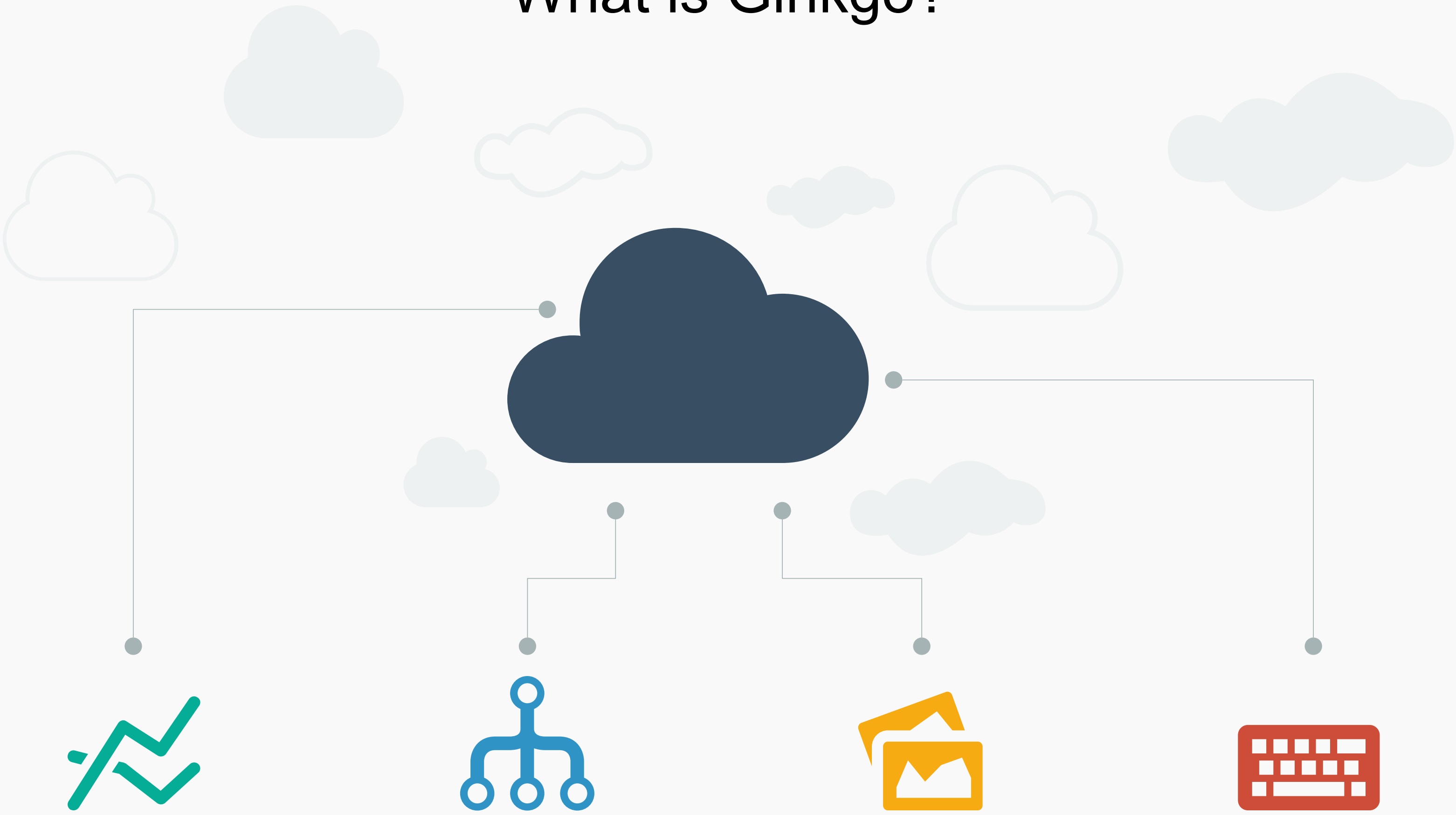
ginkgo

SINGLE-CELL COPY-NUMBER ANALYSIS

VIZBI 2015 WORKSHOP

Robert Aboukhalil
raboukha@cshl.edu

What is Ginkgo?



