 59.35 MB (3)

Home Advanced

Open Assembly Import Features

Nucleotide Direction Read Type Classic

Zoom: Variants:

Adjust Navigate Overlays

Data Open Assembly

Select assembly files:

Primary assembly file or URL:
/Users/andreas/Downloads/Galaxy36-[mac-18_plmvd.sam].sam

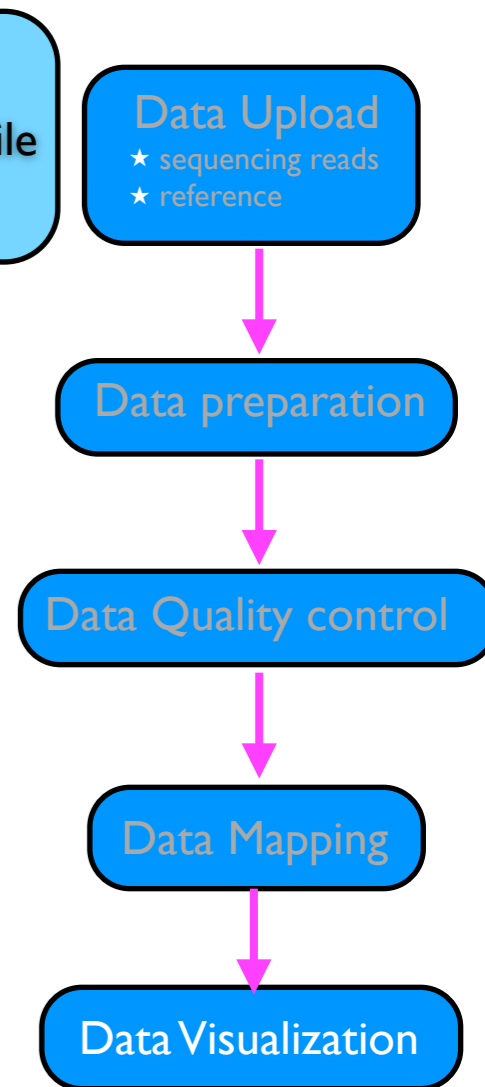
Reference/consensus file or URL:
/Users/andreas/Downloads/uppsala_class/PLMV_PC-C40.fasta

Current status: Assembly - SAM | Reference - FASTA

Notes:

- Tablet currently supports ACE, AFG, MAQ (text), SOAP, SAM, and (indexed) BAM assemblies.
- Reference files (if needed for MAQ, SOAP, SAM and BAM) can be in FASTA or FASTQ format.
- Unsure how to get started? [Click here to open an example assembly.](#)

Open Cancel Help



Tablet

Data visualization Viroid

Layout

zoom

Galaxy36-[mac-18_plmvd.sam].sam - Tablet - 1.11.03.14

gi|32139935|PC-C40 | consensus length: 419 | reads: 295 | features: 0 | Memory usage: 58.86 MB (8)

Home Advanced

Open Assembly Import Features

Data

Nucleotide Direction Read Type Classic

Layout Style

Zoom: Variants:

Adjust

Navigate Overlays

Contigs: 1 295 total reads

Contig	Length	Reads	Feat...	Mis...
gi 3...	19	295	0	3.2

Choose contig

1 to 419 (419 bp) 1 to 38 (38 bp)

1 38

Tablet Tip: Navigate around an alignment by clicking and dragging on either the overview display area or the main display area

