Next Generation Sequencing Tutorial

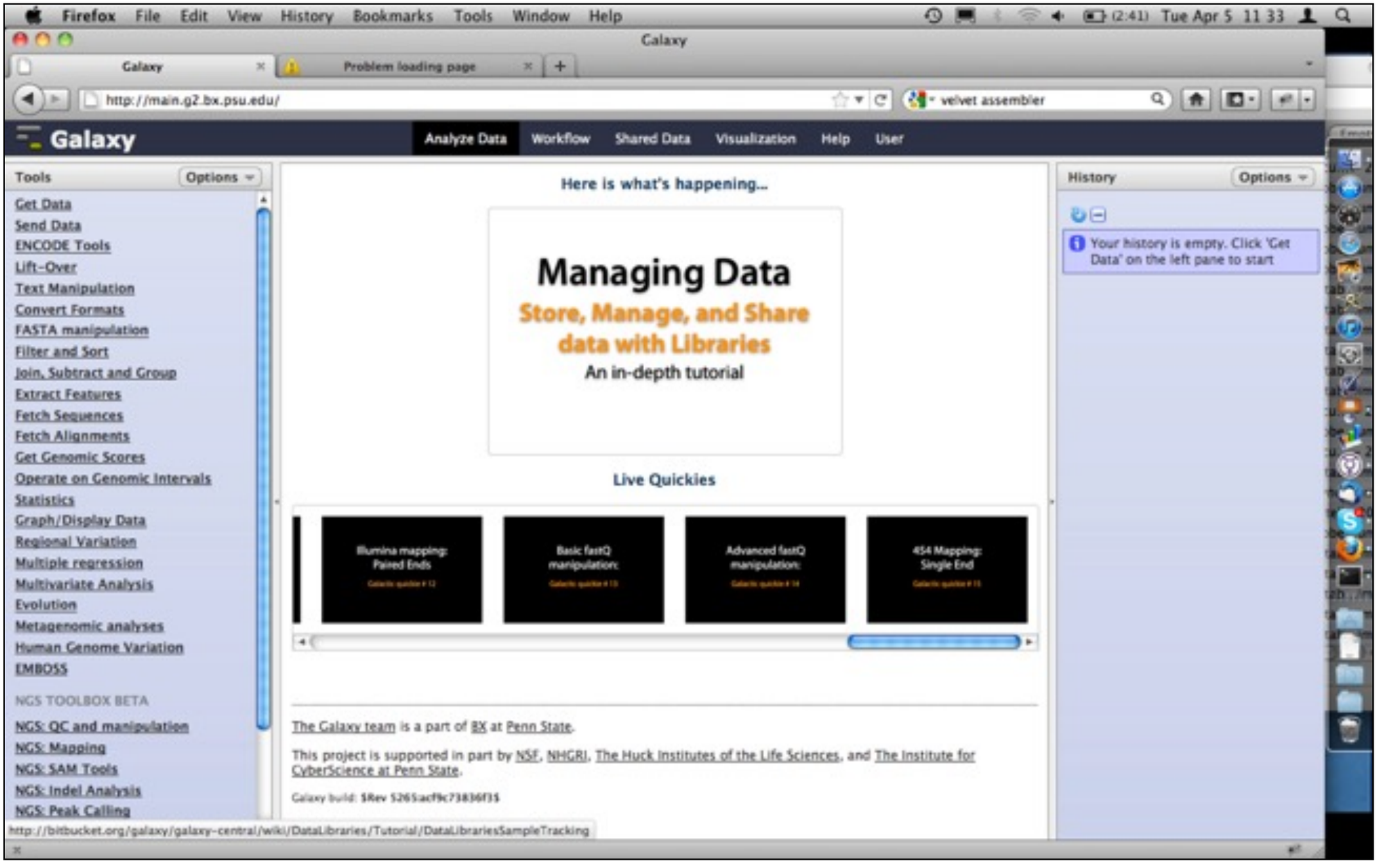
Andreas Gisel
International Institute of Tropical Agriculture (IITA)
Ibadan, Nigeria

a.gisel@cgiar.org

Working Platform
GALAXY

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

GALAXY


Firefox File Edit View History Bookmarks Tools Window Help
Galaxy
http://main.g2.bx.psu.edu/
velvet assembler
Galaxy
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Tools
Options
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
EMBOSS
NGS TOOLBOX BETA
NGS: QC and manipulation
NGS: Mapping
NGS: SAM Tools
NGS: Indel Analysis
NGS: Peak Calling
Here is what's happening...
Managing Data
Store, Manage, and Share data with Libraries
An in-depth tutorial
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
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: $Rev: 5265:acf9c73836f3$
http://bitbucket.org/galaxy/galaxy-central/wiki/DataLibraries/Tutorial/DataLibrariesSampleTracking

GALAXY

Workflow

Register user

Data Upload
★ sequencing reads
★ reference sequence

Data preparation

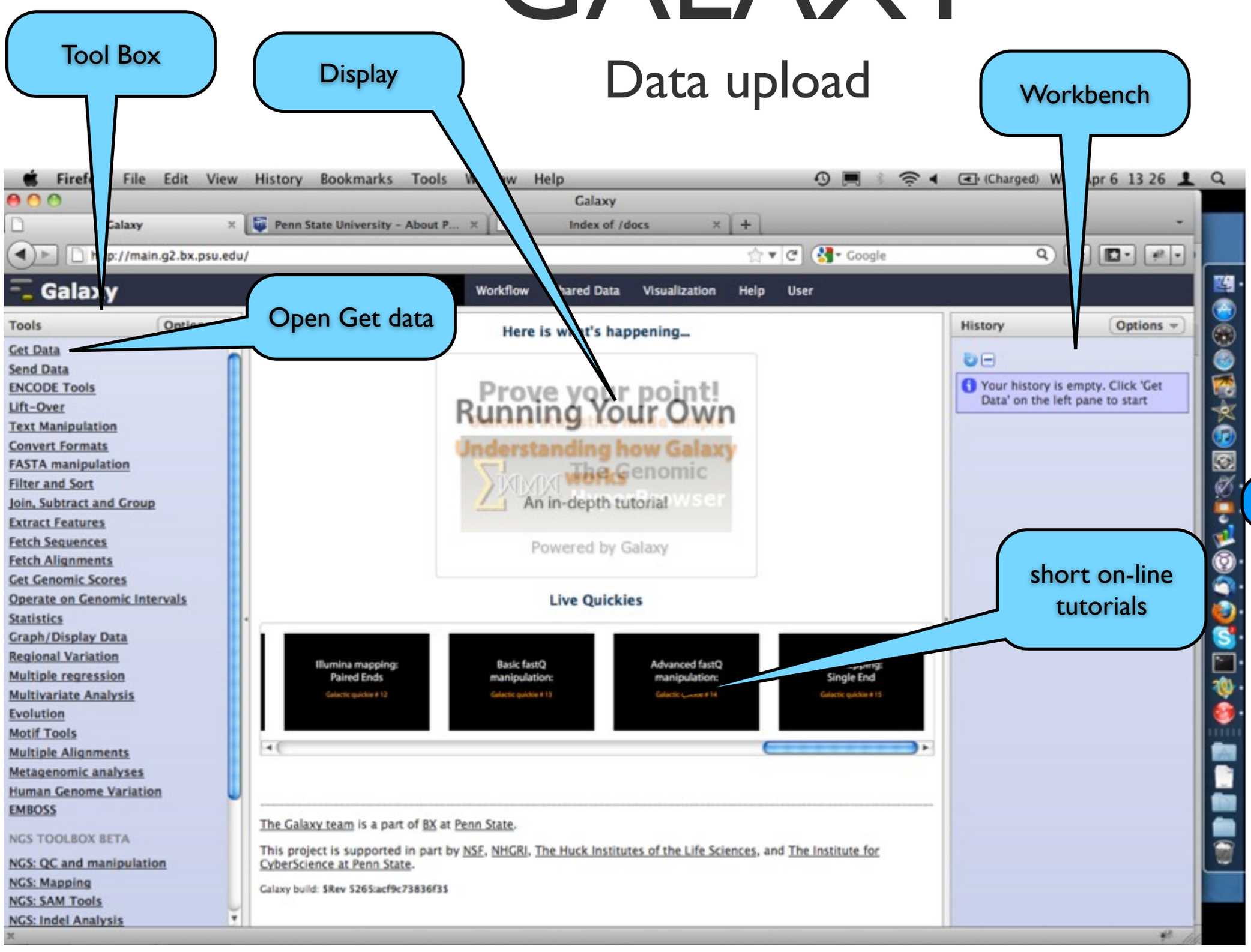
Data Quality control

Data Mapping

Data Visualization

GALAXY

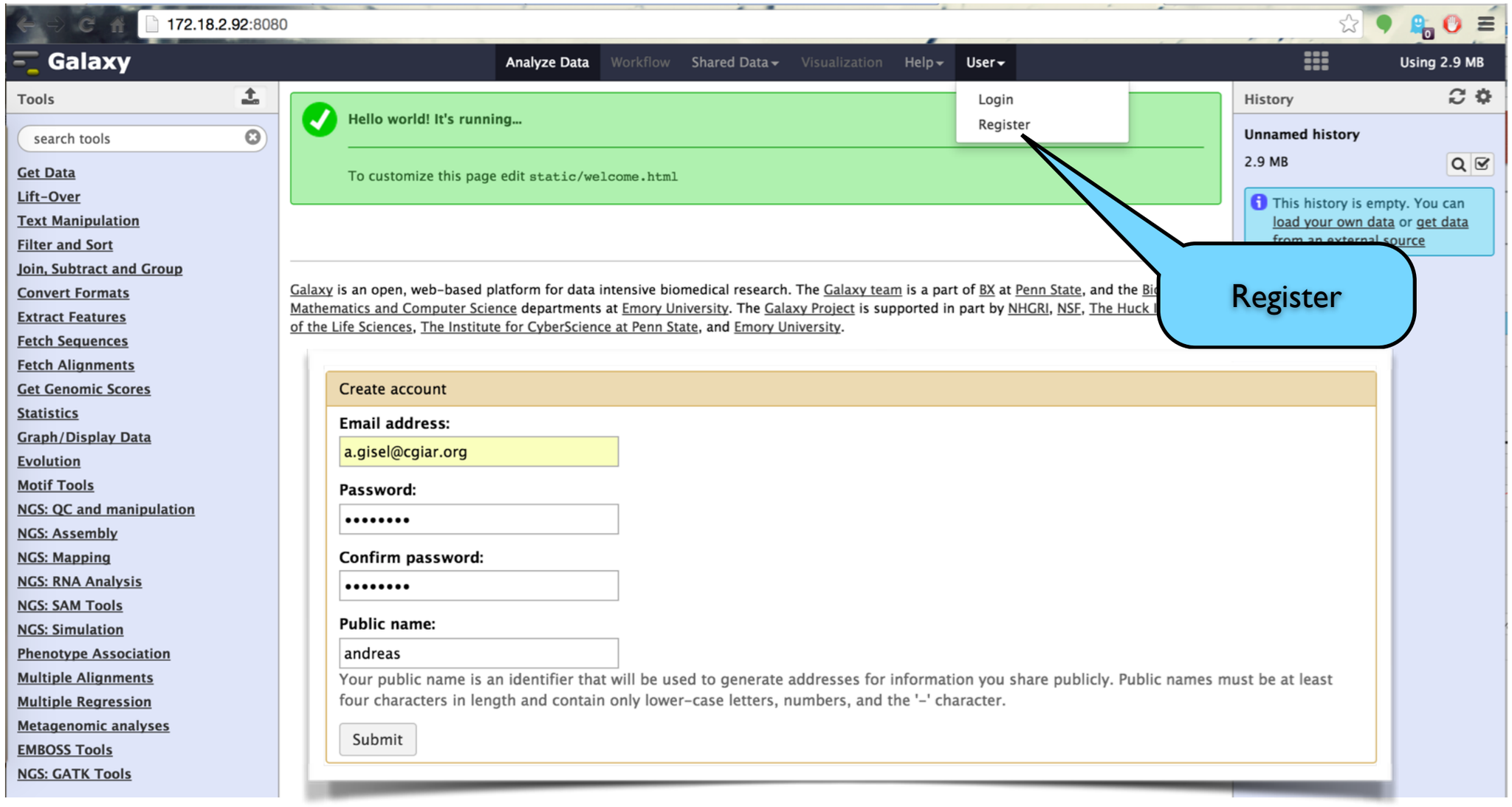
Data upload



- Data Upload
  - ★ sequencing reads
  - ★ reference
- Data preparation
- Data Quality control
- Data Mapping
- Data Visualization