Single-cell CNV analysis

Phylogenetic analysis

Data visualization tools

Share with collaborators

Outline

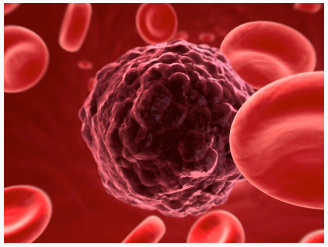
Experimental approaches

Computational approaches for CNV analysis

Demo

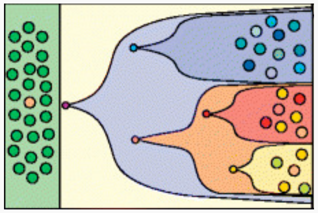
Hands-on

Single-cell sequencing

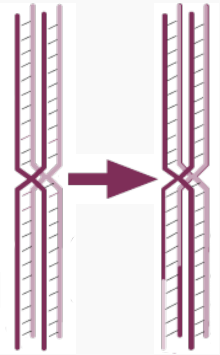


Circulating tumor cells

Neuronal mosaicism

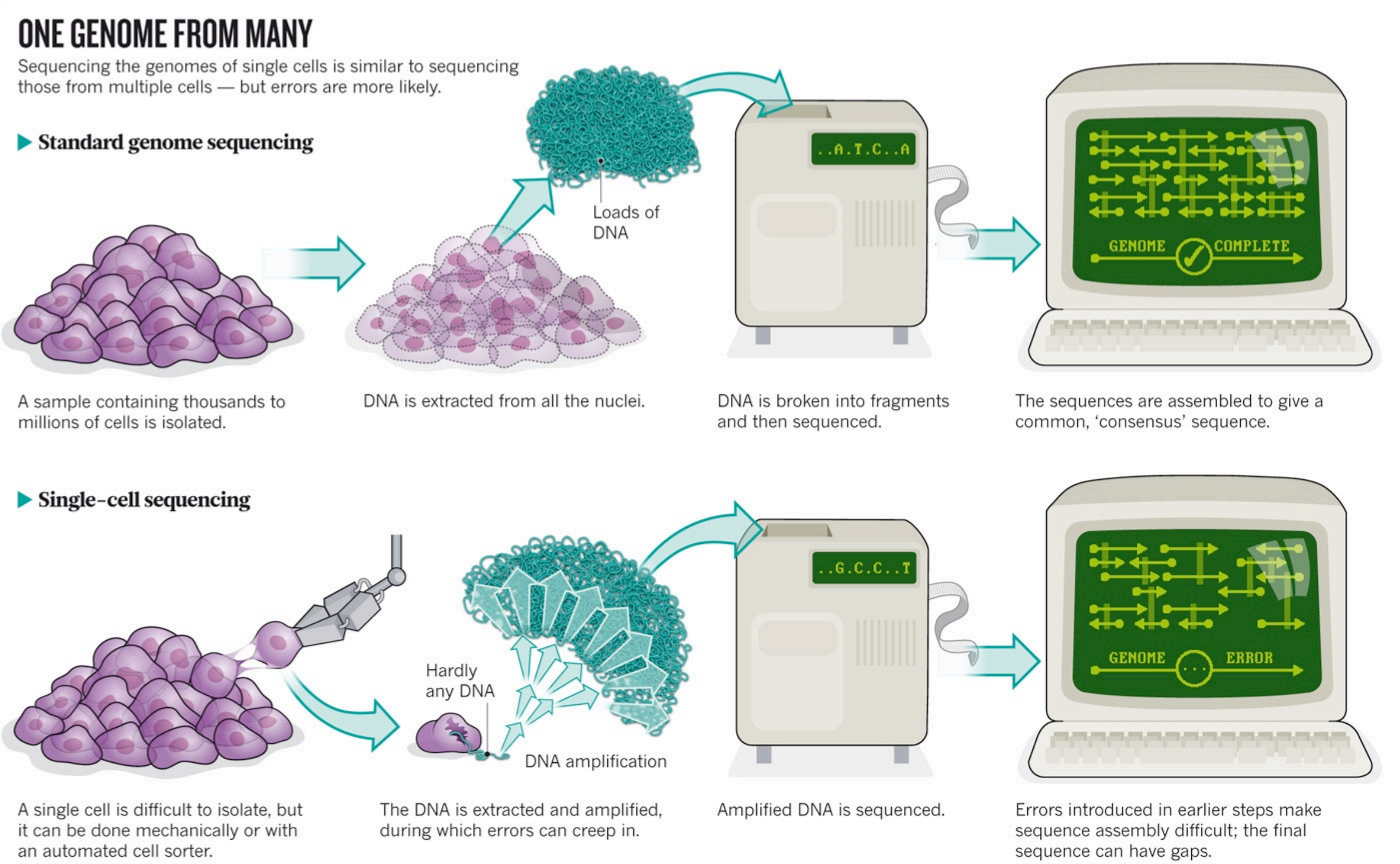


Clonal evolution in tumors



Recombination/ crossover in germ cells

Single-cell vs. bulk sequencing



ONE GENOME FROM MANY

Sequencing the genomes of single cells is similar to sequencing those from multiple cells — but errors are more likely.

▶ Standard genome sequencing

A sample containing thousands to millions of cells is isolated.

DNA is extracted from all the nuclei.

DNA is broken into fragments and then sequenced.

The sequences are assembled to give a common, 'consensus' sequence.

▶ Single-cell sequencing

A single cell is difficult to isolate, but it can be done mechanically or with an automated cell sorter.

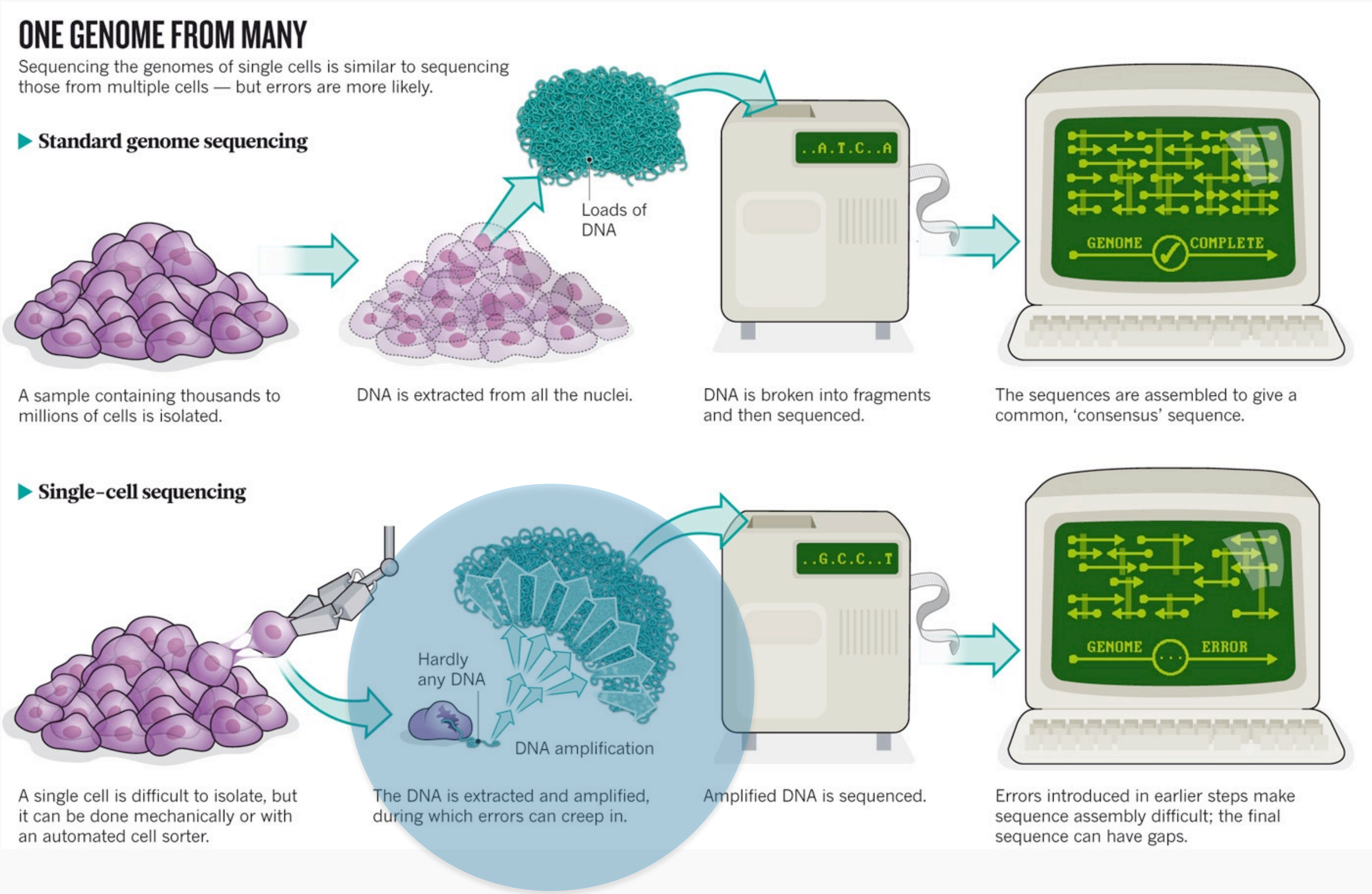
The DNA is extracted and amplified, during which errors can creep in.

Amplified DNA is sequenced.

Errors introduced in earlier steps make sequence assembly difficult; the final sequence can have gaps.

Brian Owens, Nature News 2012

Single-cell vs. bulk sequencing



ONE GENOME FROM MANY

Sequencing the genomes of single cells is similar to sequencing those from multiple cells — but errors are more likely.

Standard genome sequencing

A sample containing thousands to millions of cells is isolated.

DNA is extracted from all the nuclei.

DNA is broken into fragments and then sequenced.

The sequences are assembled to give a common, 'consensus' sequence.