Data Upload

- ★ sequencing reads
- ★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

Tablet

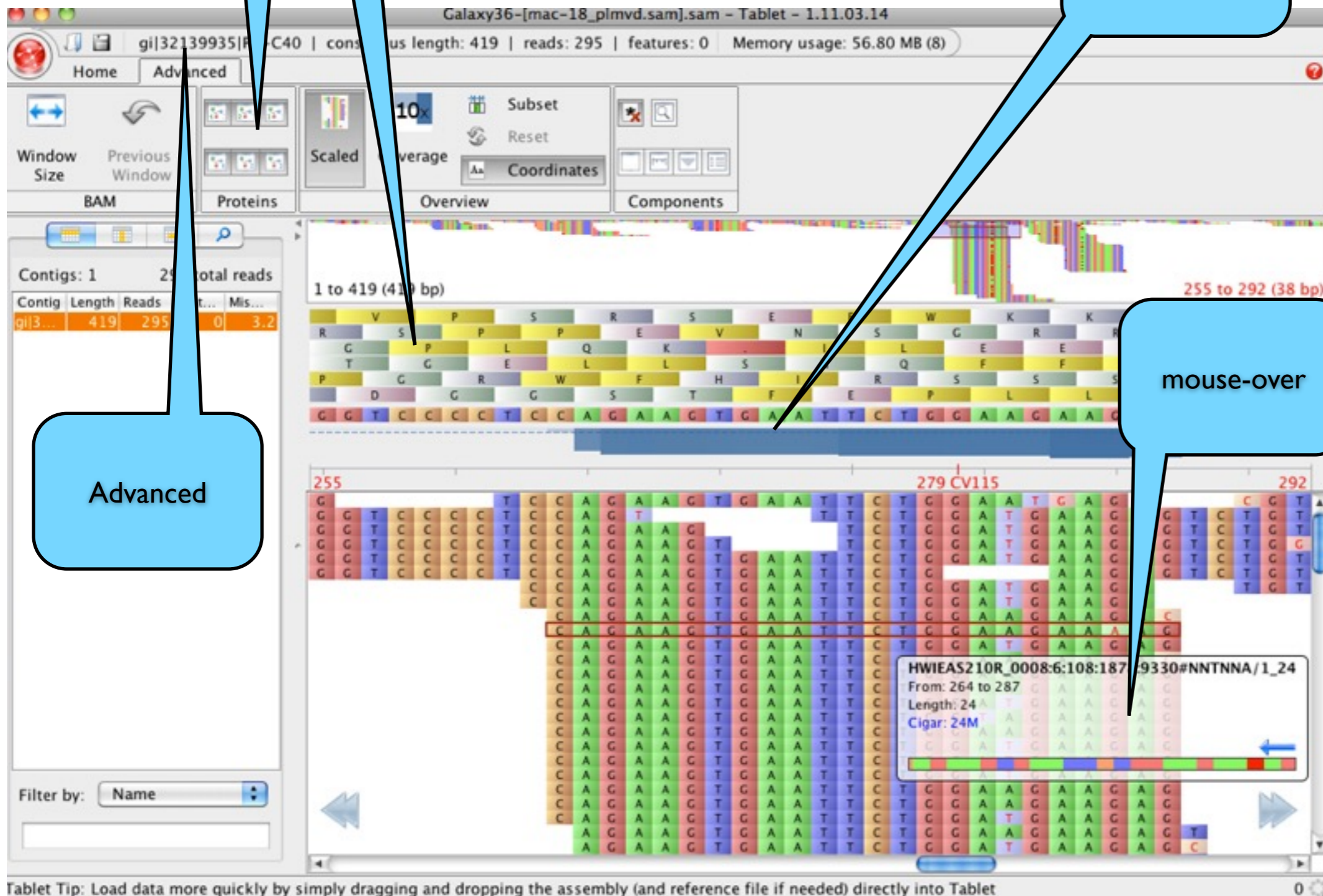
Data visualization
Viroid

Protein reading
frame

coverage

mouse-over

Advanced



Data Upload

- ★ sequencing reads
- ★ reference