GALAXY

Data upload



Register user

Data Upload
★ sequencing reads
★ reference

Data preparation

Data Quality control

Data Mapping

Data Visualization

GALAXY

Data upload

Data Upload
★ sequencing reads
★ reference

Data preparation

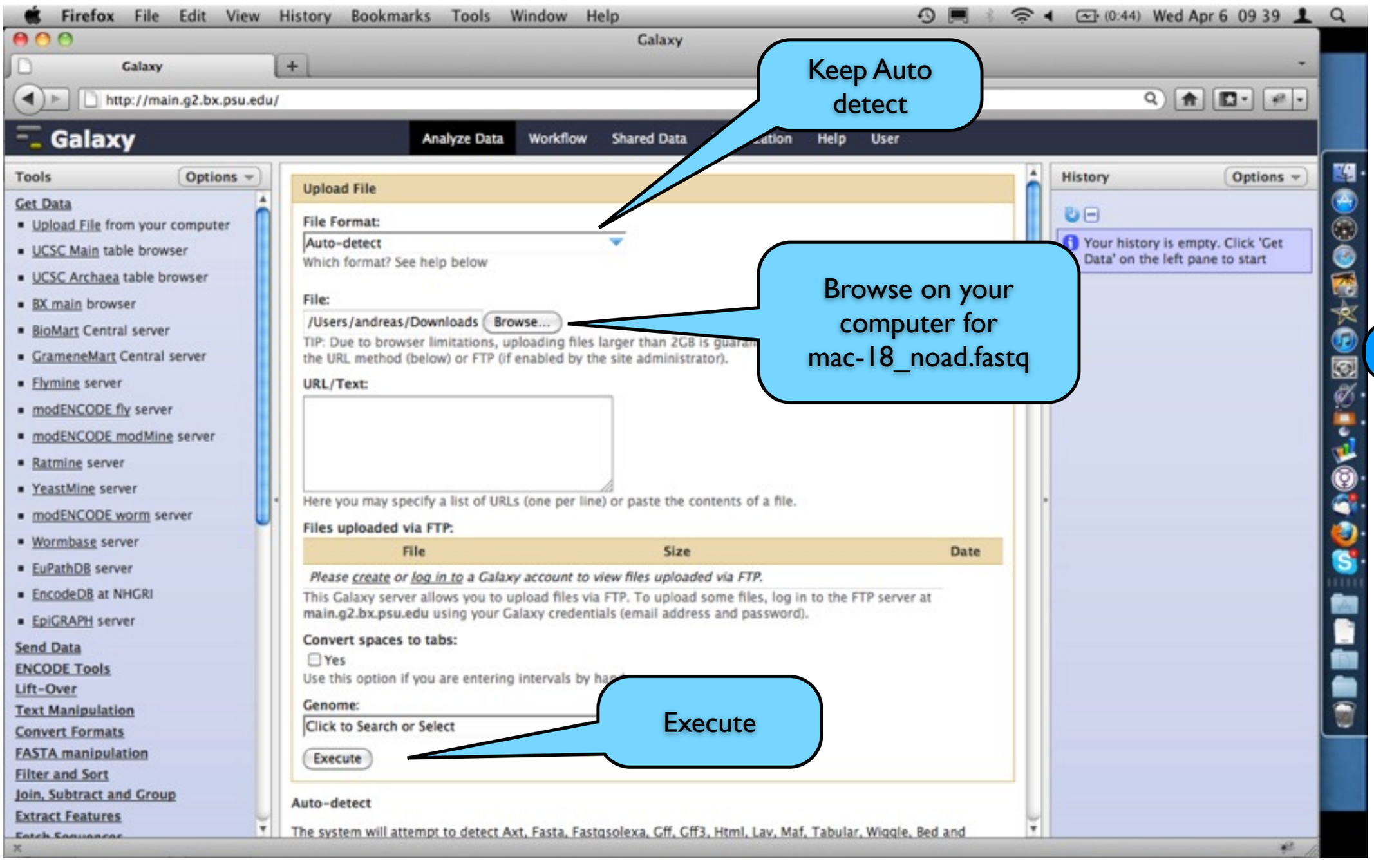
Data Quality control

Data Mapping

Data Visualization

GALAXY

Data upload



Data Upload
★ sequencing reads
★ reference

Data preparation

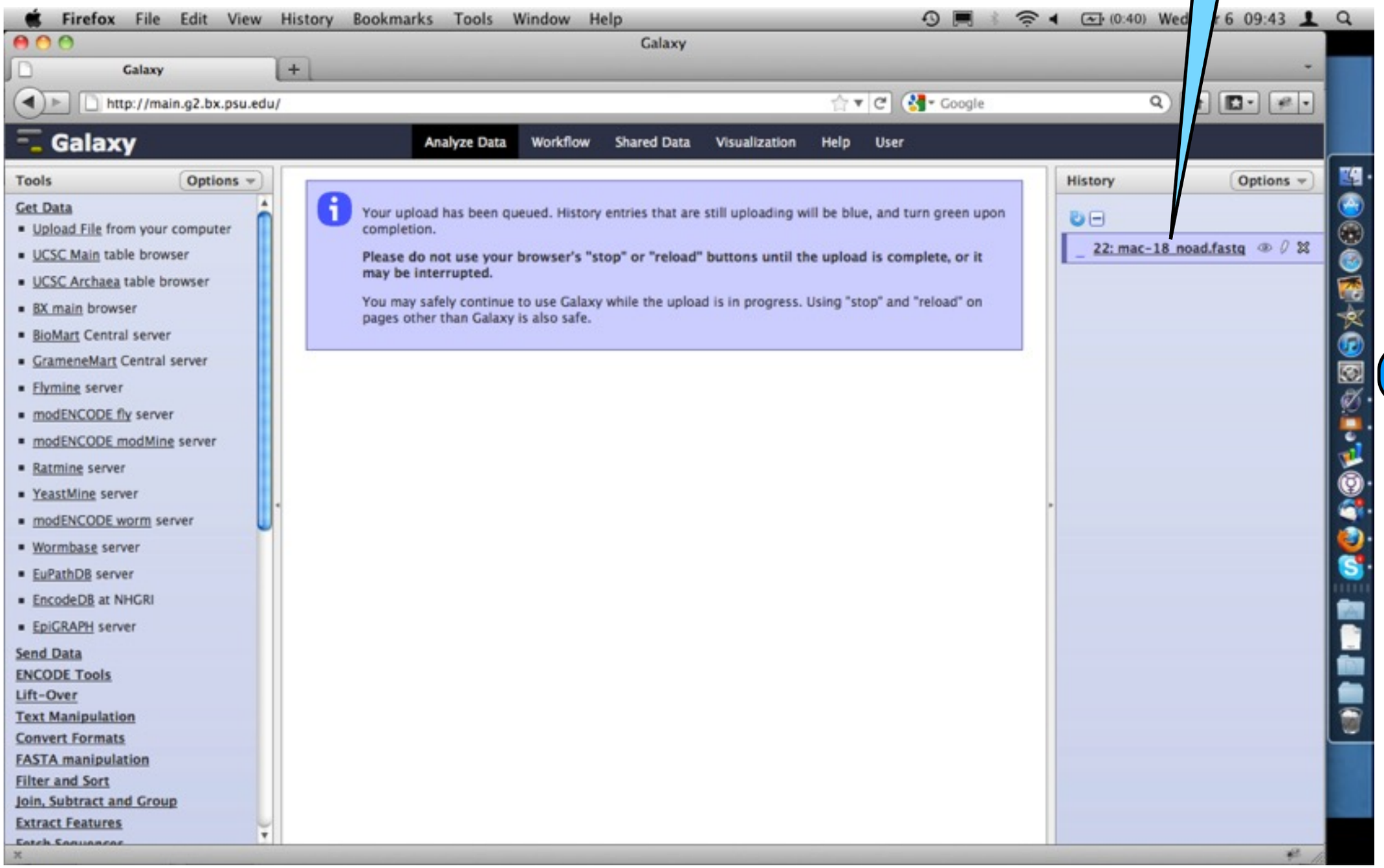
Data Quality control

Data Mapping

Data Visualization

GALAXY

Data upload



Data Upload
★ sequencing reads
★ reference

Data preparation

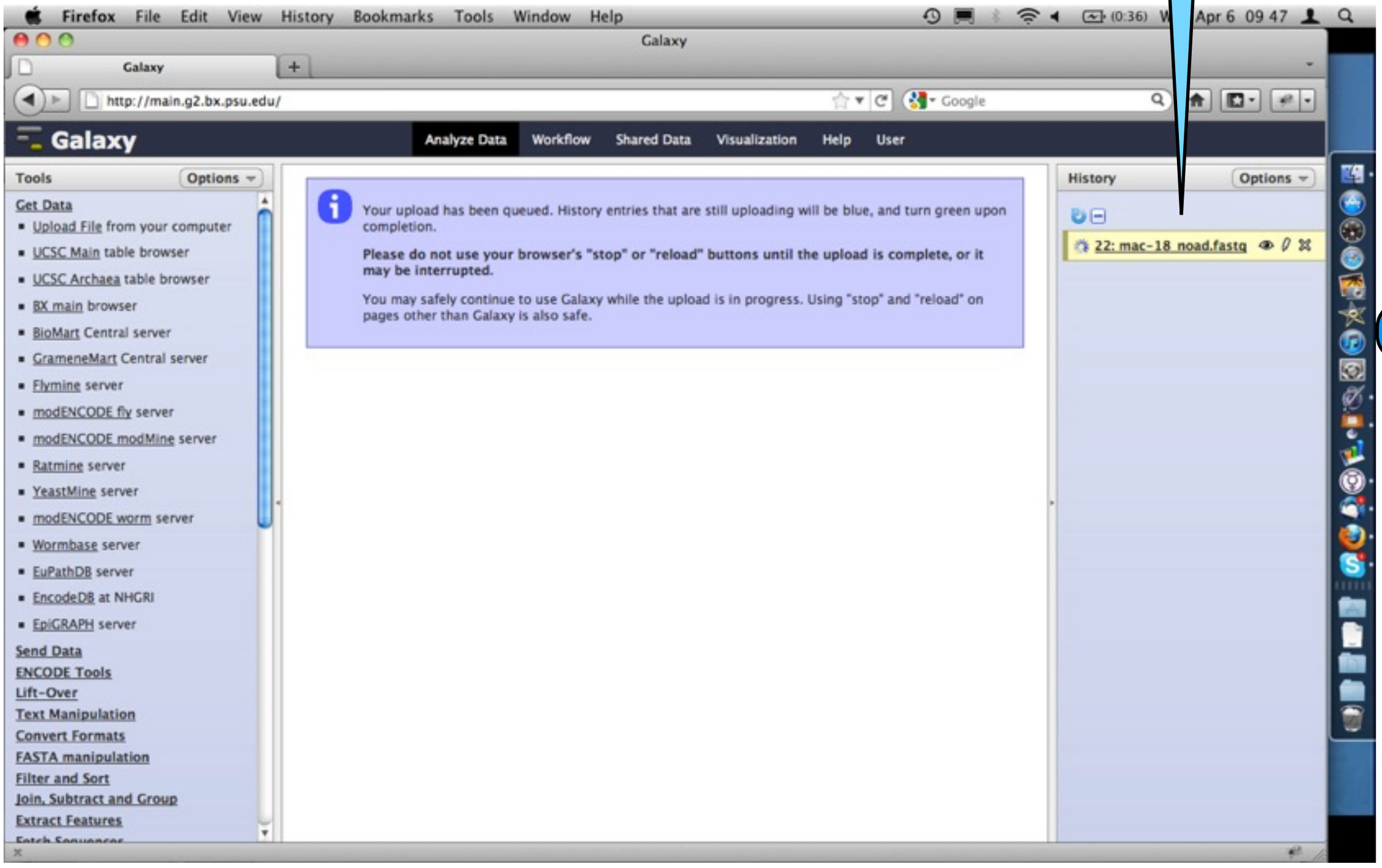
Data Quality control

Data Mapping

Data Visualization

GALAXY

Data upload



Data Upload
★ sequencing reads
★ reference

Data preparation

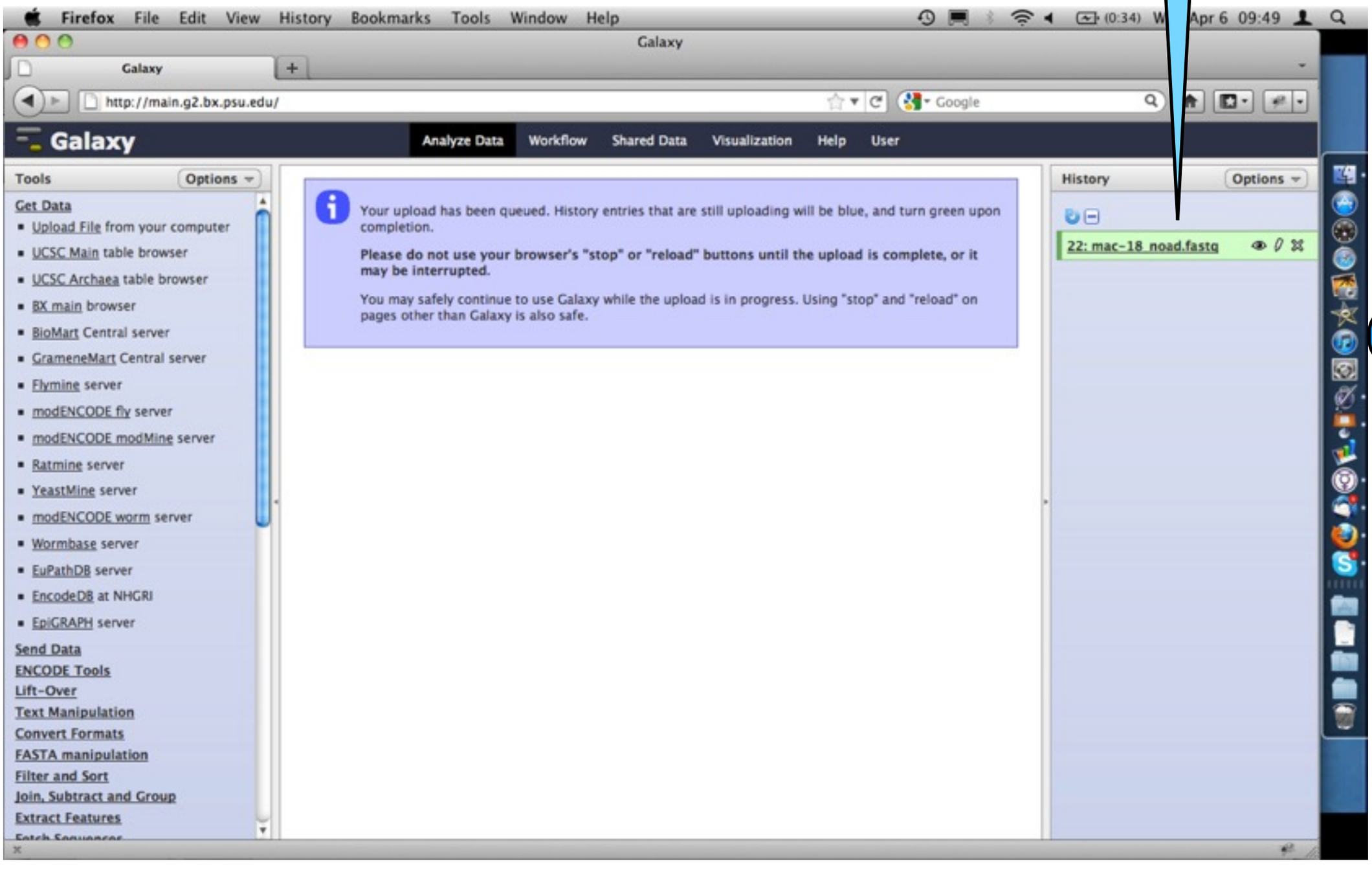
Data Quality control

Data Mapping

Data Visualization

GALAXY

Data upload



Data Upload
★ sequencing reads
★ reference

Data preparation

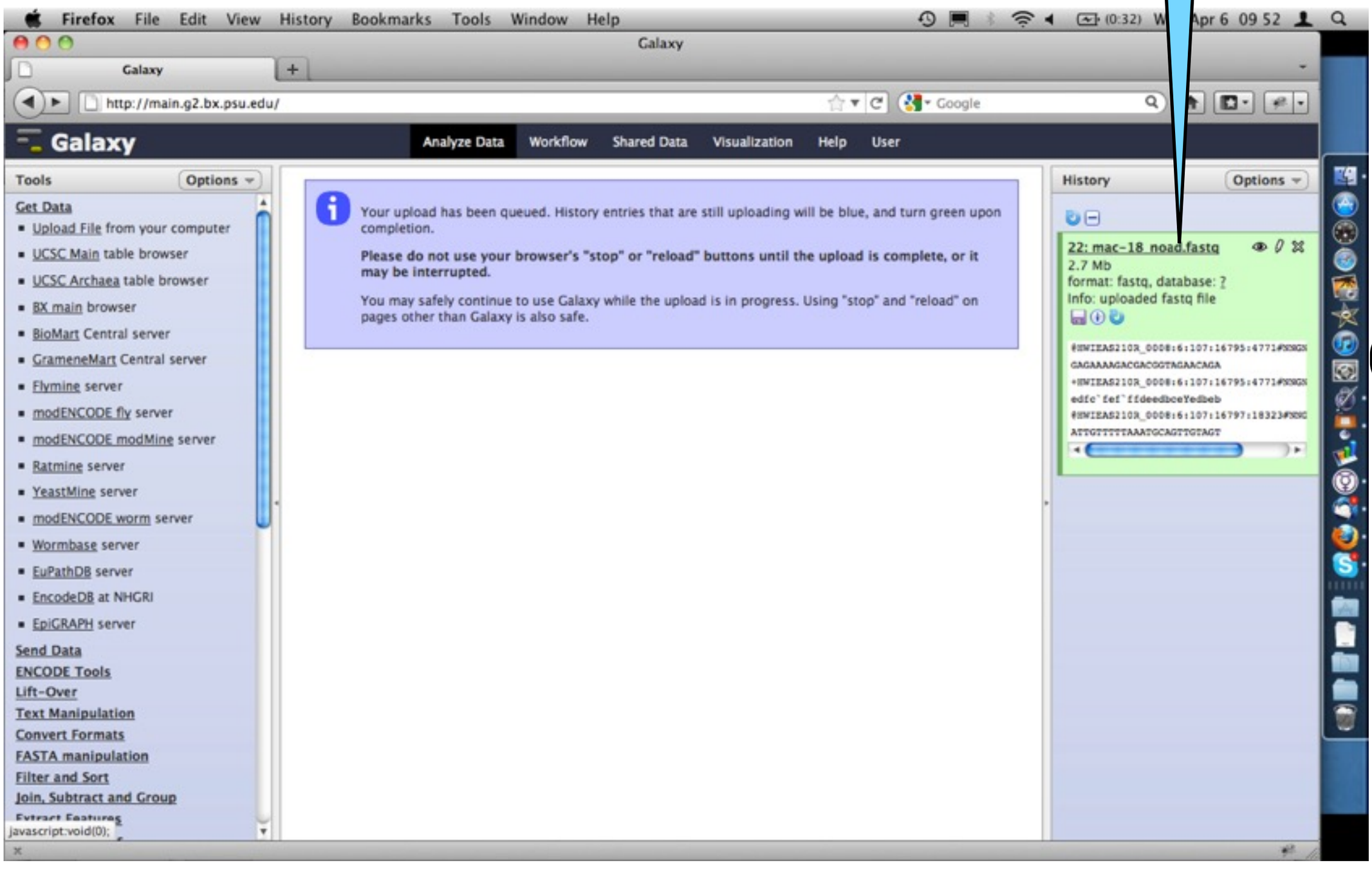
Data Quality control

Data Mapping

Data Visualization

GALAXY

Data upload