Single-cell sequencing

A single cell is difficult to isolate, but it can be done mechanically or with an automated cell sorter.

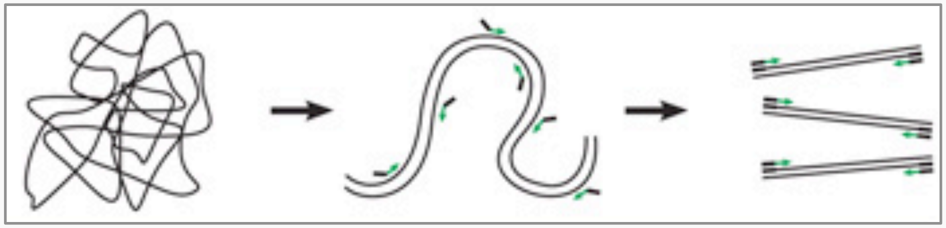
The DNA is extracted and amplified, during which errors can creep in.

Amplified DNA is sequenced.

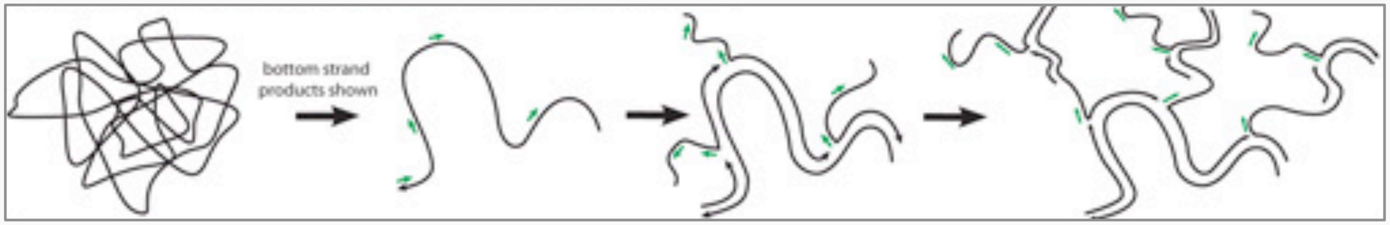
Errors introduced in earlier steps make sequence assembly difficult; the final sequence can have gaps.

Brian Owens, Nature News 2012

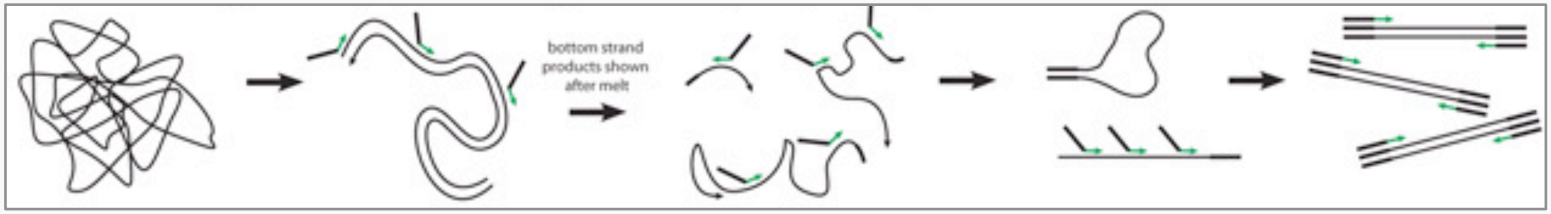
Whole Genome Amplification (WGA) methods



DOP-PCR (Degenerate Oligonucleotide Primed PCR)



MDA (Multiple Displacement Amplification)

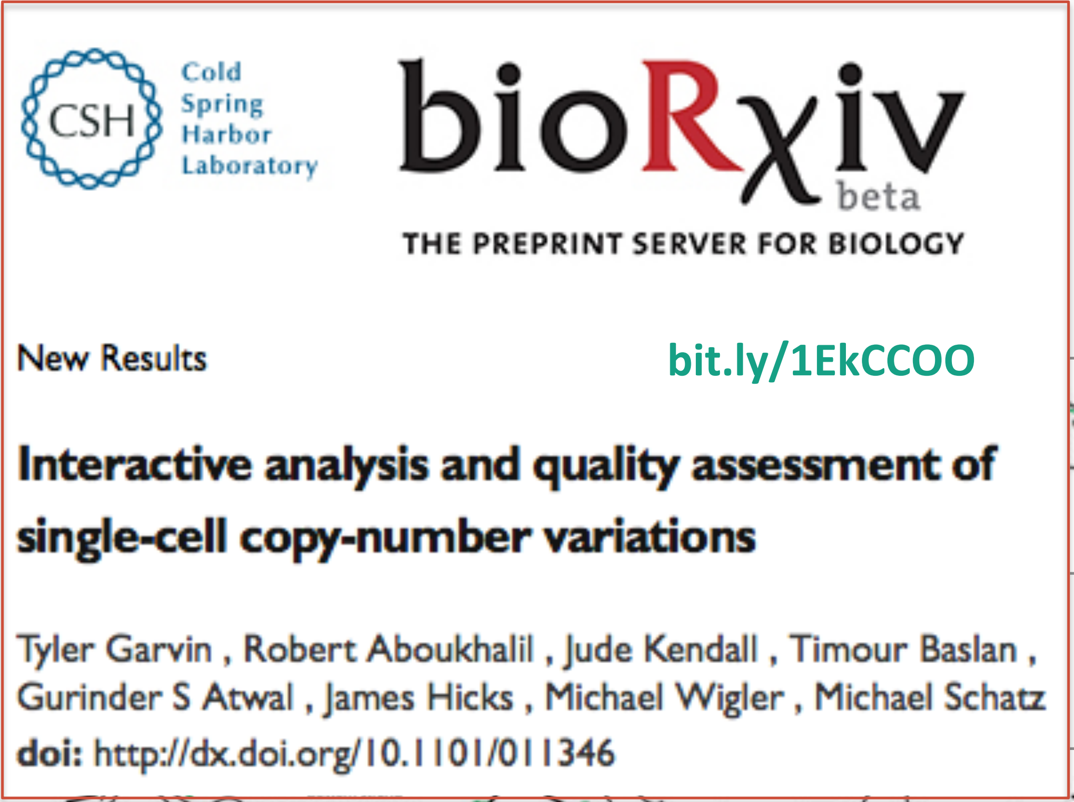


MALBAC (Multiple Annealing and Looping Based Amplification Cycles)

Paul Blainey, FEMS Microbiol Rev. 2013

Whole Genome Amplification (WGA) methods

DOP-PCR (Deg



Cold Spring Harbor Laboratory

bioRxiv beta

THE PREPRINT SERVER FOR BIOLOGY

New Results

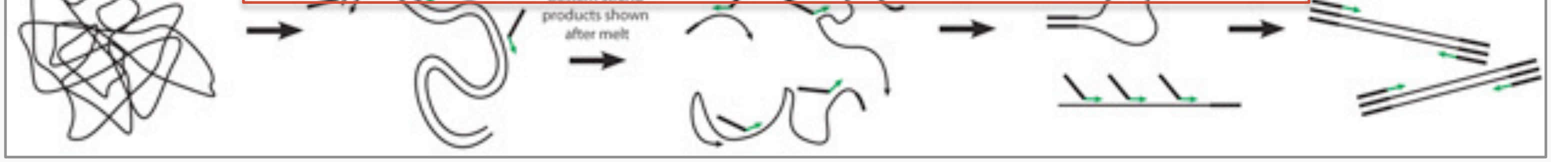
bit.ly/1EkCCOO

Interactive analysis and quality assessment of single-cell copy-number variations