Data preparation

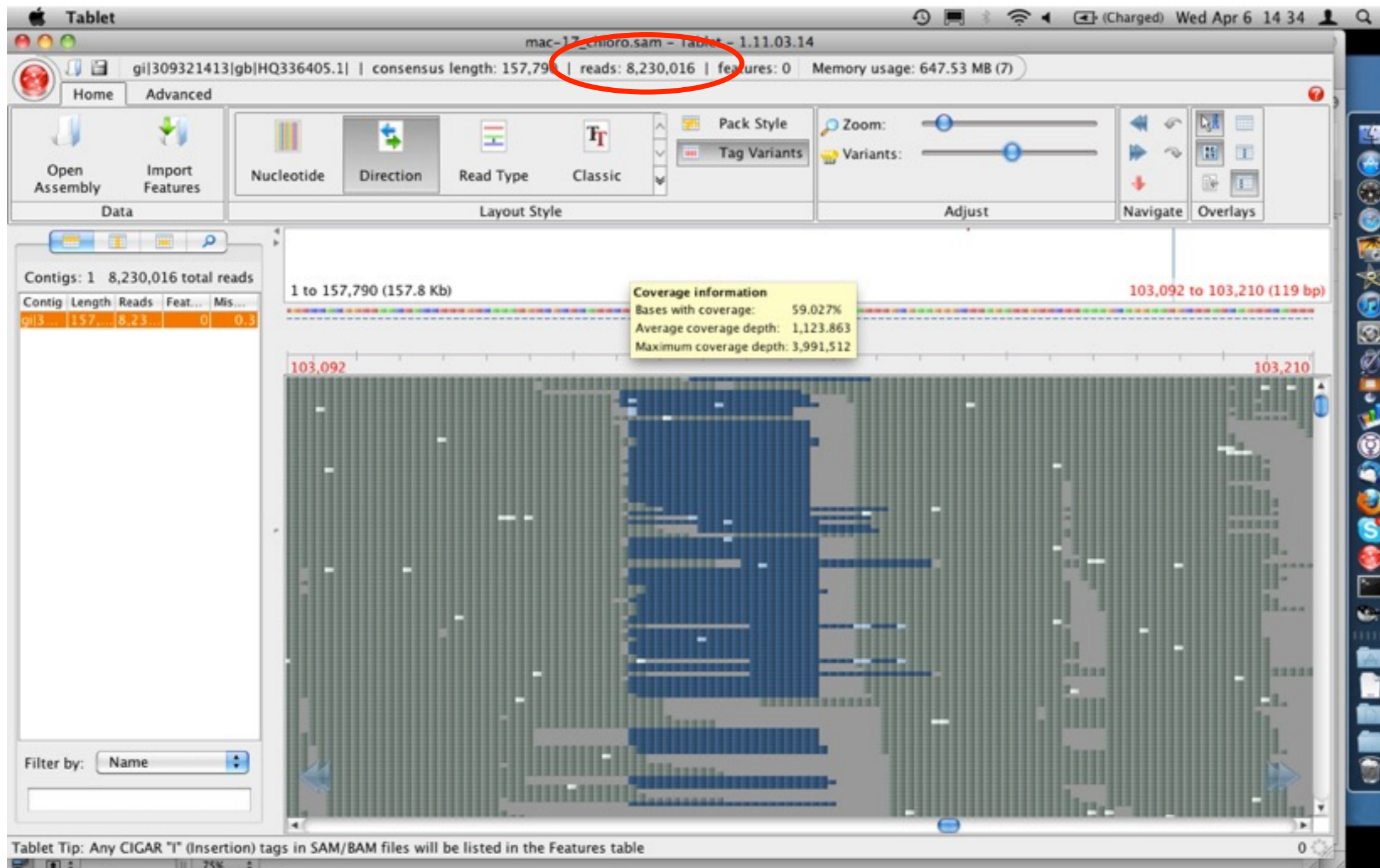
Data Quality control

Data Mapping

Data Visualization

Tablet

Data visualization
Peach tree chloroplast



Data Upload

- ★ sequencing reads
- ★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

GALAXY/Tablet

DONE !