Data Upload
★ sequencing reads
★ reference

Data preparation

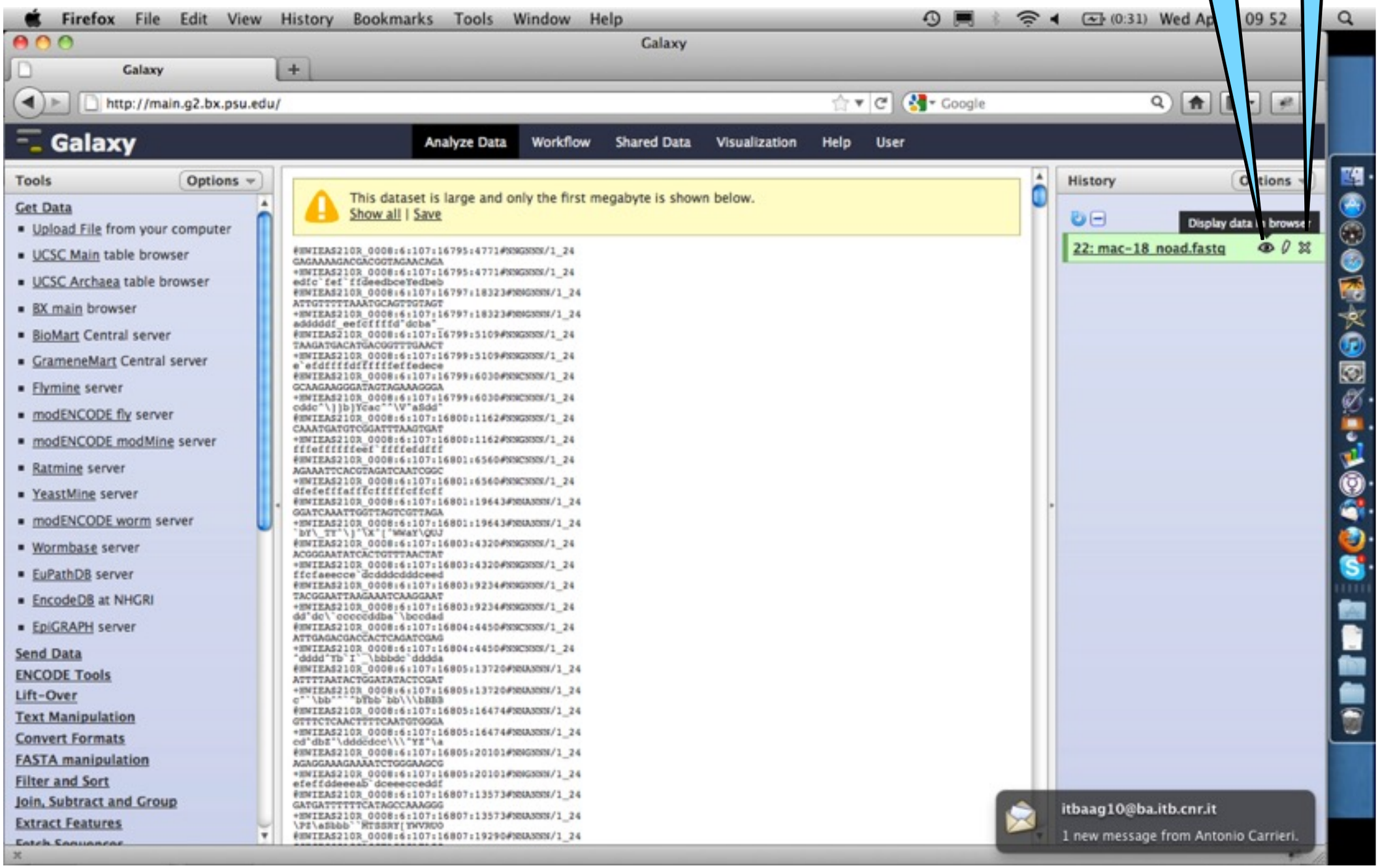
Data Quality control

Data Mapping

Data Visualization

GALAXY

Data upload



Data Upload
★ sequencing reads
★ reference

Data preparation

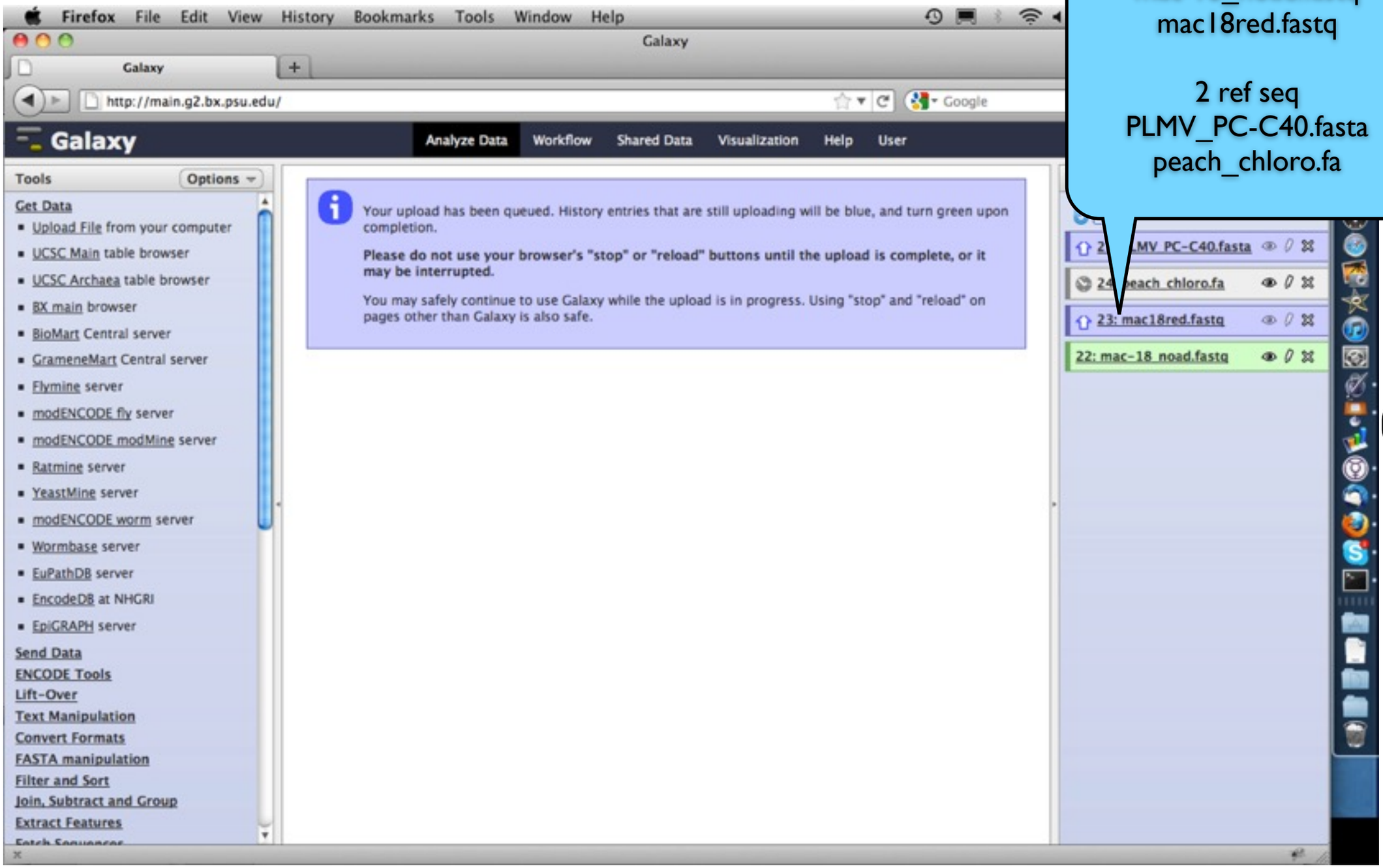
Data Quality control

Data Mapping

Data Visualization

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

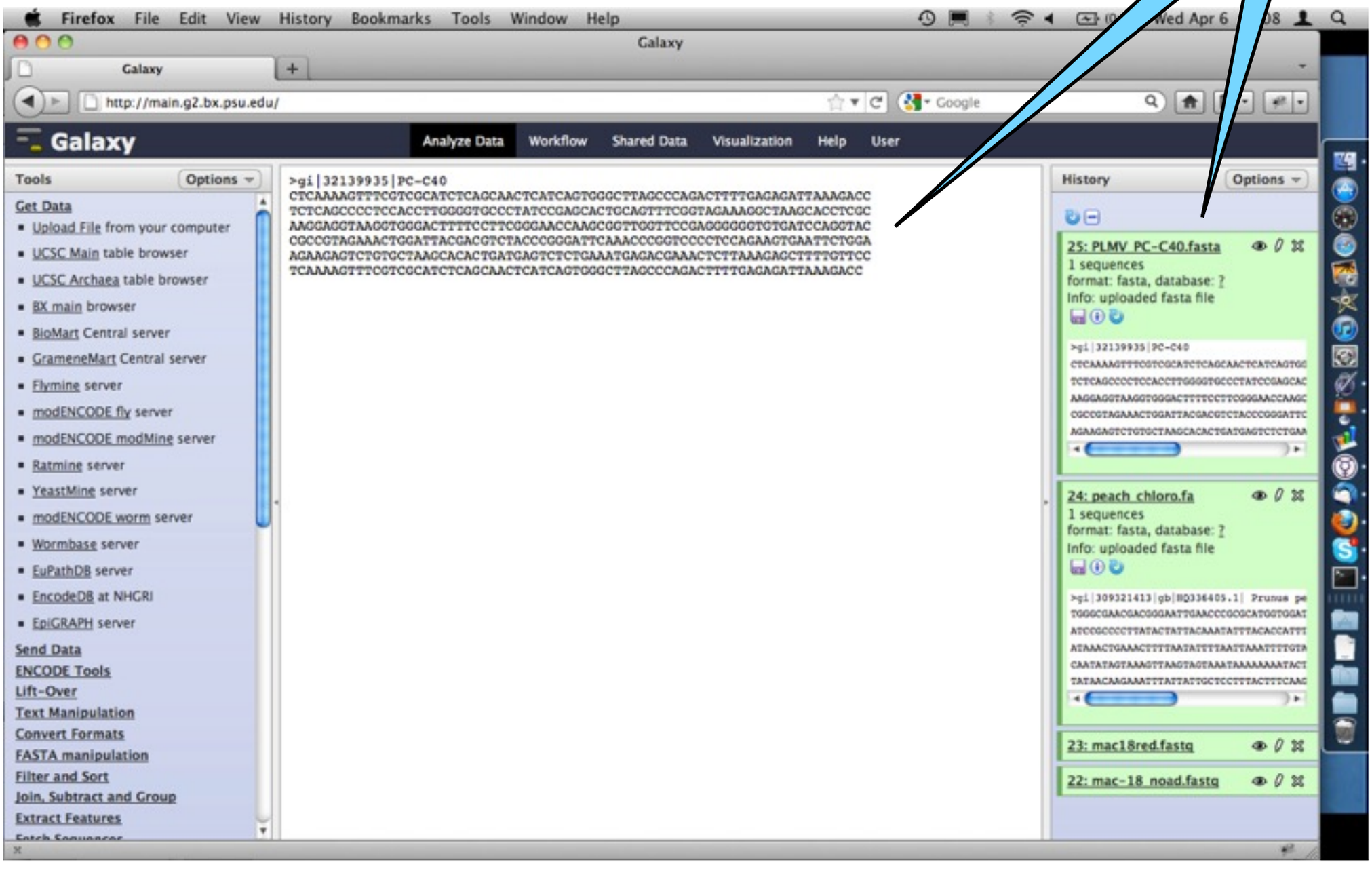
Data Quality control

Data Mapping

Data Visualization

GALAXY

Data upload



2 fastq files
2 fasta files

Data Upload
★ sequencing reads
★ reference

Data preparation

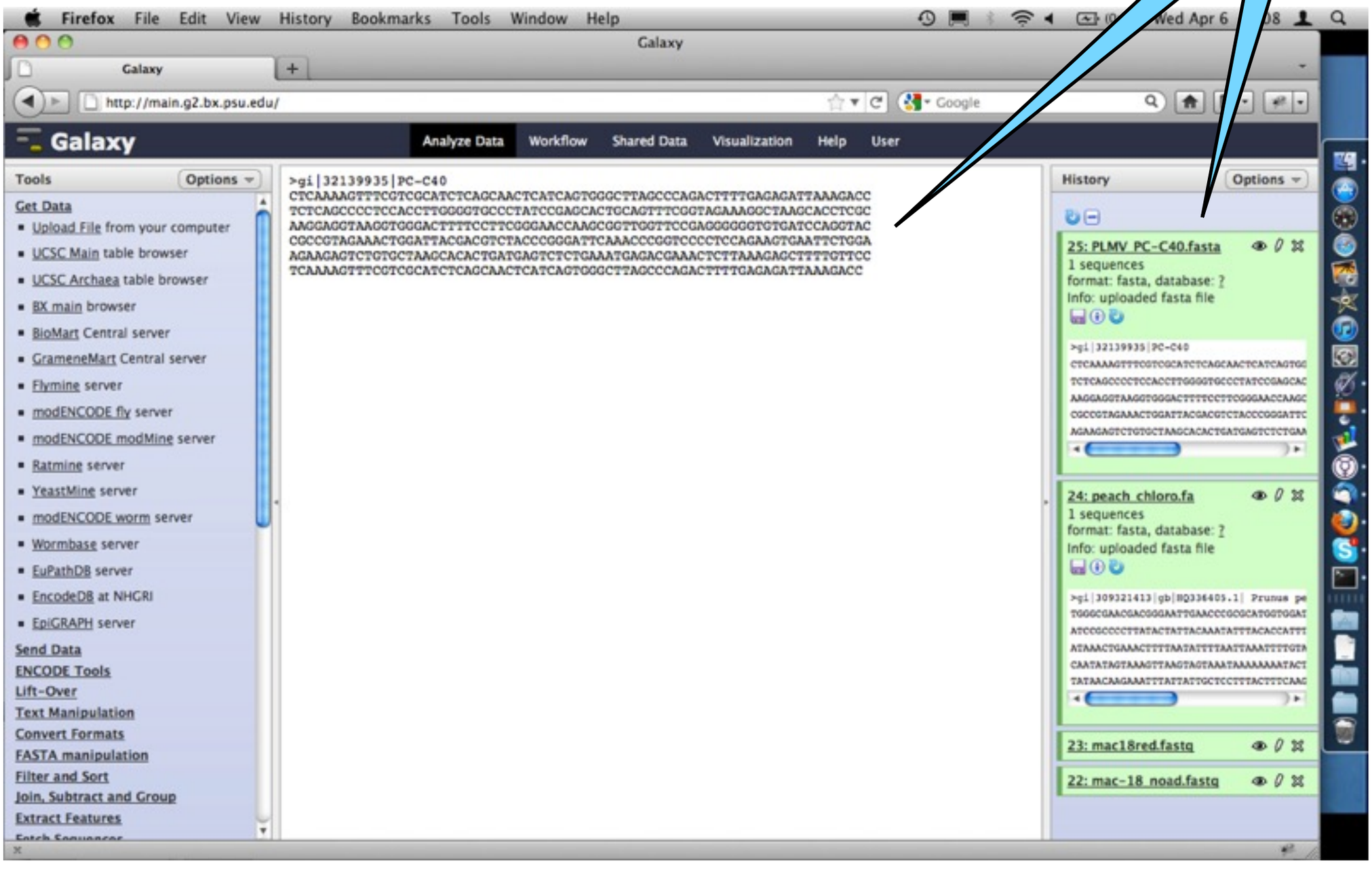
Data Quality control

Data Mapping

Data Visualization

GALAXY

Data upload



2 fastq files
2 fasta files

Data Upload
★ sequencing reads
★ reference

Data preparation

Data Quality control

Data Mapping

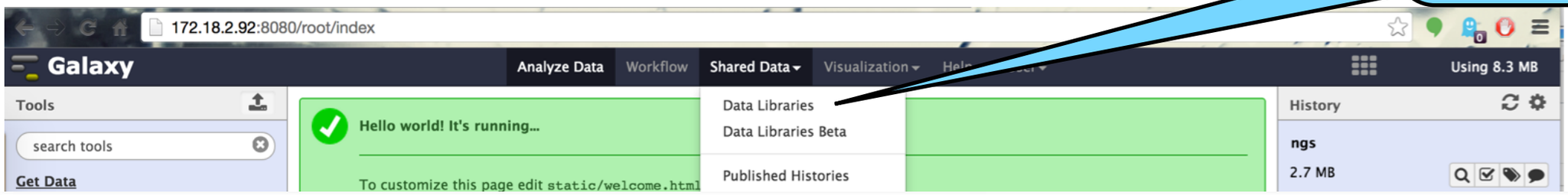
Data Visualization

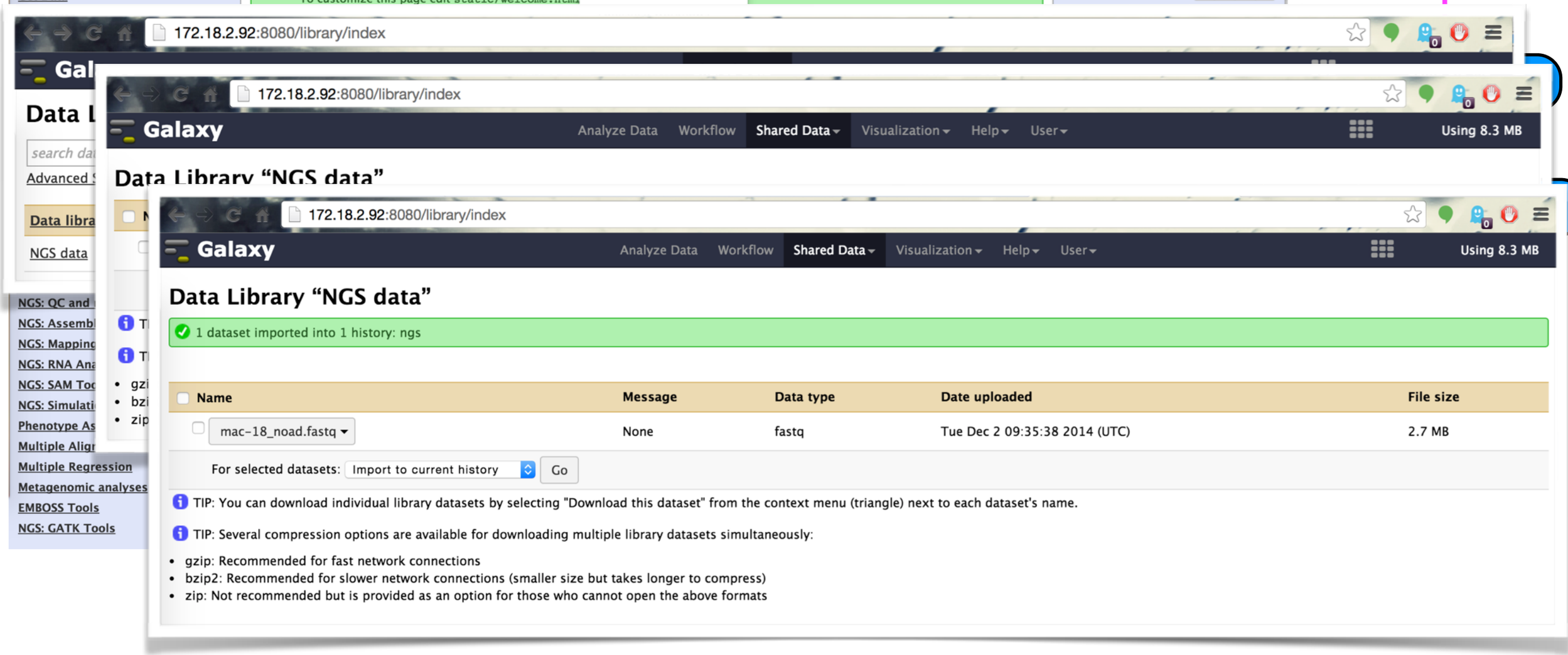
GALAXY

Data upload

Data Libraries

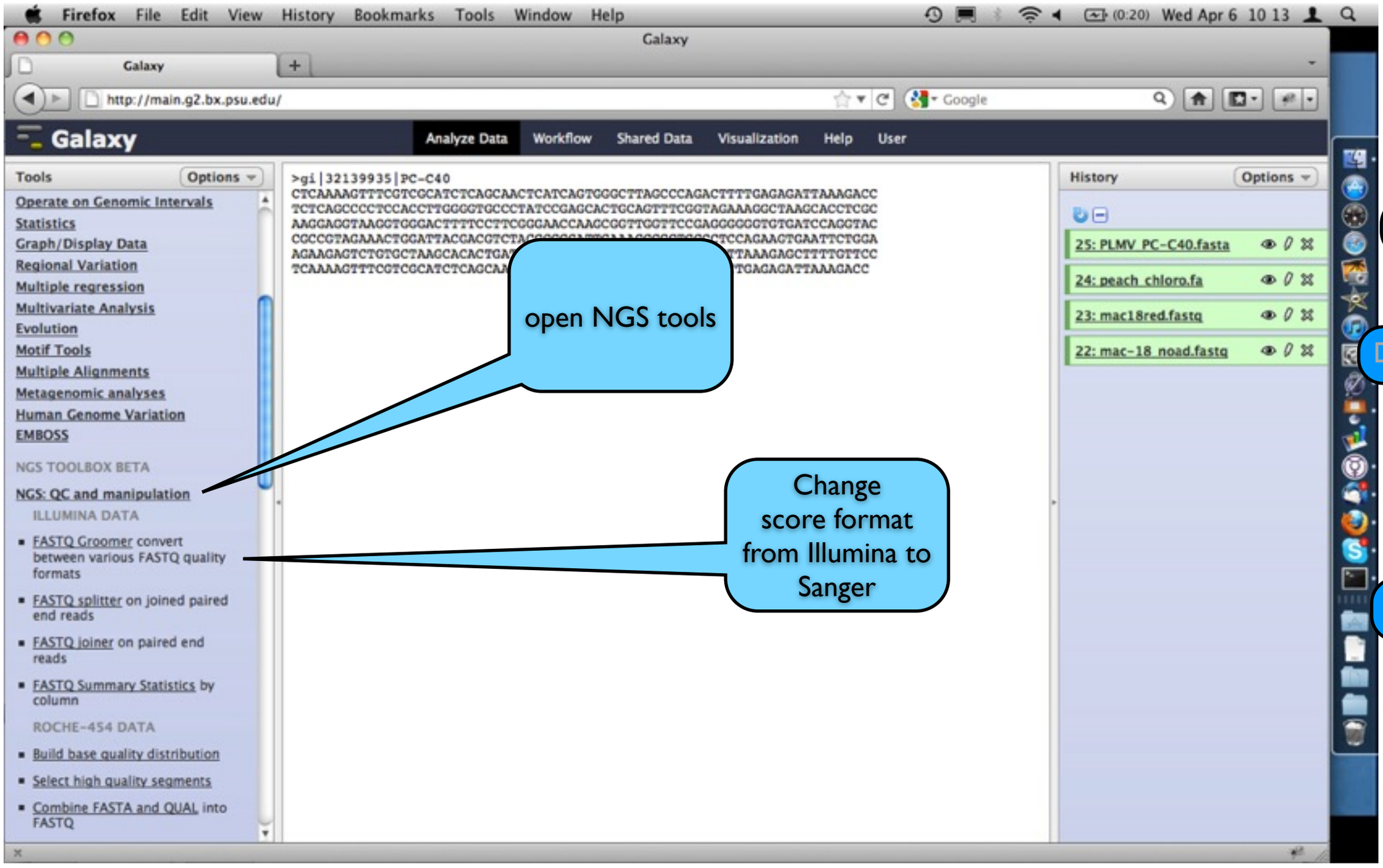
Data Upload
★ sequencing reads
★ reference