Tyler Garvin , Robert Aboukhalil , Jude Kendall , Timour Baslan , Gurinder S Atwal , James Hicks , Michael Wigler , Michael Schatz

doi: http://dx.doi.org/10.1101/011346

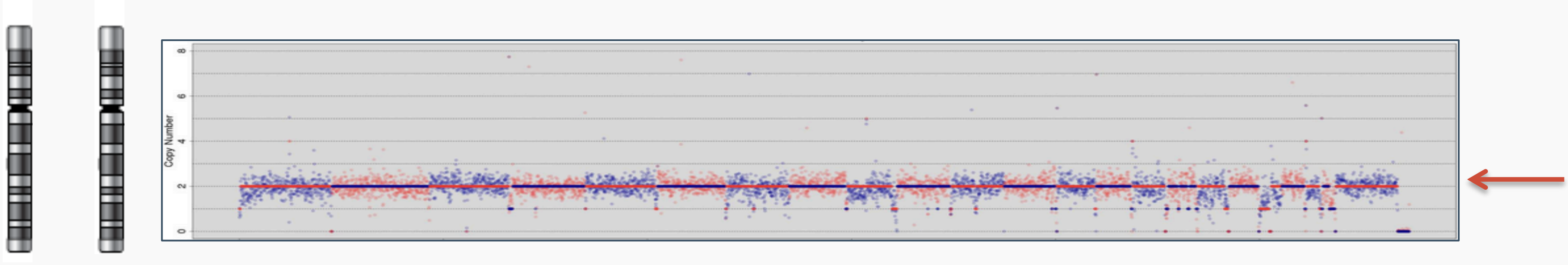
MDA (Multiple I

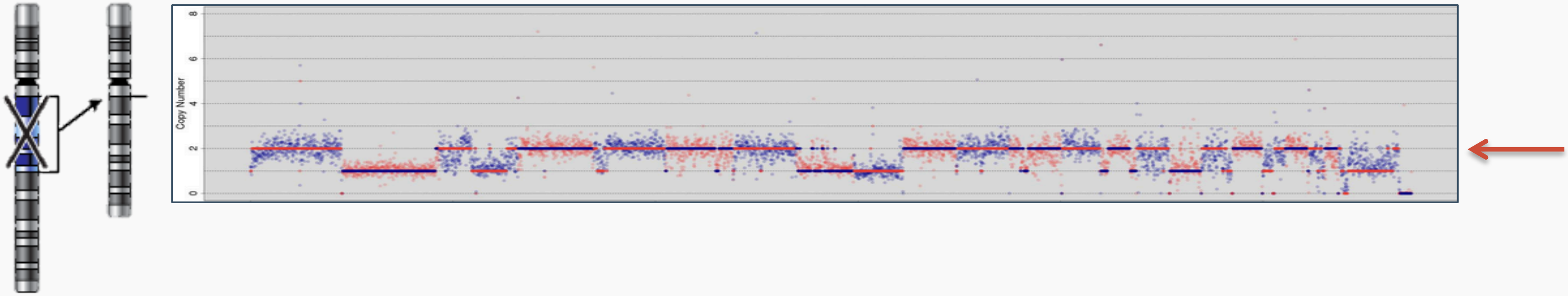


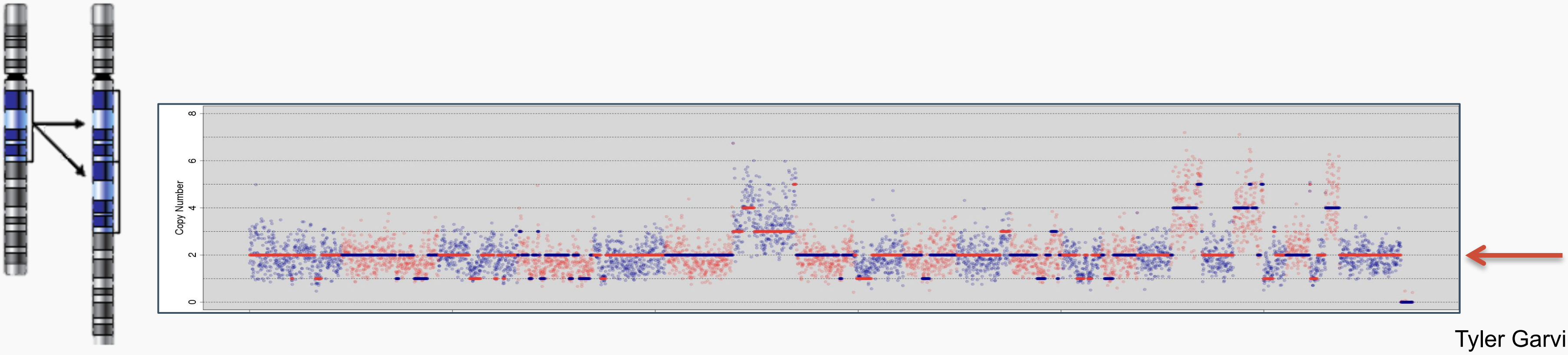
MALBAC (Multiple Annealing and Looping Based Amplification Cycles)