GALAXY

Data preparation



Data Upload
★ sequencing reads
★ reference

Data preparation

Data Quality control

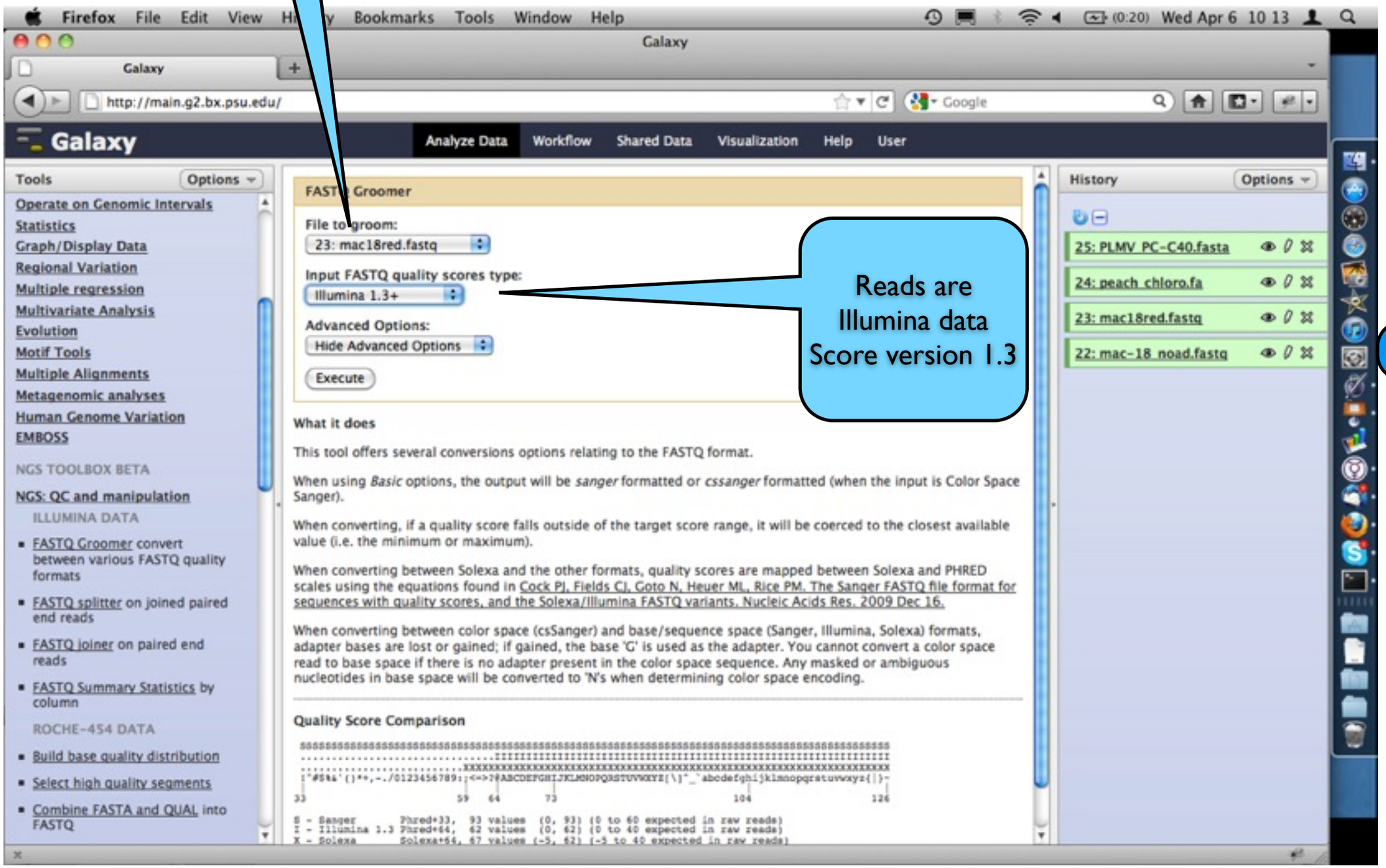
Data Mapping

Data Visualization

GALAXY

Data preparation

Choose data to work on



Reads are Illumina data Score version 1.3

Data Upload
★ sequencing reads
★ reference

Data preparation

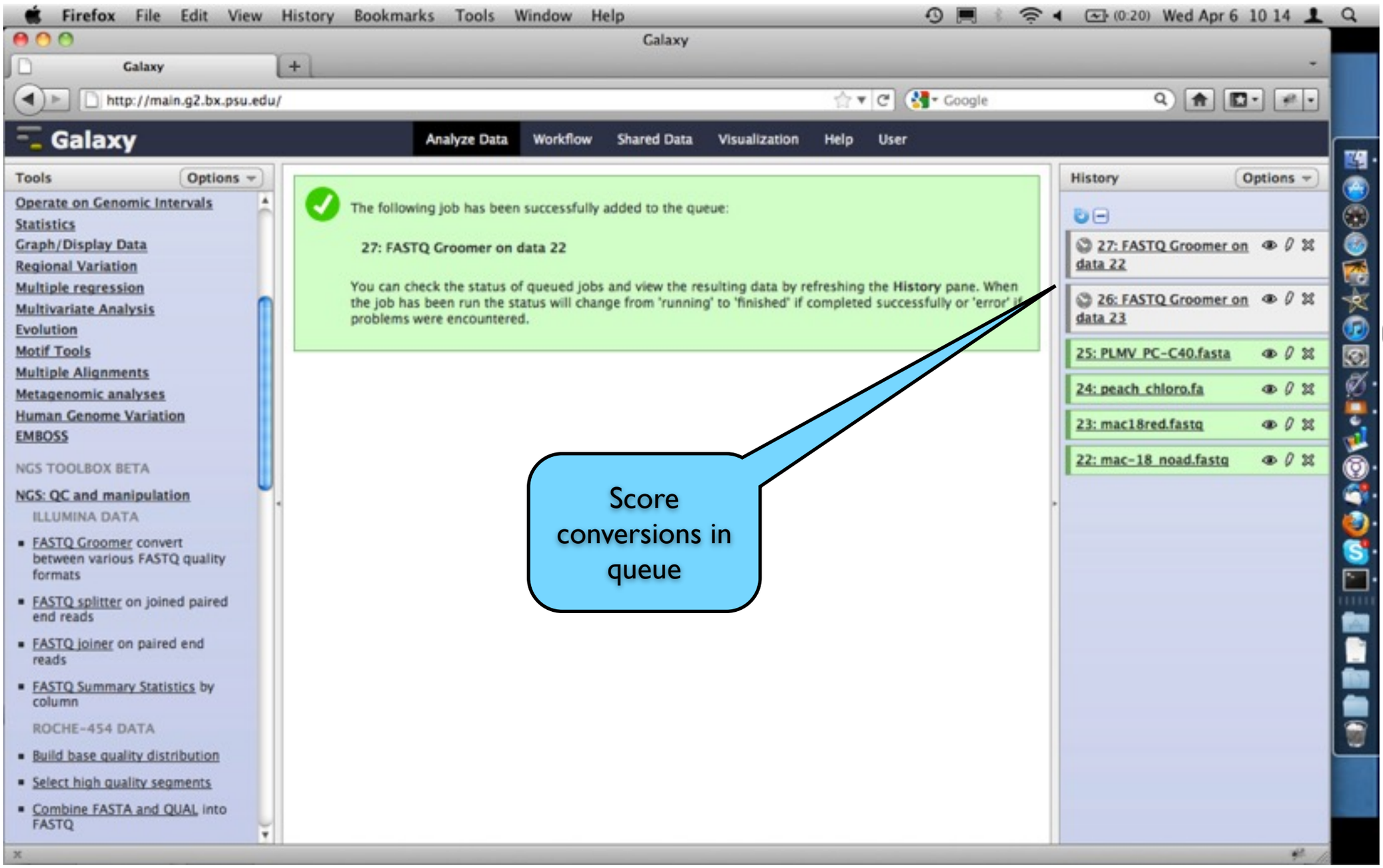
Data Quality control

Data Mapping

Data Visualization

GALAXY

Data preparation



Data Upload
★ sequencing reads
★ reference

Data preparation

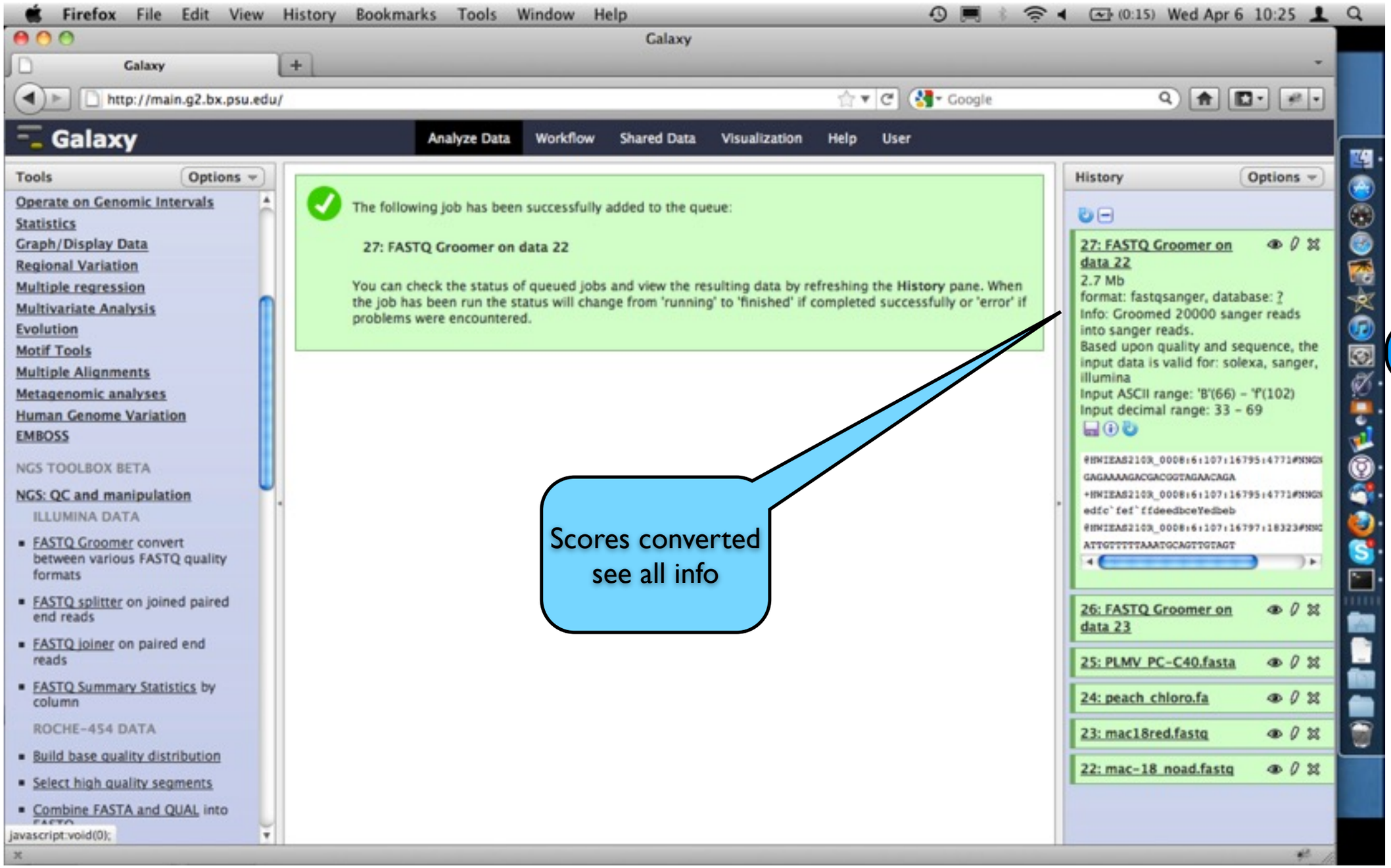
Data Quality control

Data Mapping

Data Visualization

GALAXY

Data preparation



Data Upload
★ sequencing reads
★ reference

Data preparation

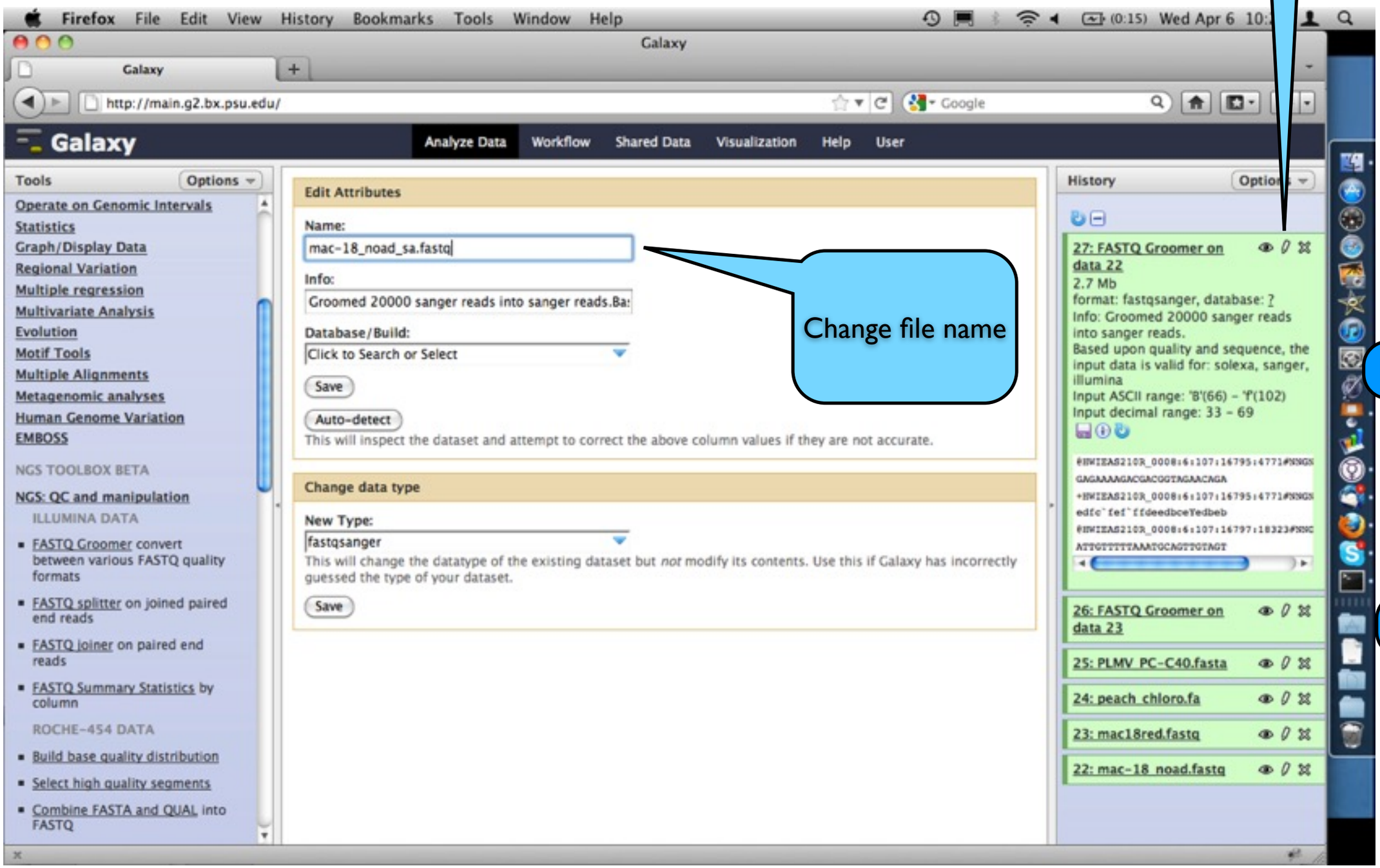
Data Quality control

Data Mapping

Data Visualization

GALAXY

Data preparation



Data Upload
★ sequencing reads
★ reference

Data preparation

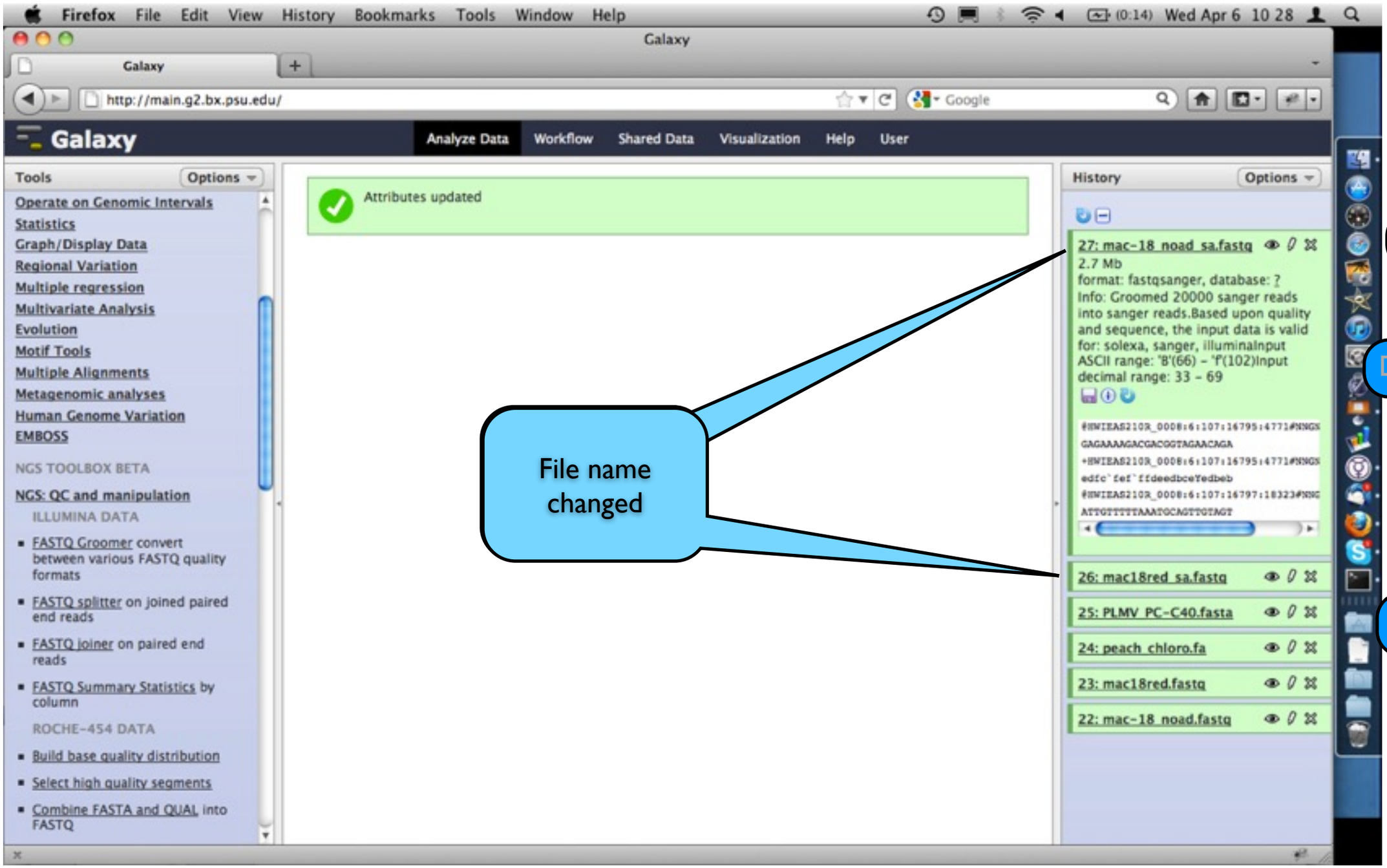
Data Quality control

Data Mapping

Data Visualization

GALAXY

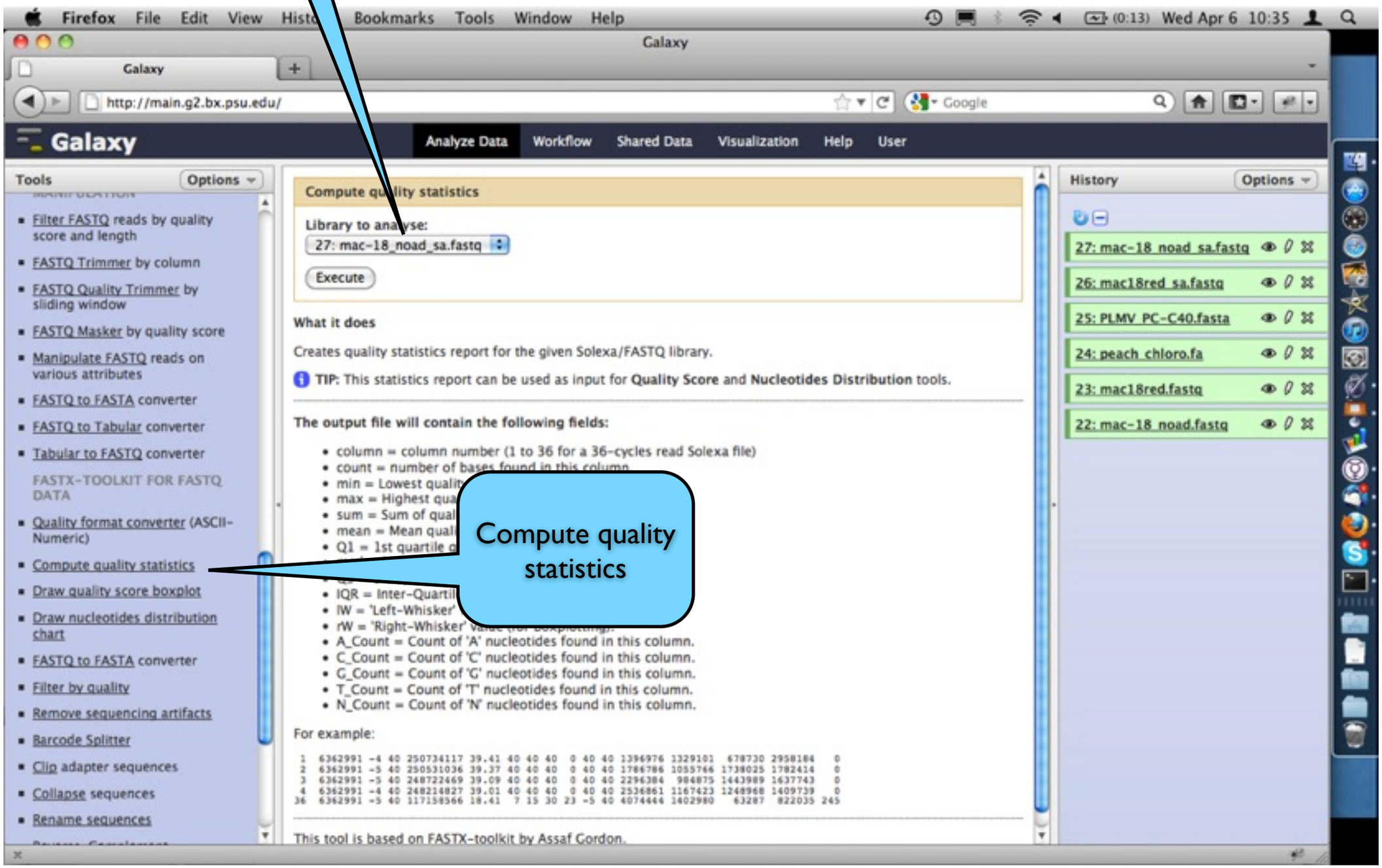
Data preparation



GALAXY

Data quality control

Choose data set



Data Upload
★ sequencing reads
★ reference

Data preparation

Data Quality control

Data Mapping

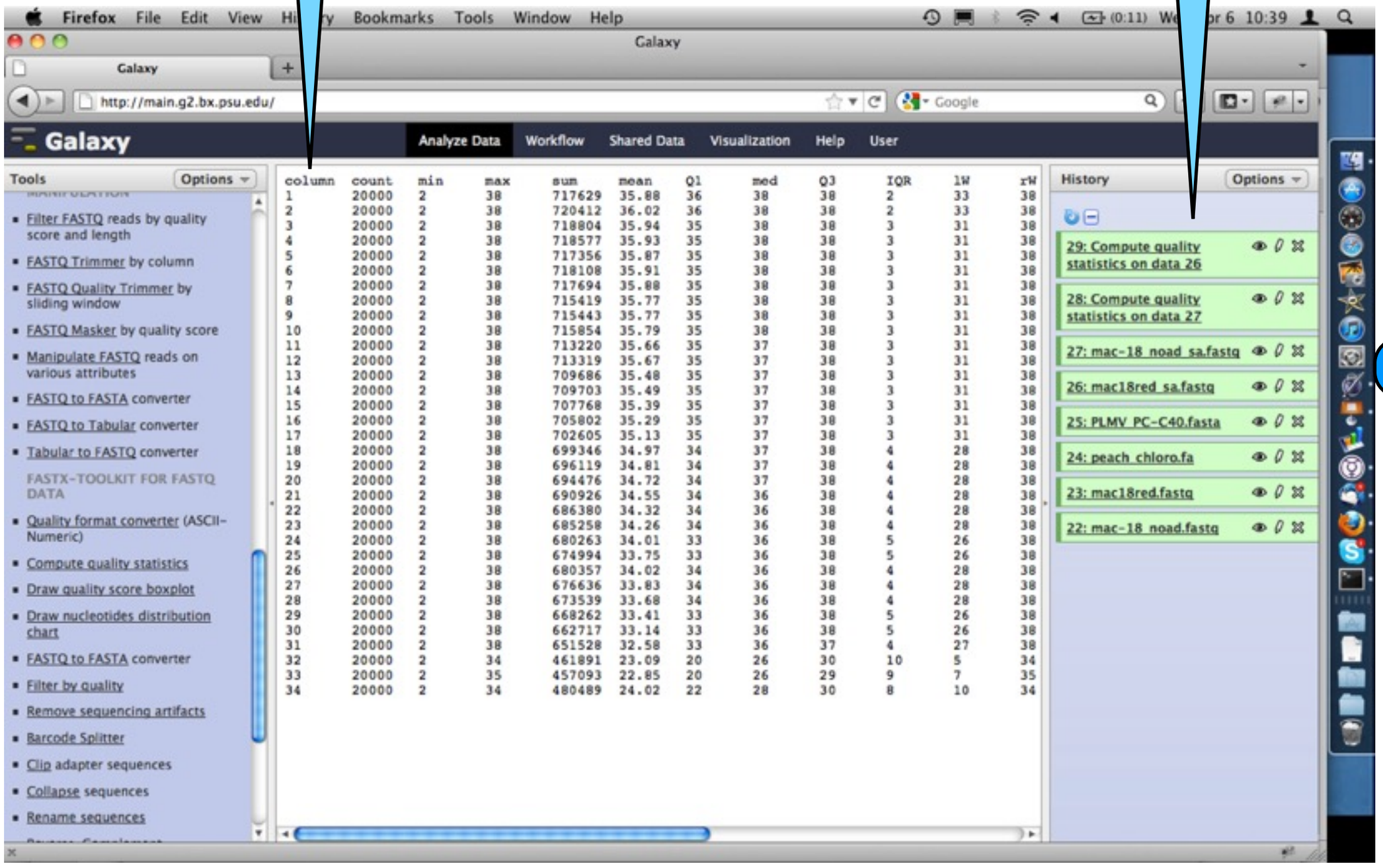
Data Visualization

GALAXY

Data quality control

nuleotide position

Statistics



column	count	min	max	sum	mean	Q1	med	Q3	IQR	lW	rW
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