Paul Blainey, FEMS Microbiol Rev. 2013

Copy-number variant analysis

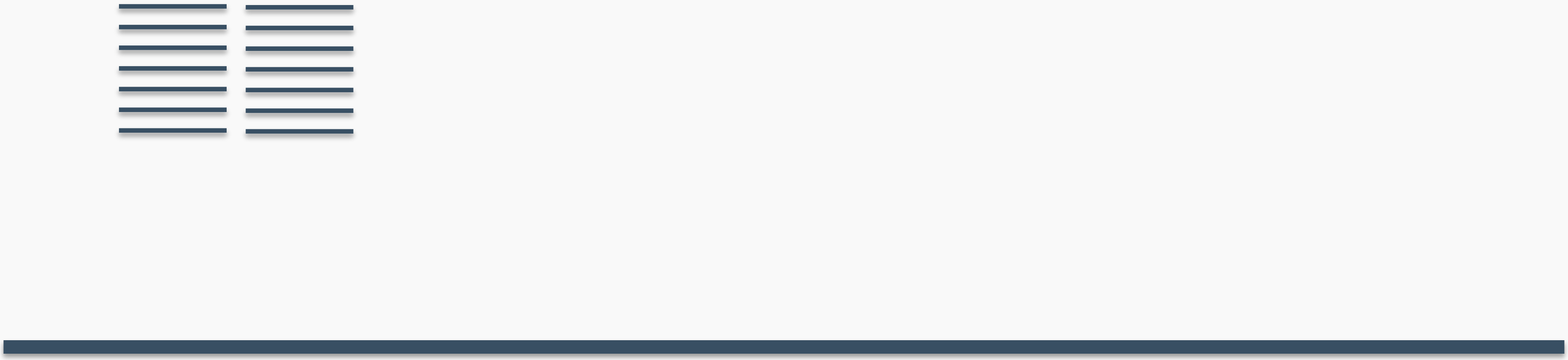






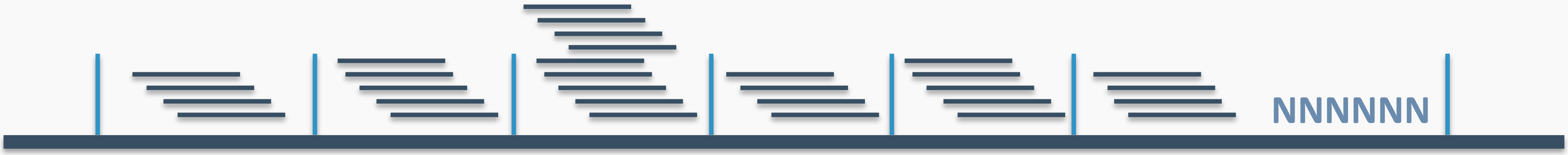
Tyler Garvin

Copy-number variant analysis



Low coverage allows us to study copy-number variants

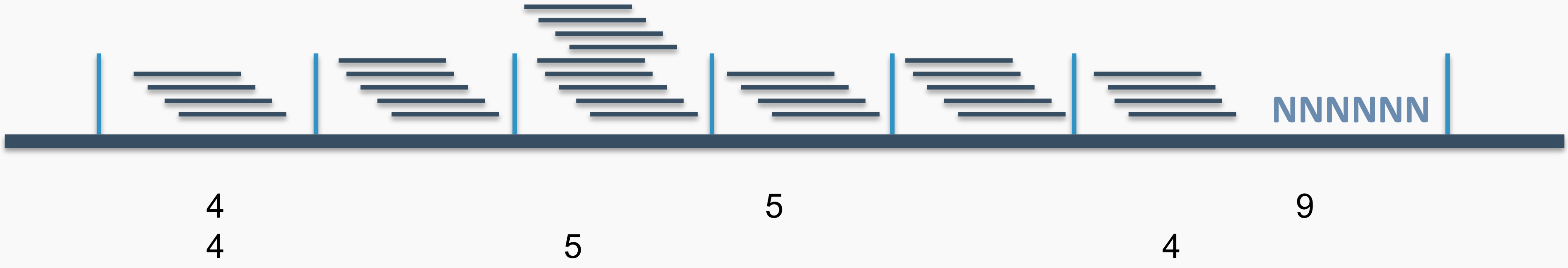
Copy-number variant analysis



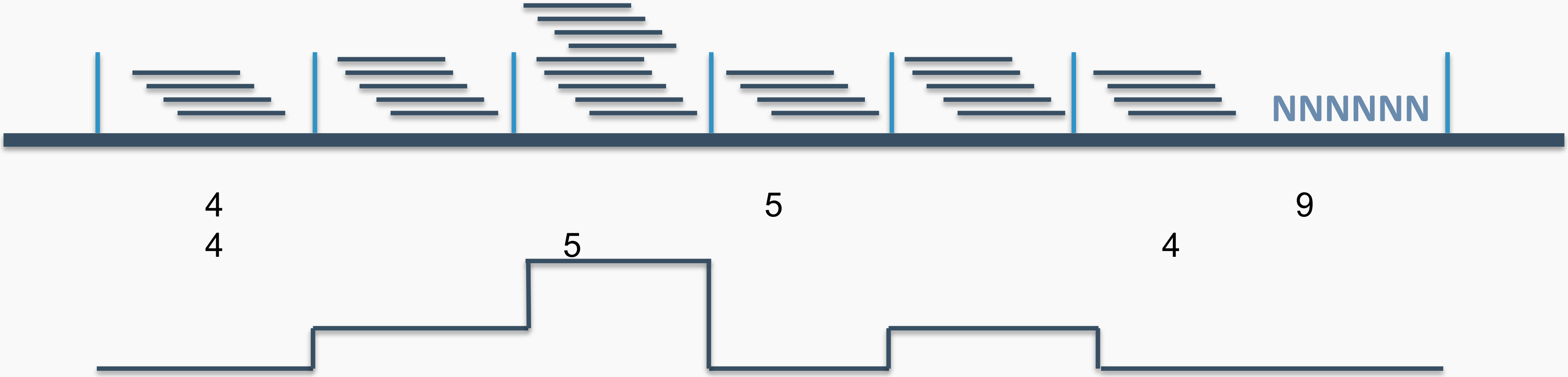
Divide genome into “bins” with ~50 – 100 reads / bin

Baslan et al., Nature Protocols 2012

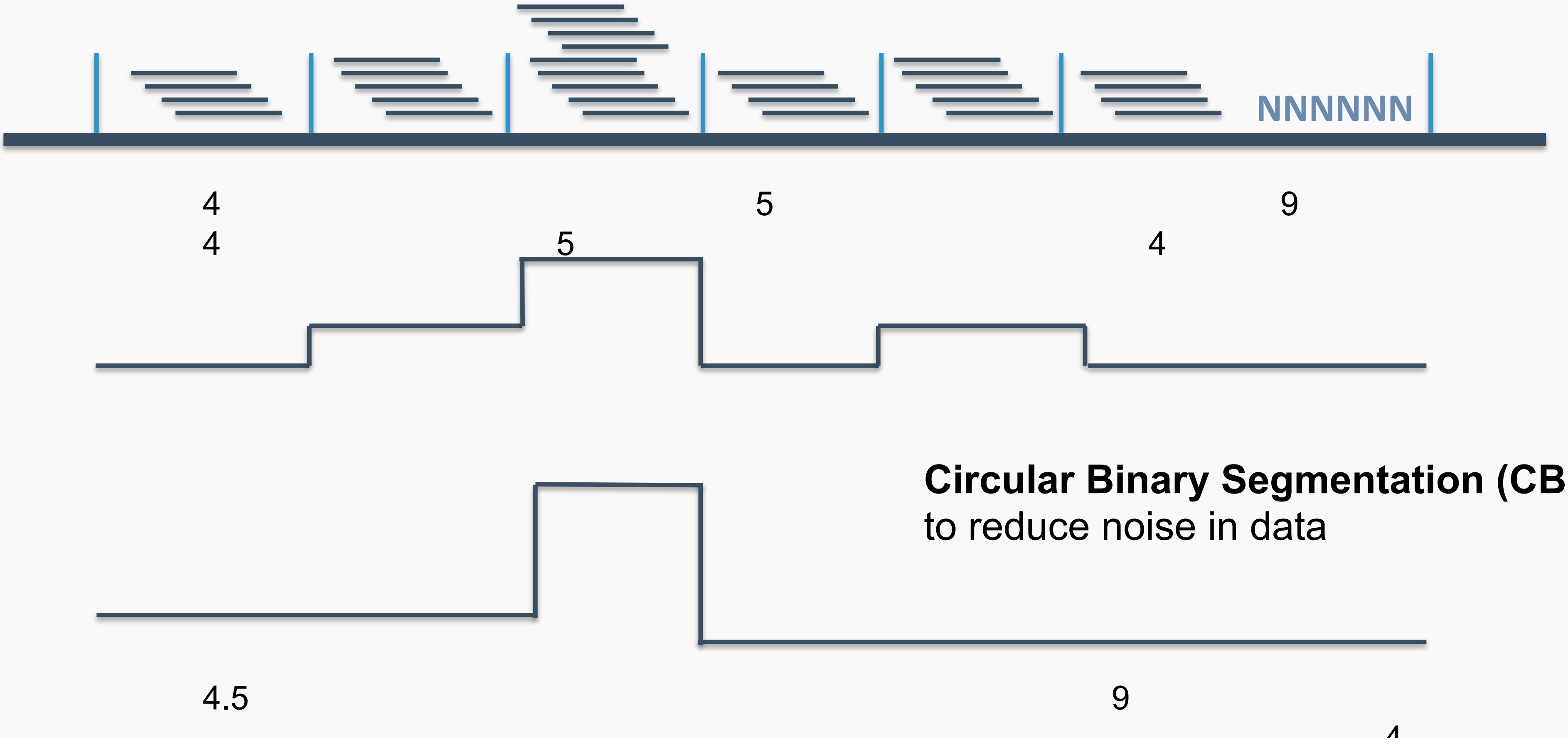
Copy-number variant analysis



Copy-number variant analysis



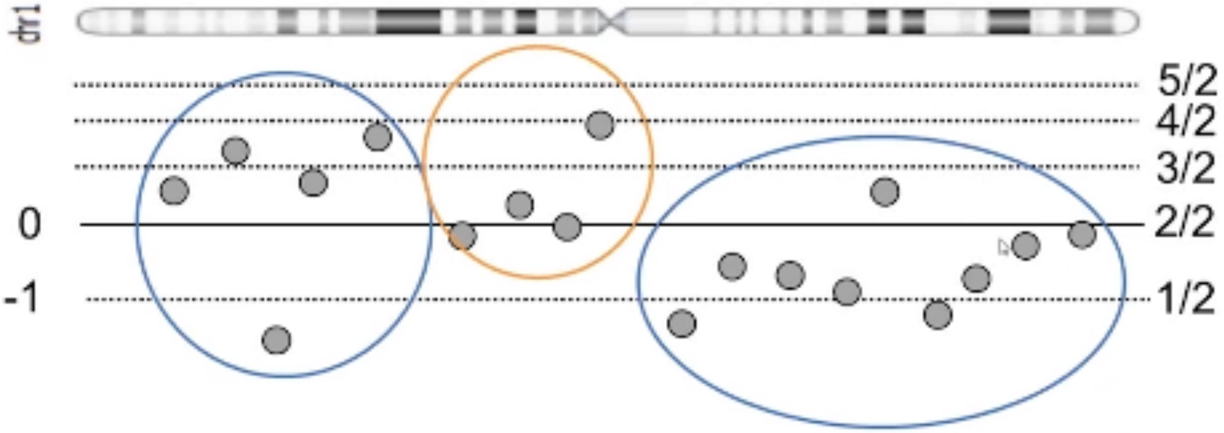
Copy-number variant analysis

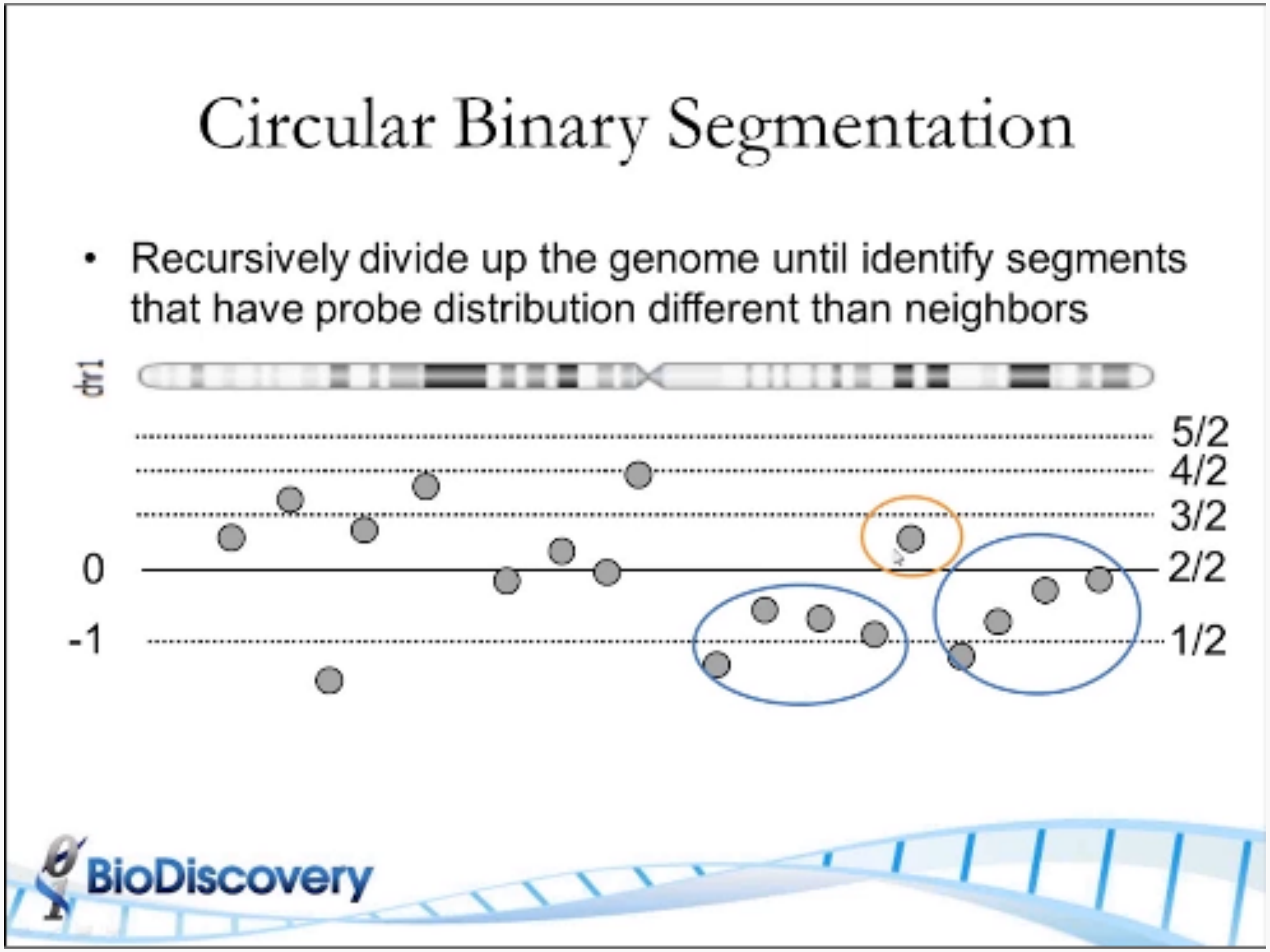


Circular Binary Segmentation (CBS
to reduce noise in data

Circular Binary Segmentation

- Recursively divide up the genome until identify segments that have probe distribution different than neighbors