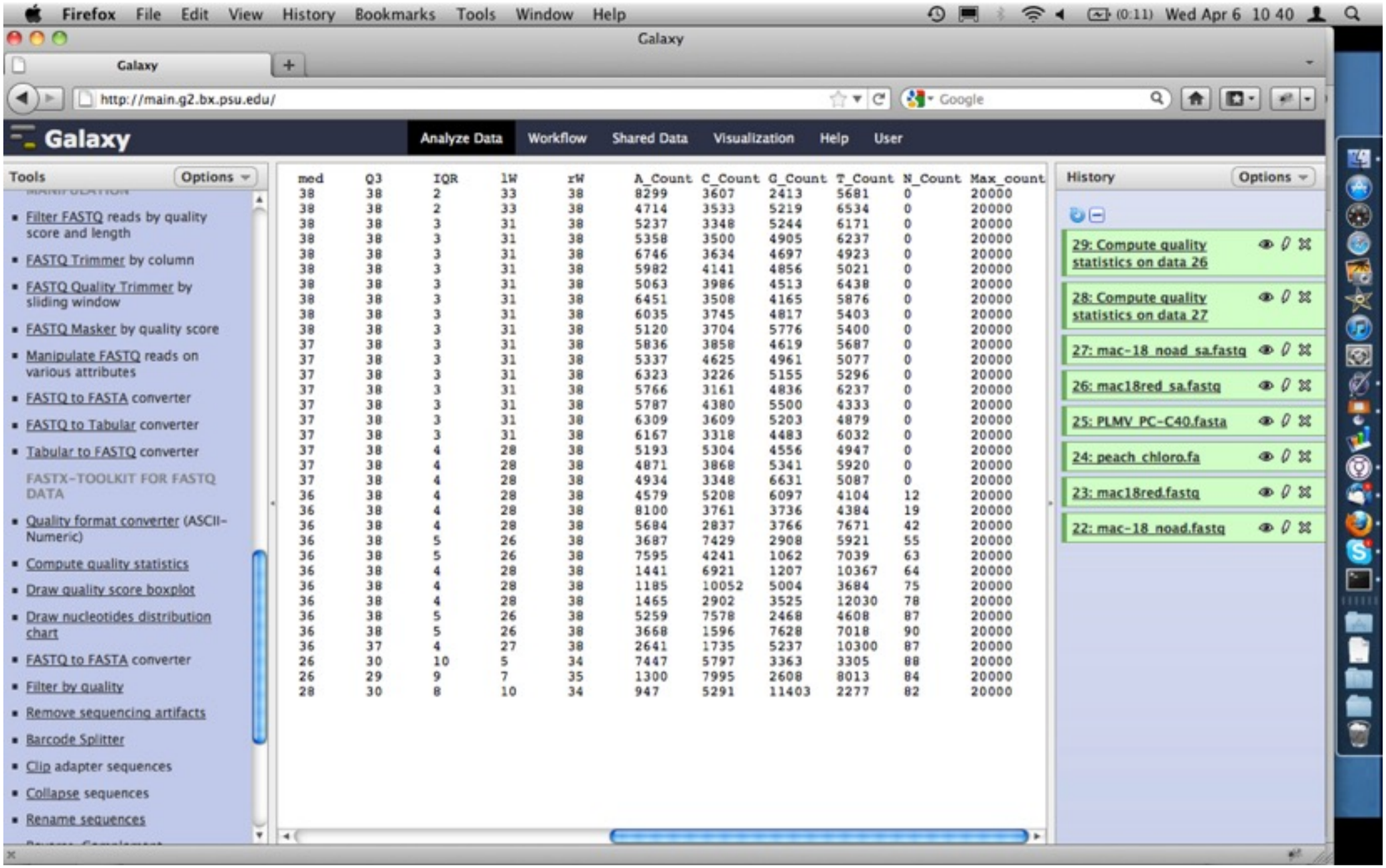
Data Quality control

Data Mapping

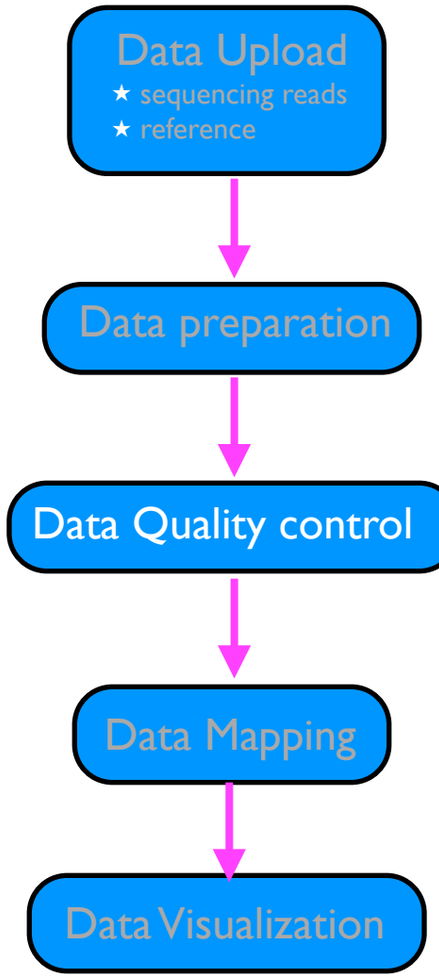
Data Visualization

GALAXY

Data quality control

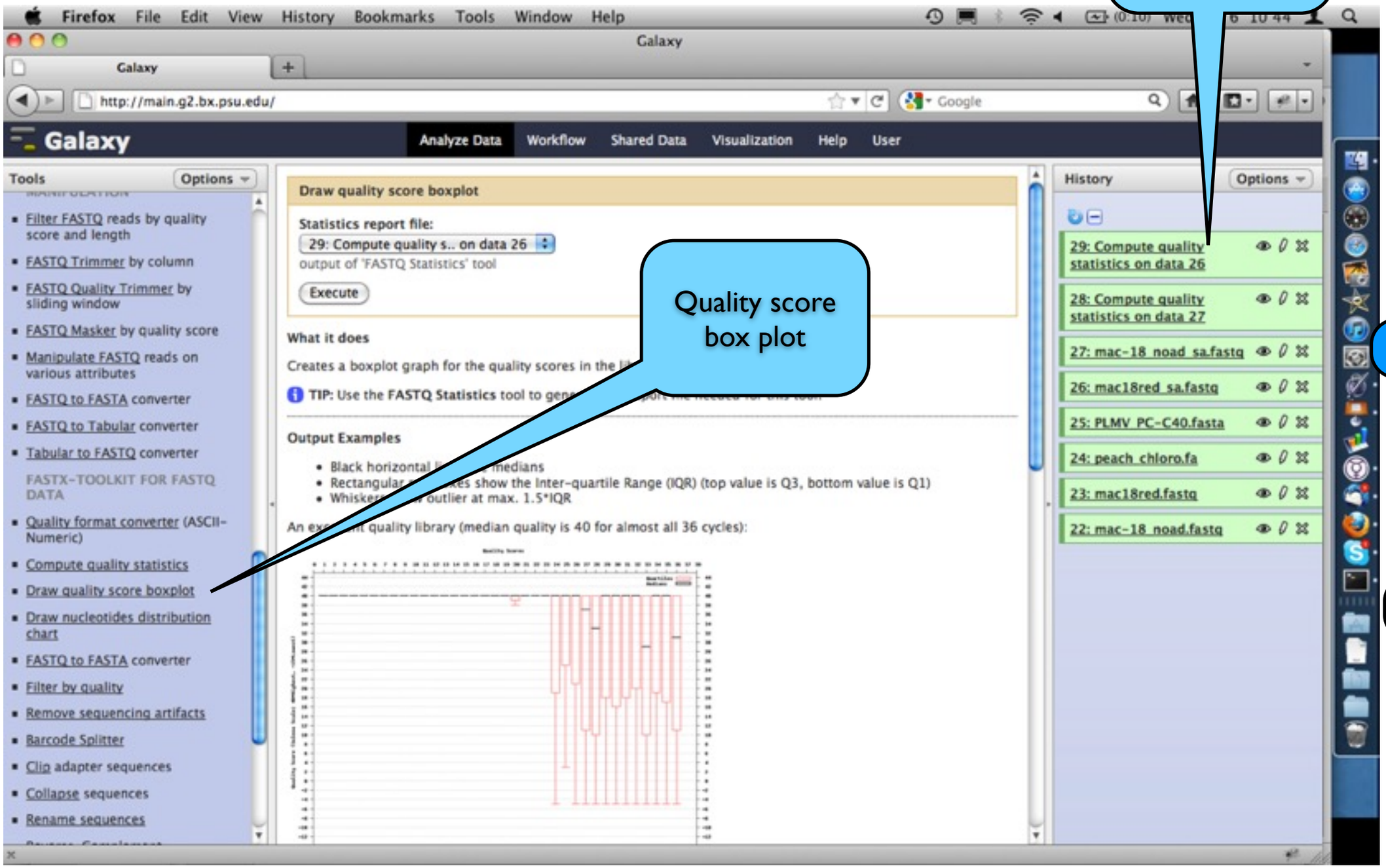


med	Q3	IQR	lW	rW	A_Count	C_Count	G_Count	T_Count	N_Count	Max_count
38	38	2	33	38	8299	3607	2413	5681	0	20000
38	38	2	33	38	4714	3533	5219	6534	0	20000
38	38	3	31	38	5237	3348	5244	6171	0	20000
38	38	3	31	38	5358	3500	4905	6237	0	20000
38	38	3	31	38	6746	3634	4697	4923	0	20000
38	38	3	31	38	5982	4141	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



GALAXY

Data quality control



Data Upload
★ sequencing reads
★ reference

Data preparation

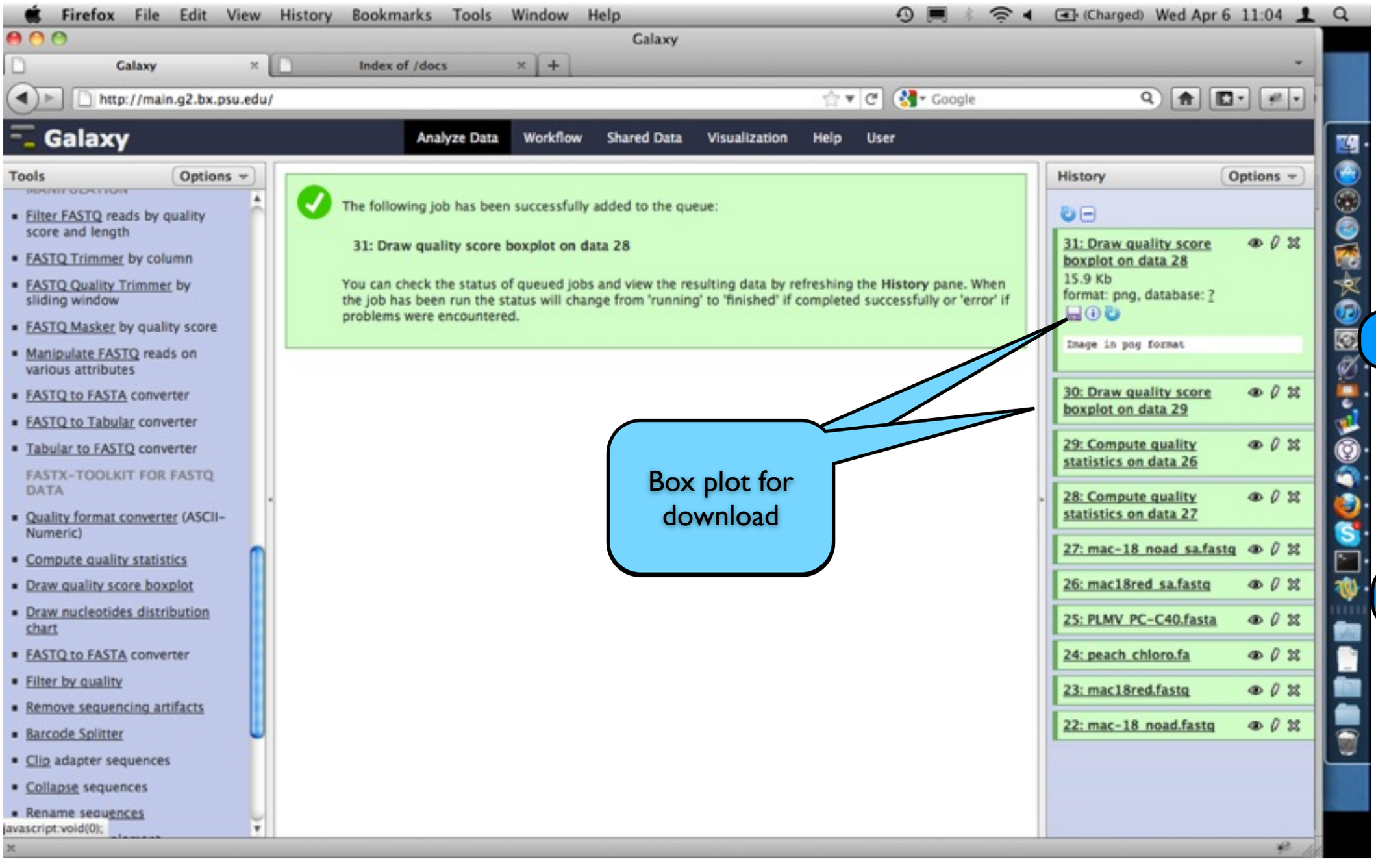
Data Quality control

Data Mapping

Data Visualization

GALAXY

Data quality control



Data Upload
★ sequencing reads
★ reference

Data preparation

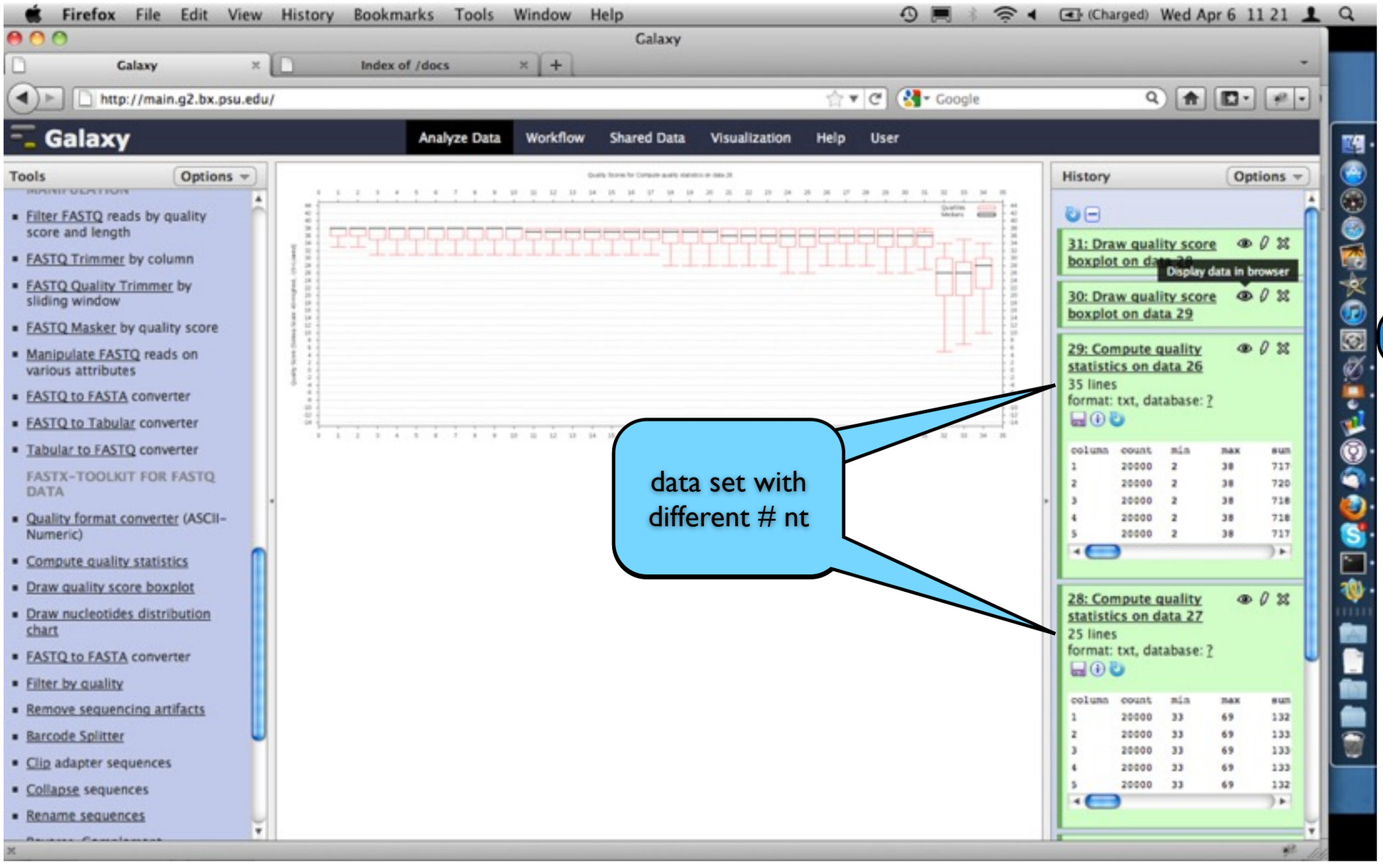
Data Quality control

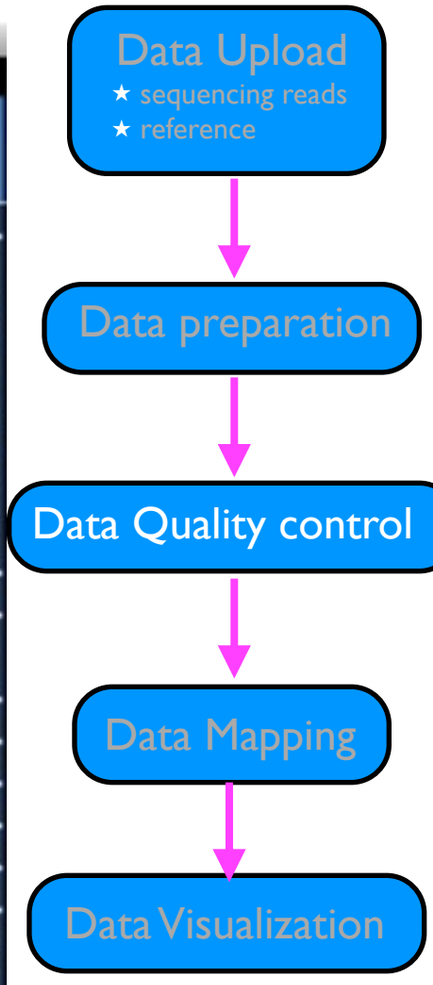
Data Mapping

Data Visualization

GALAXY

Data quality control





GALAXY

Data mapping

NGS Mapping

Use our own reference

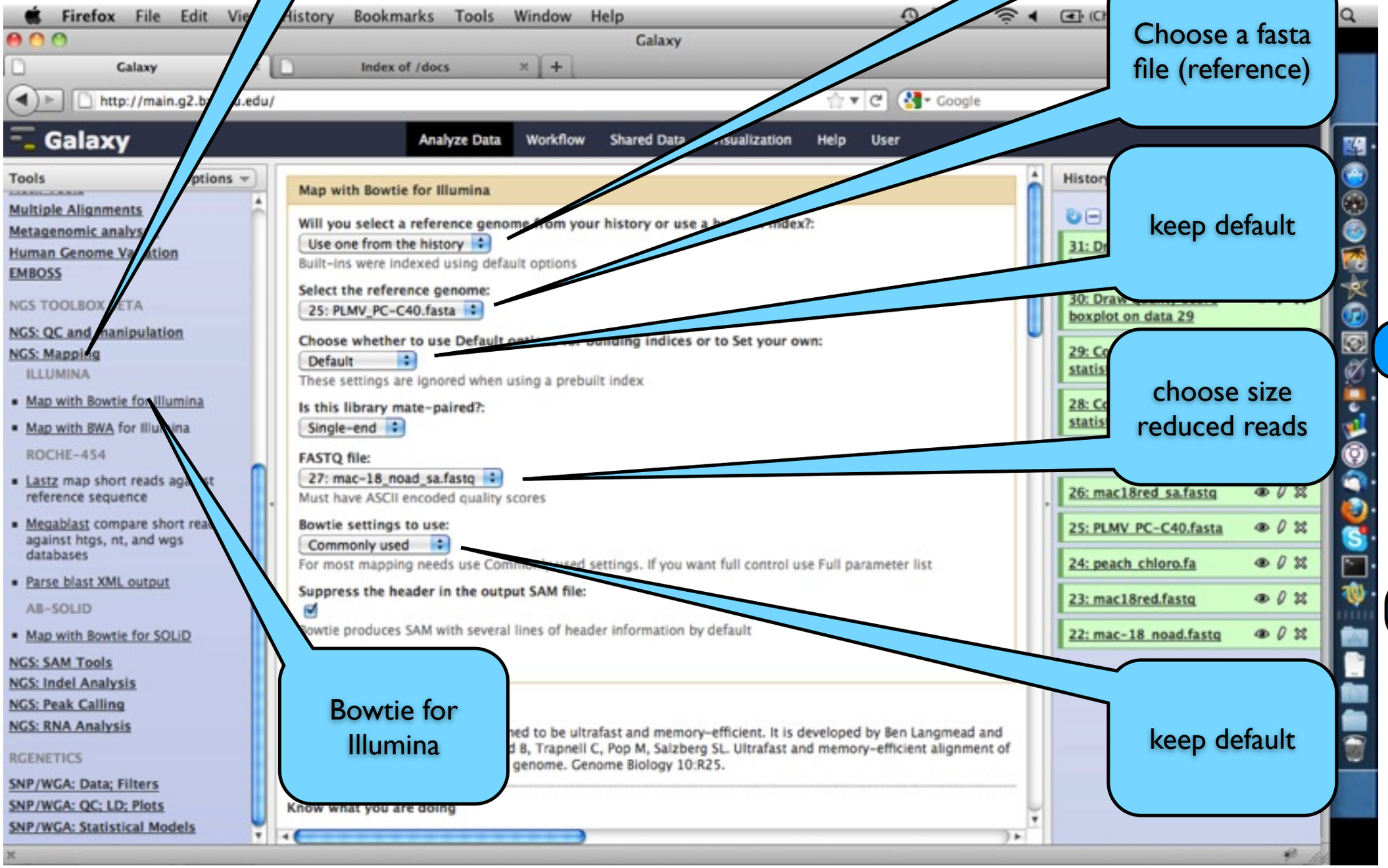
Choose a fasta file (reference)