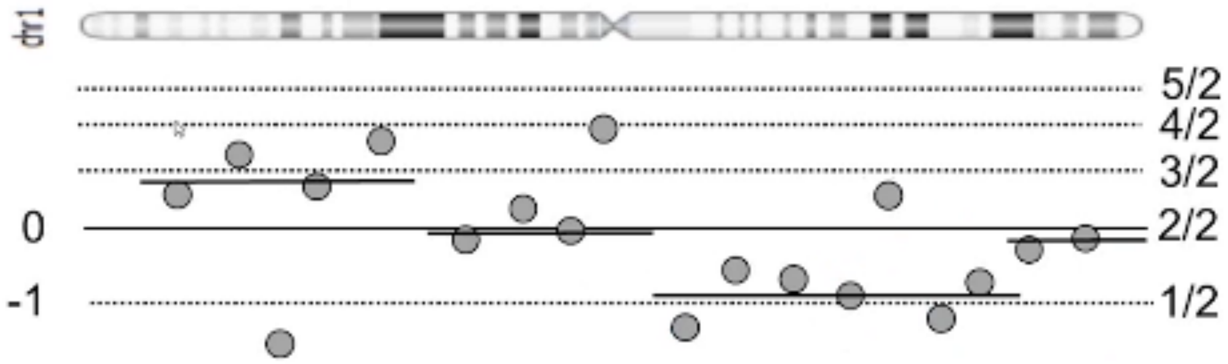


Circular Binary Segmentation

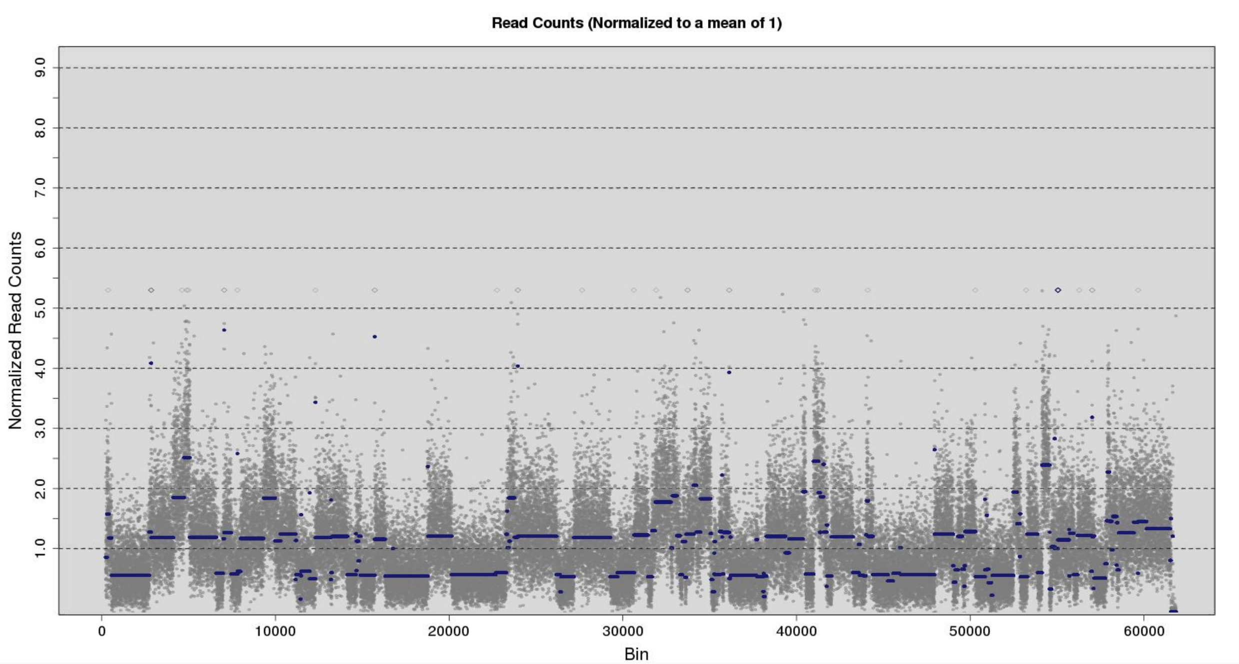
- Recursively divide up the genome until identify segments that have probe distribution different than neighbors

Circular Binary Segmentation

- Recursively divide up the genome until identify segments that have probe distribution different than neighbors