keep default

choose size reduced reads

Bowtie for Illumina

keep default



Data Upload
★ sequencing reads
★ reference

Data preparation

Data Quality control

Data Mapping

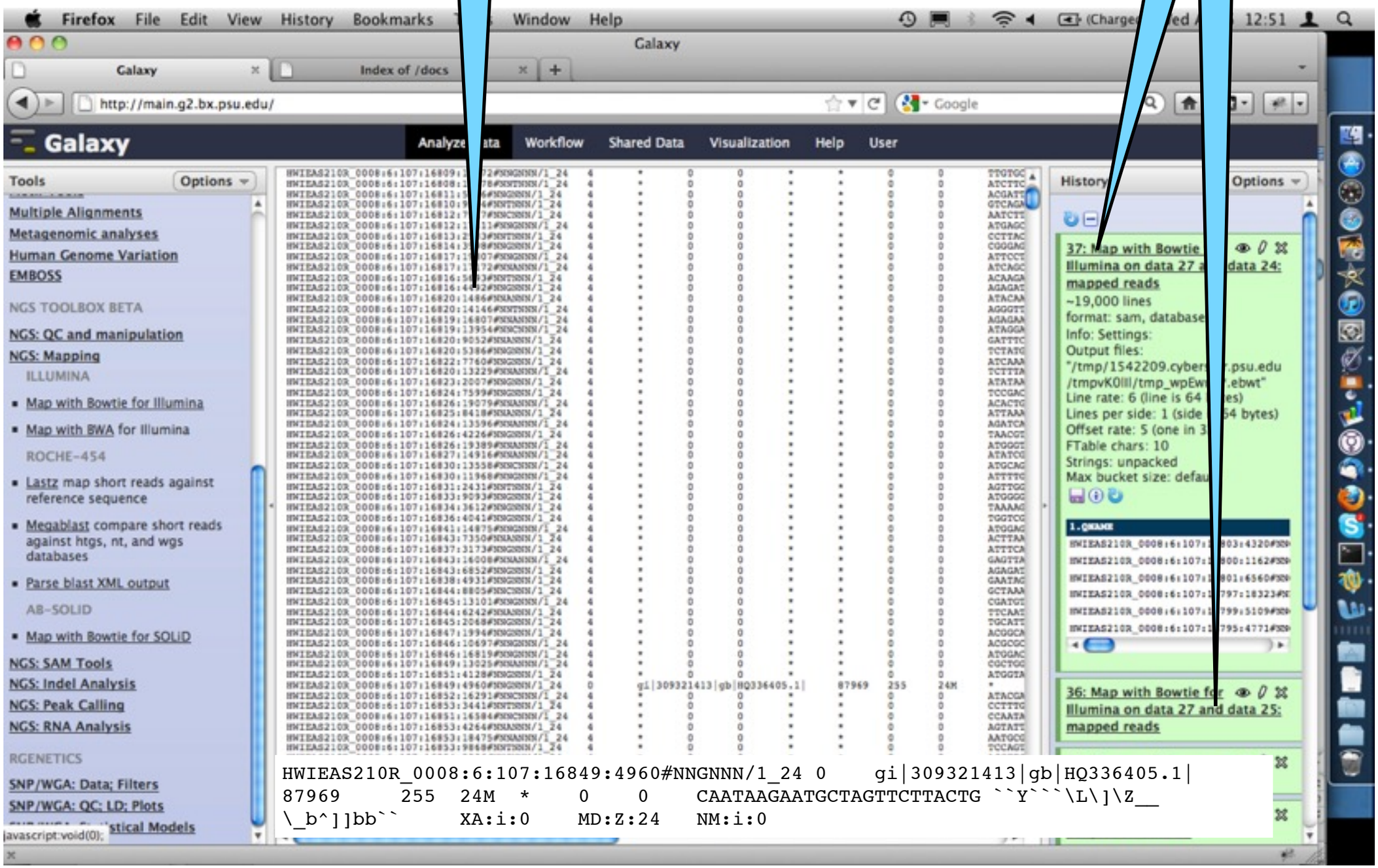
Data Visualization

GALAXY

Data mapping

Mapping file in SAM format

NGS Mapping



Data Upload
★ sequencing reads
★ reference

Data preparation

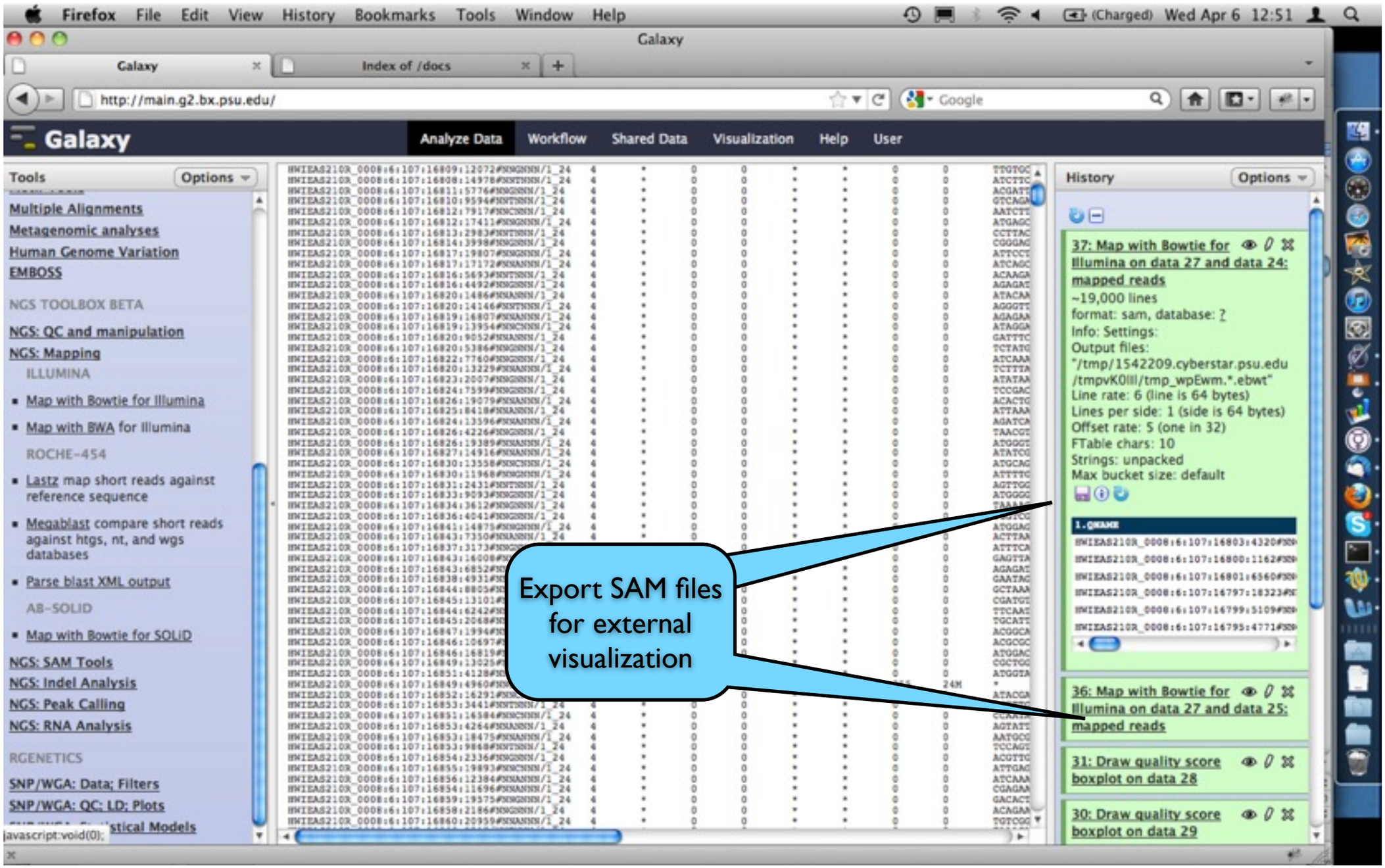
Data Quality control

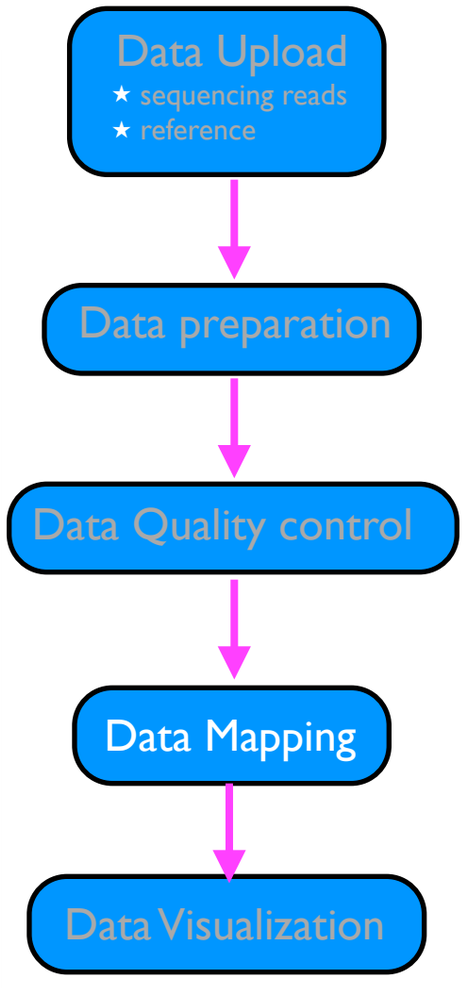
Data Mapping

Data Visualization

GALAXY

Data mapping

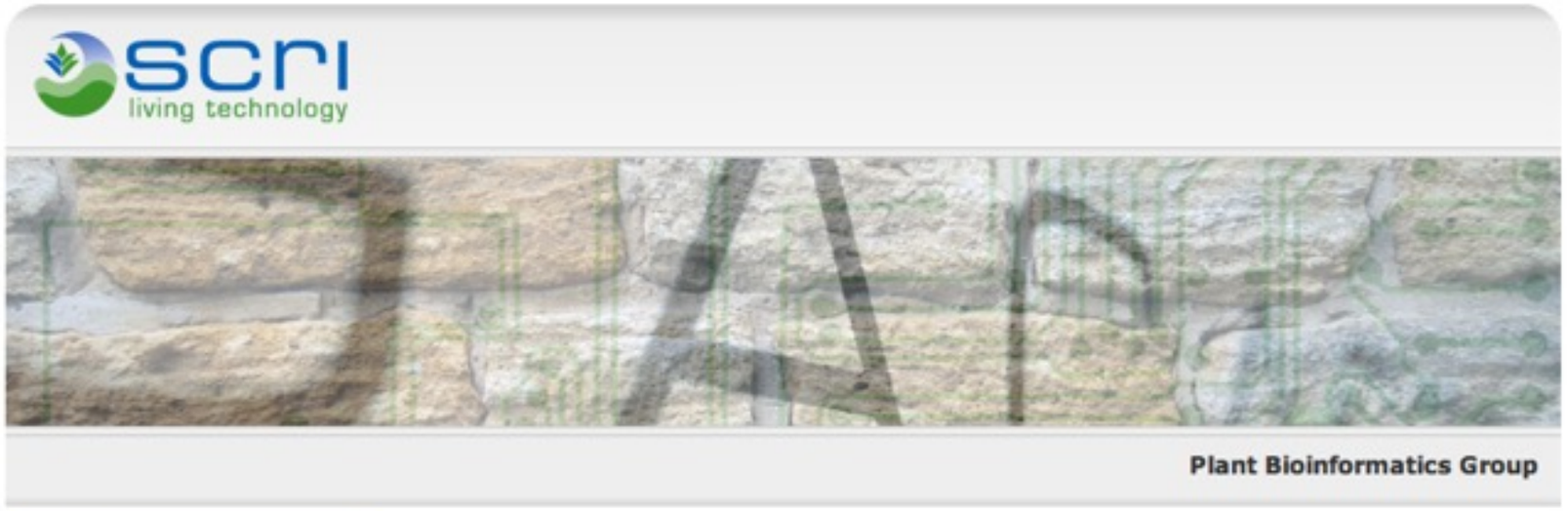




NOT GALAXY

Data visualization

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



Tablet

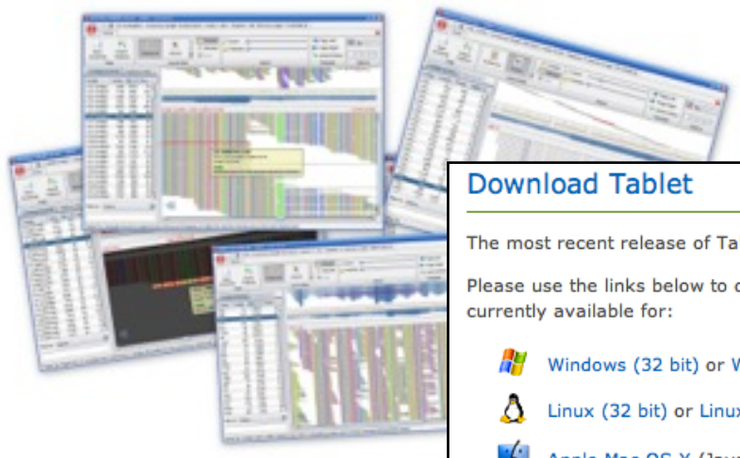
- Tablet Homepage
- Download Tablet
- Screenshots
- Tablet FAQ
- Sample Data
- Assembly Conversion
- Papers and Presentations
- Privacy Policy
- Tablet World Map
- Online Help

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.

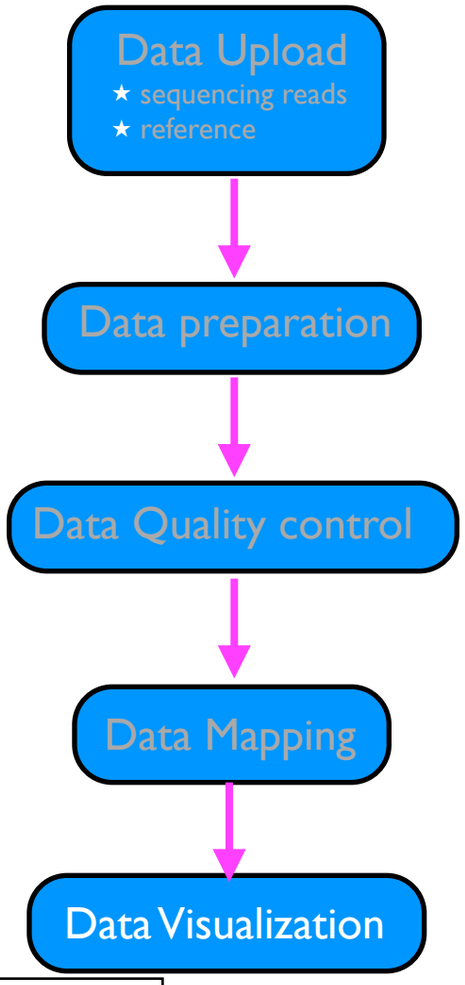


Download Tablet

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

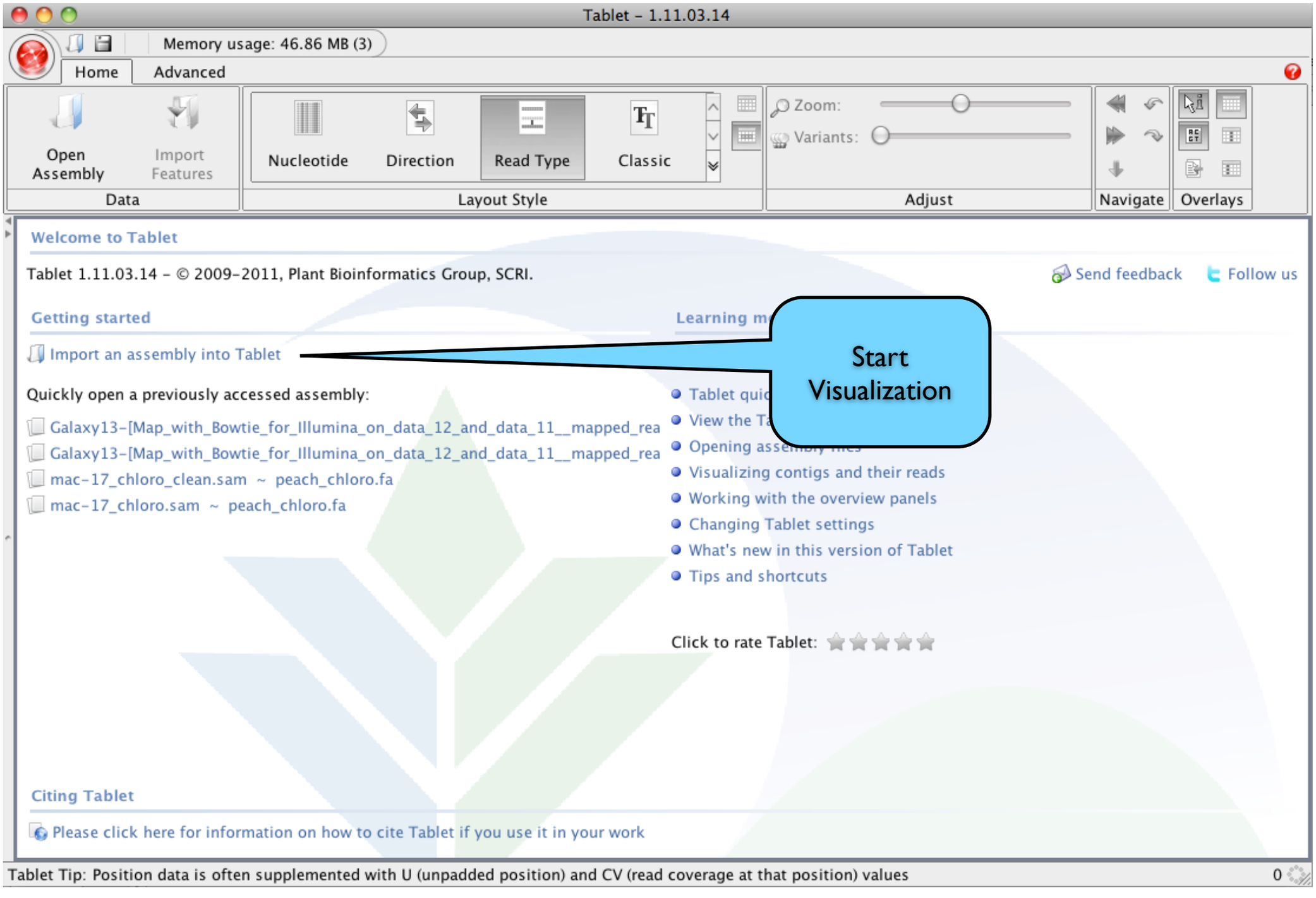
Please use the links below to download the Tablet installer most suitable for your operating system. Tablet is currently available for:

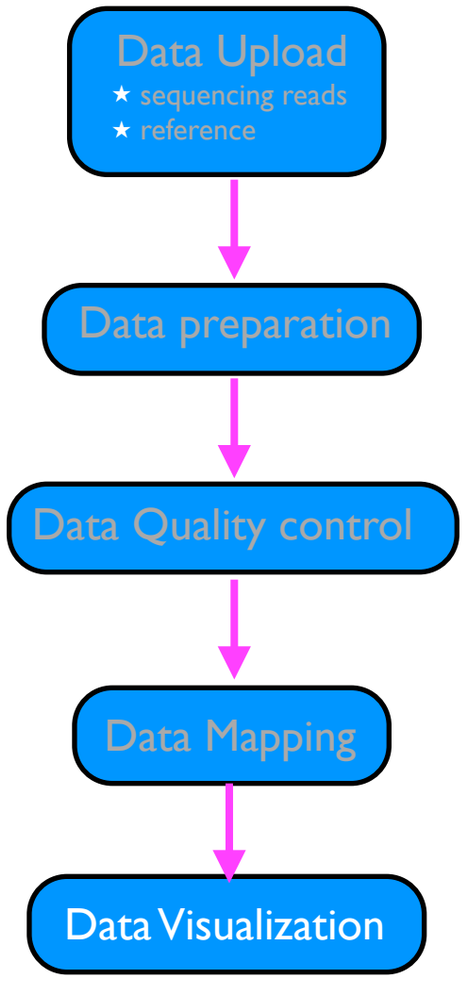
- Windows (32 bit) or Windows (64 bit)
- Linux (32 bit) or Linux (64 bit)
- Apple Mac OS X (Java 6 required)
- Solaris (Sparc)



Tablet

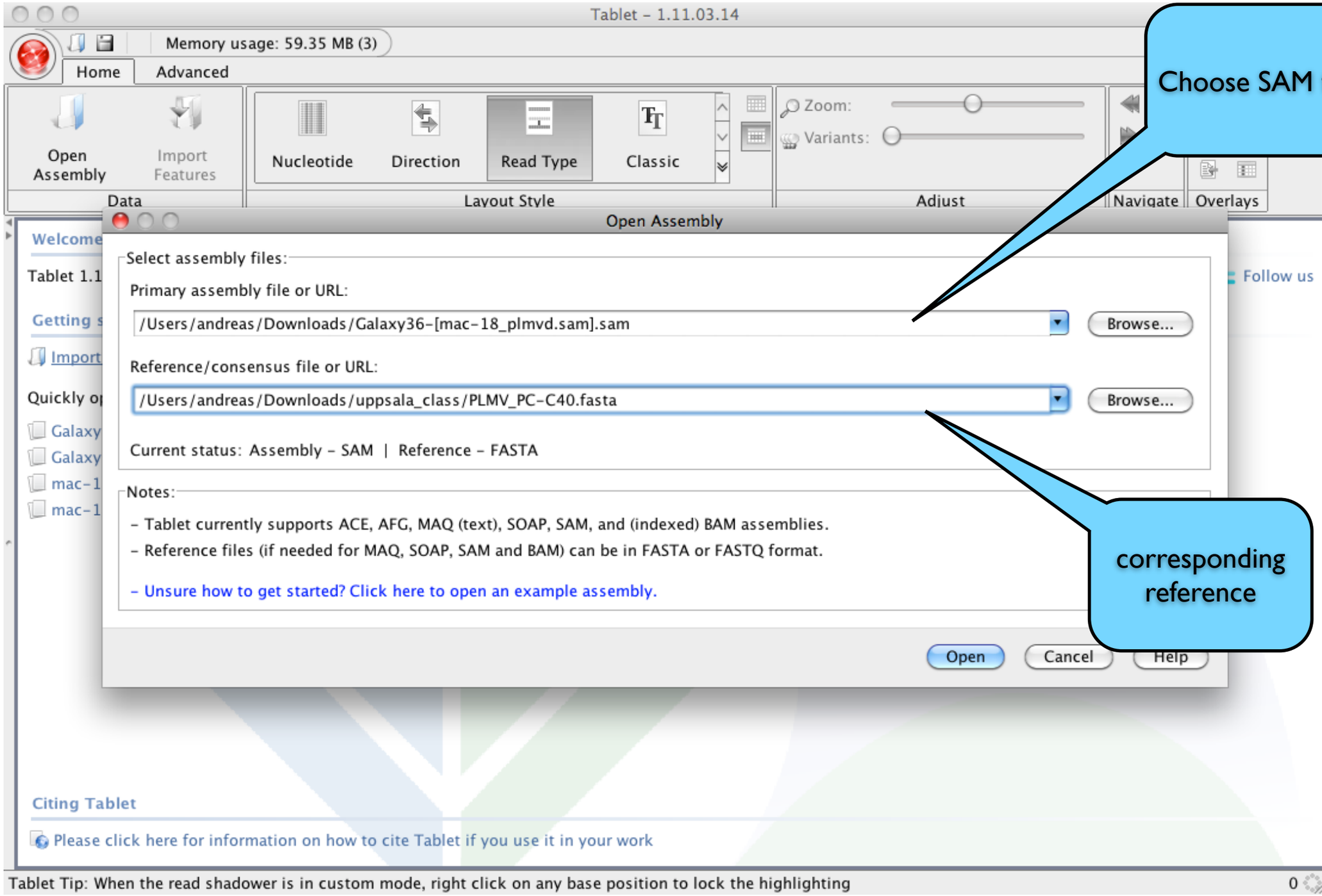
Data visualization





Tablet

Data visualization

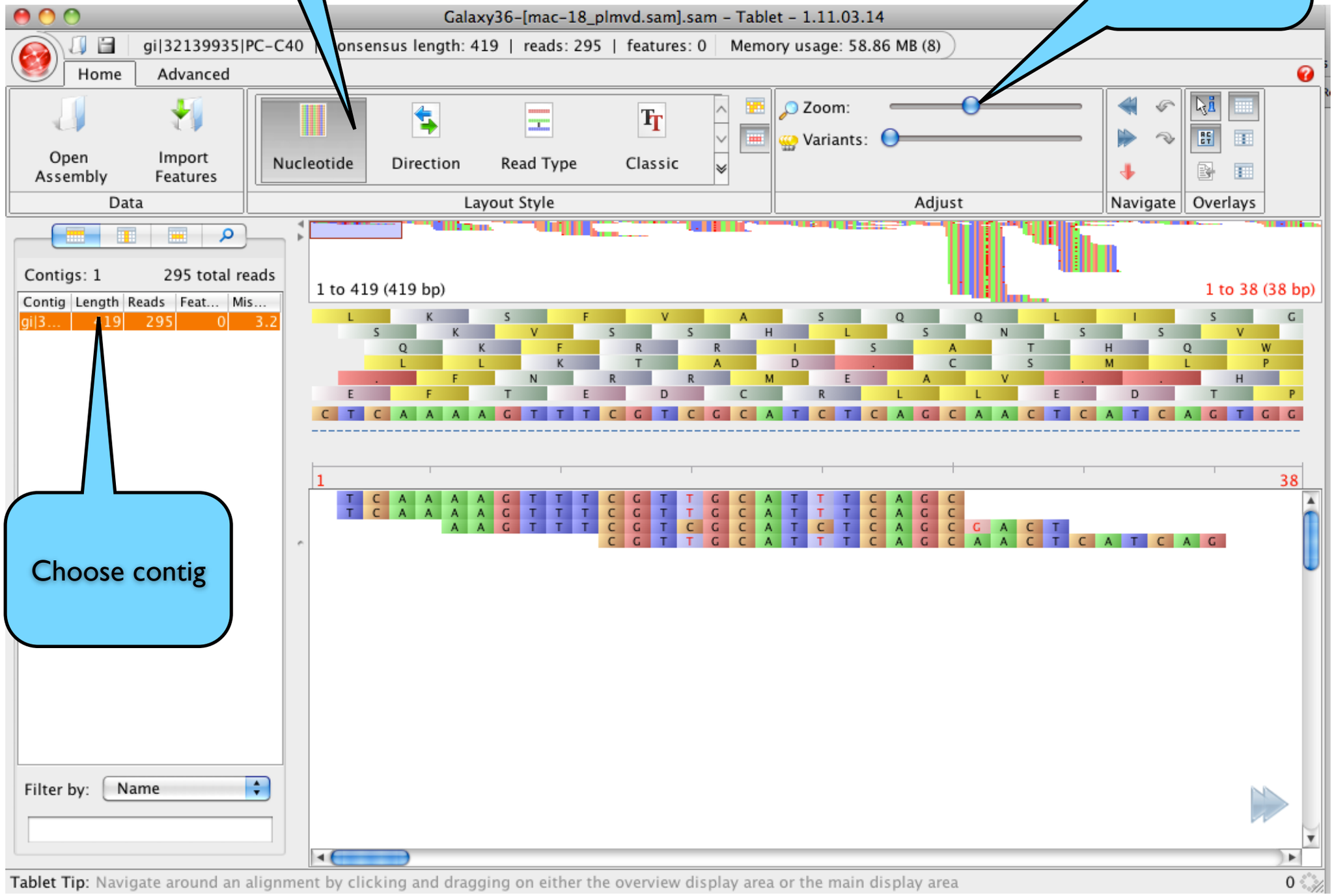


Tablet

Data visualization
Viroid

Layout

zoom



Choose contig

Data Upload
★ sequencing reads
★ reference

Data preparation

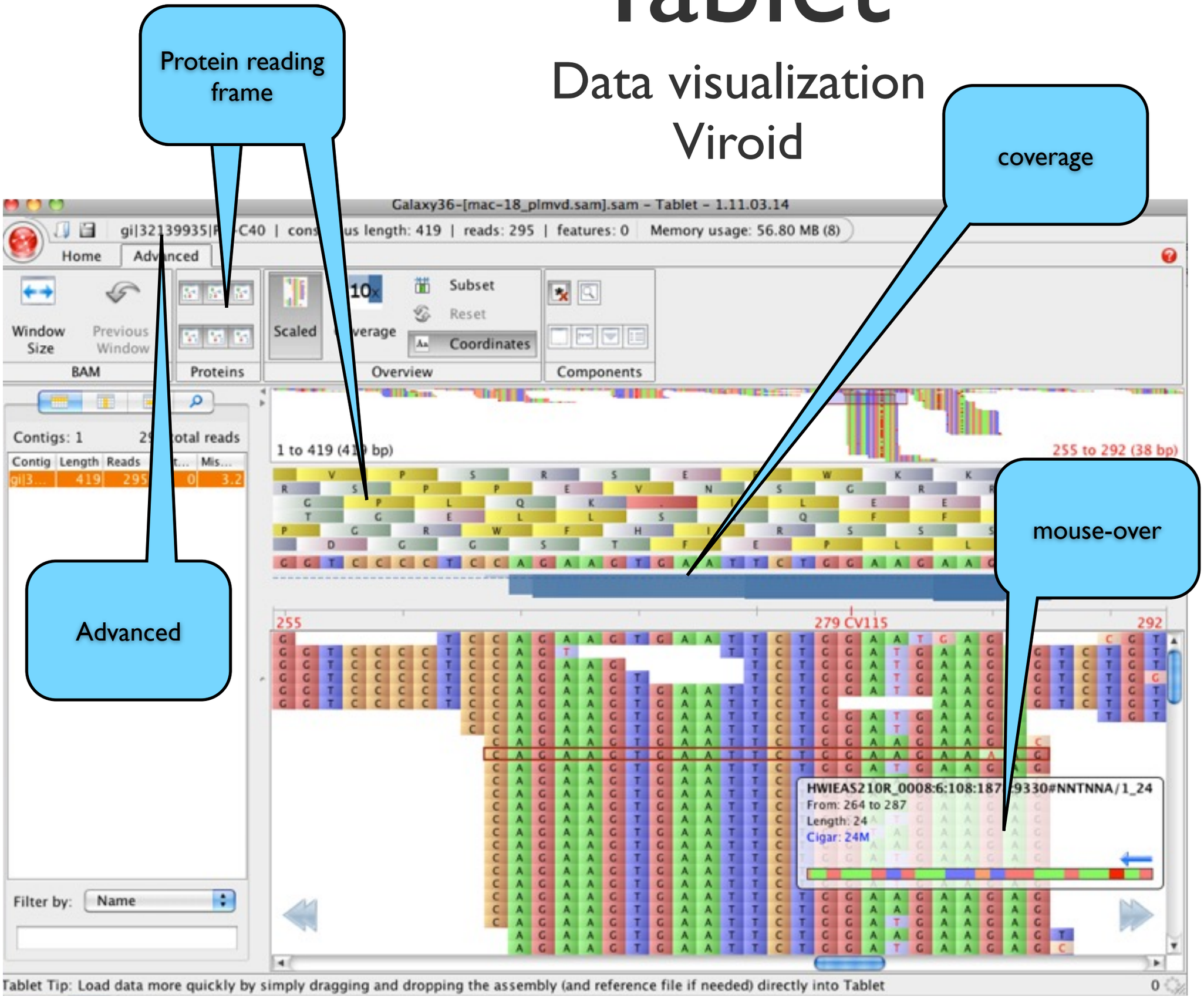
Data Quality control

Data Mapping

Data Visualization

Tablet

Data visualization
Viroid



Data Upload
★ sequencing reads
★ reference

Data preparation

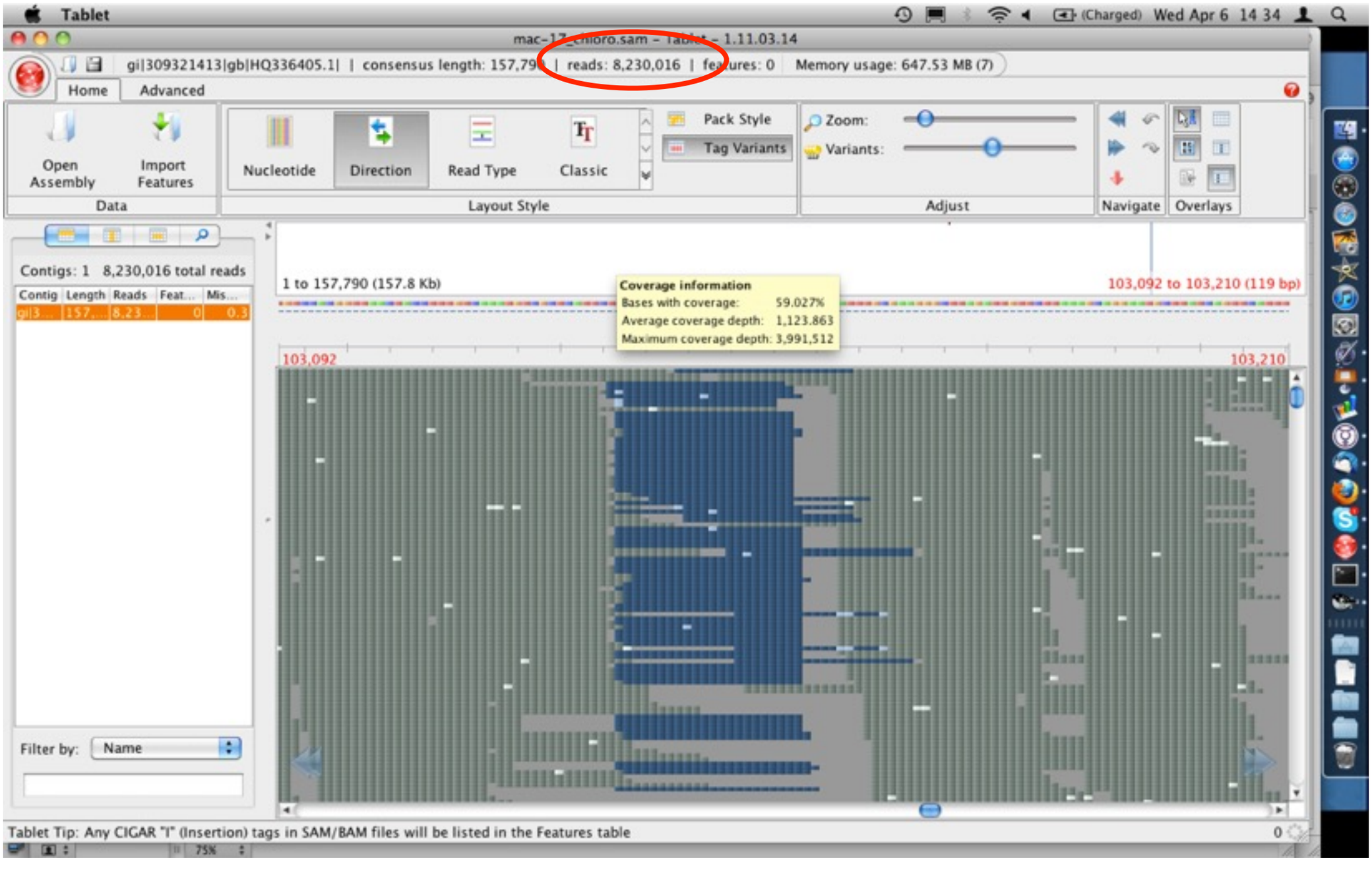
Data Quality control

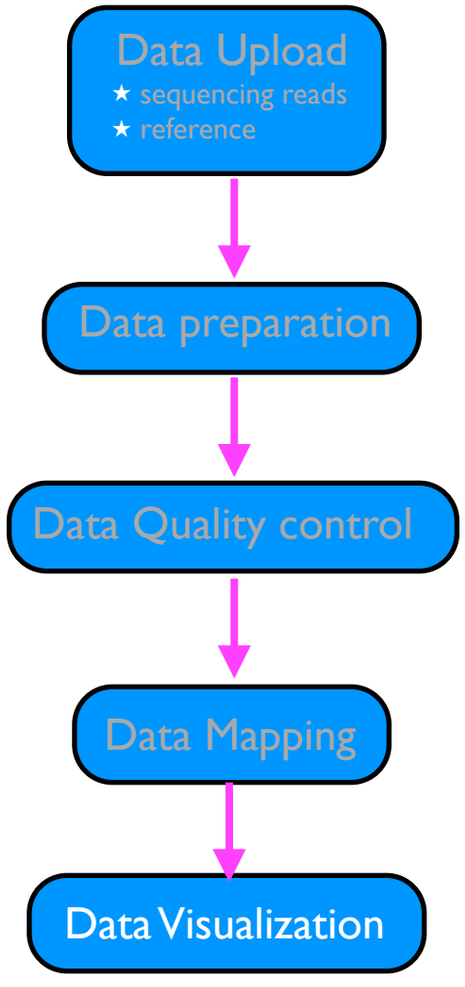
Data Mapping

Data Visualization

Tablet

Data visualization
Peach tree chloroplast





GALAXY/Tablet

DONE !