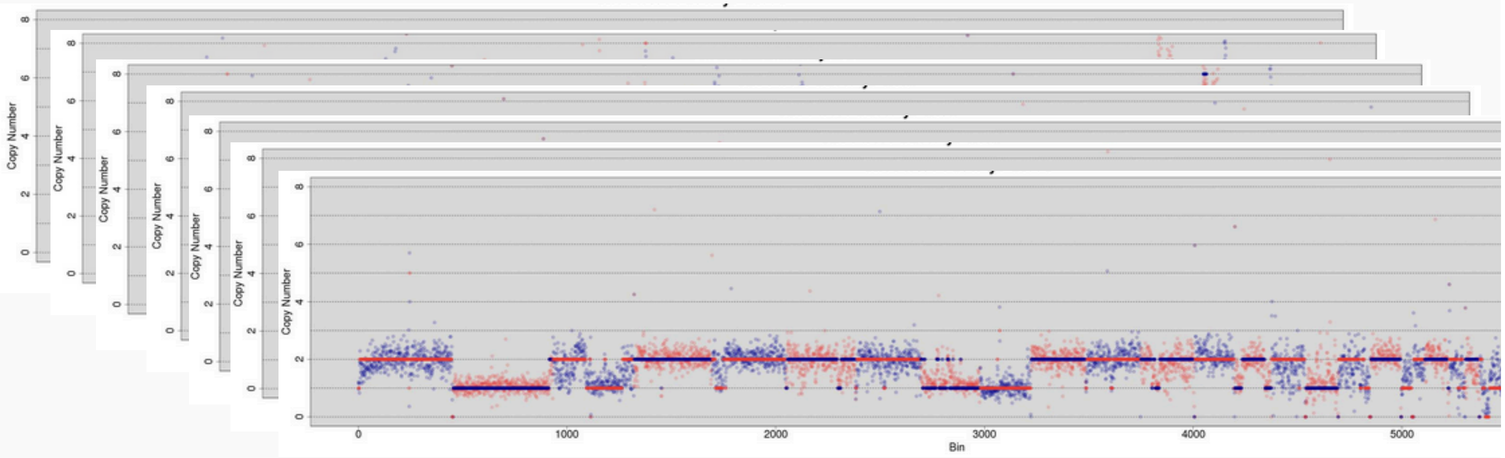


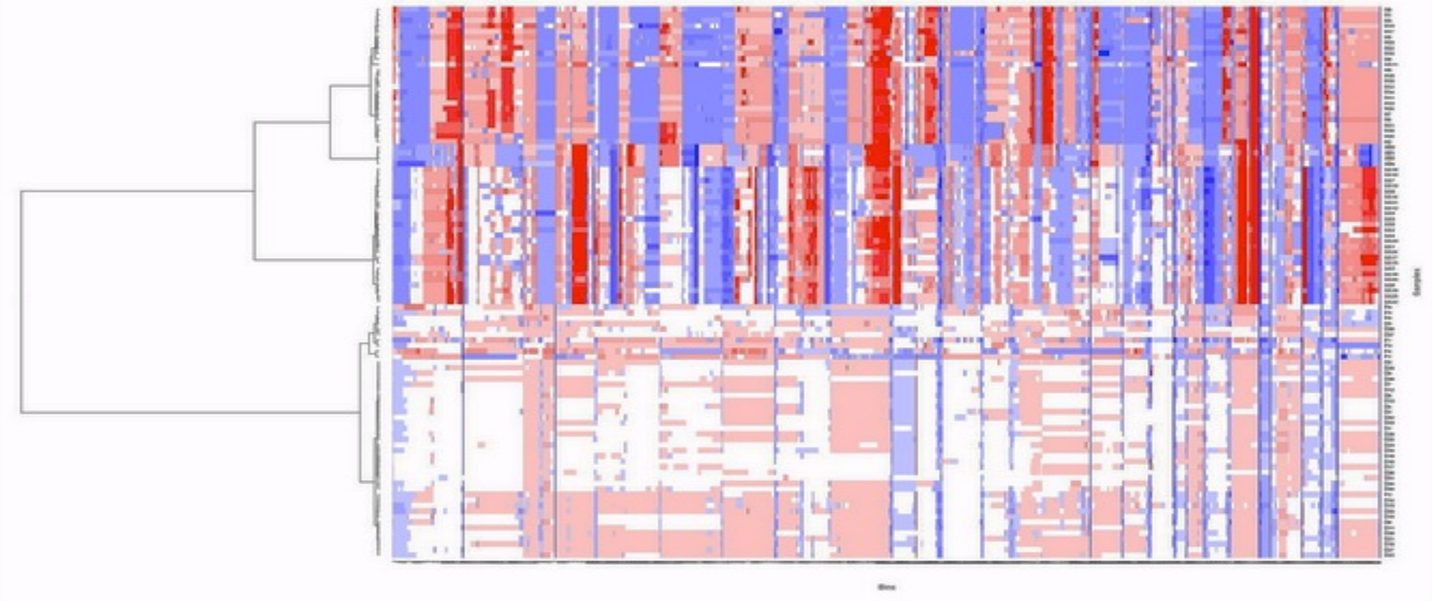
Copy-number variant analysis



Tyler Garvin

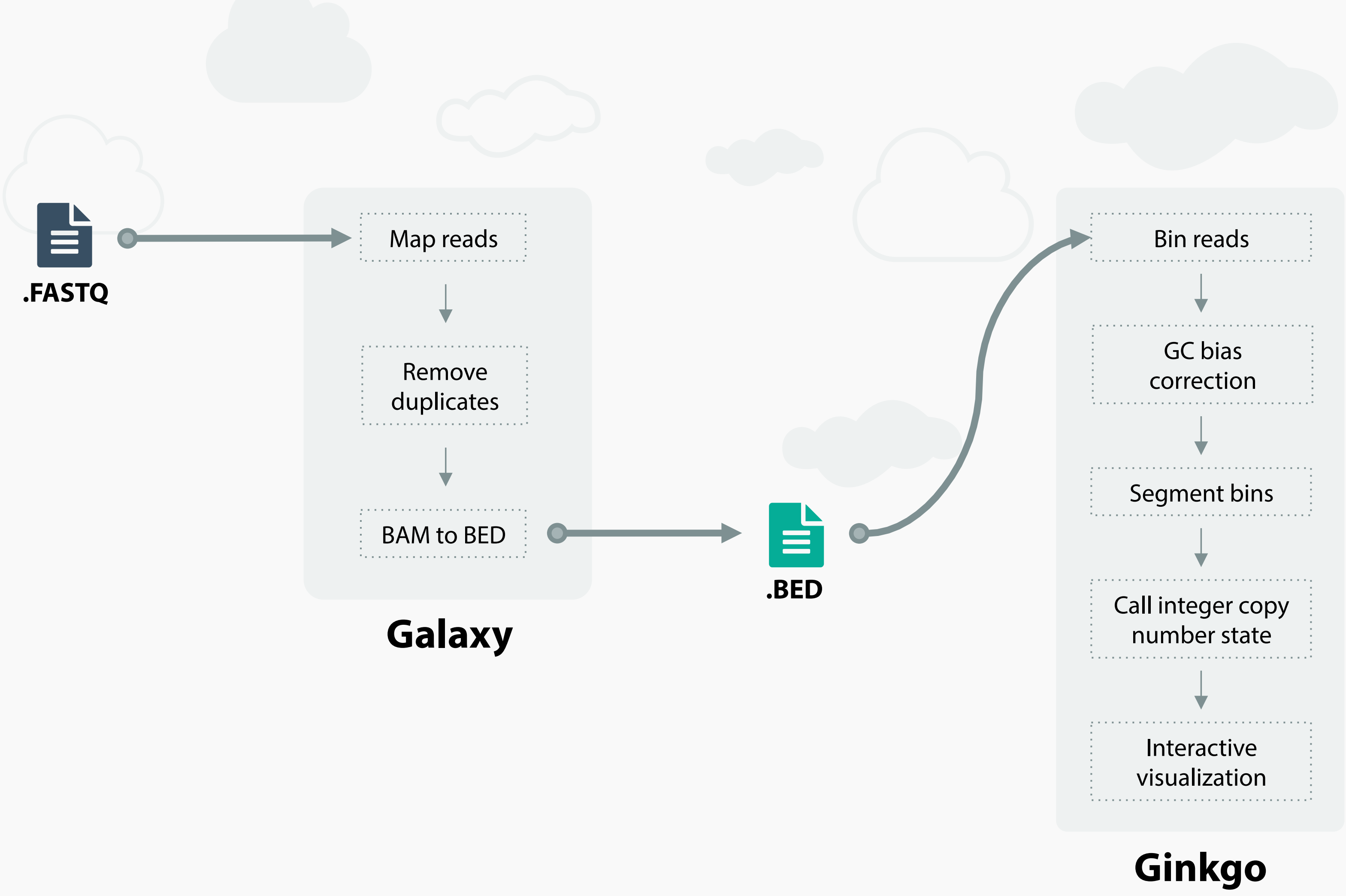
Copy-number variant analysis





Mike Schatz

Analysis Pipeline



.FASTQ

Map reads
Remove duplicates
BAM to BED

Galaxy

.BED

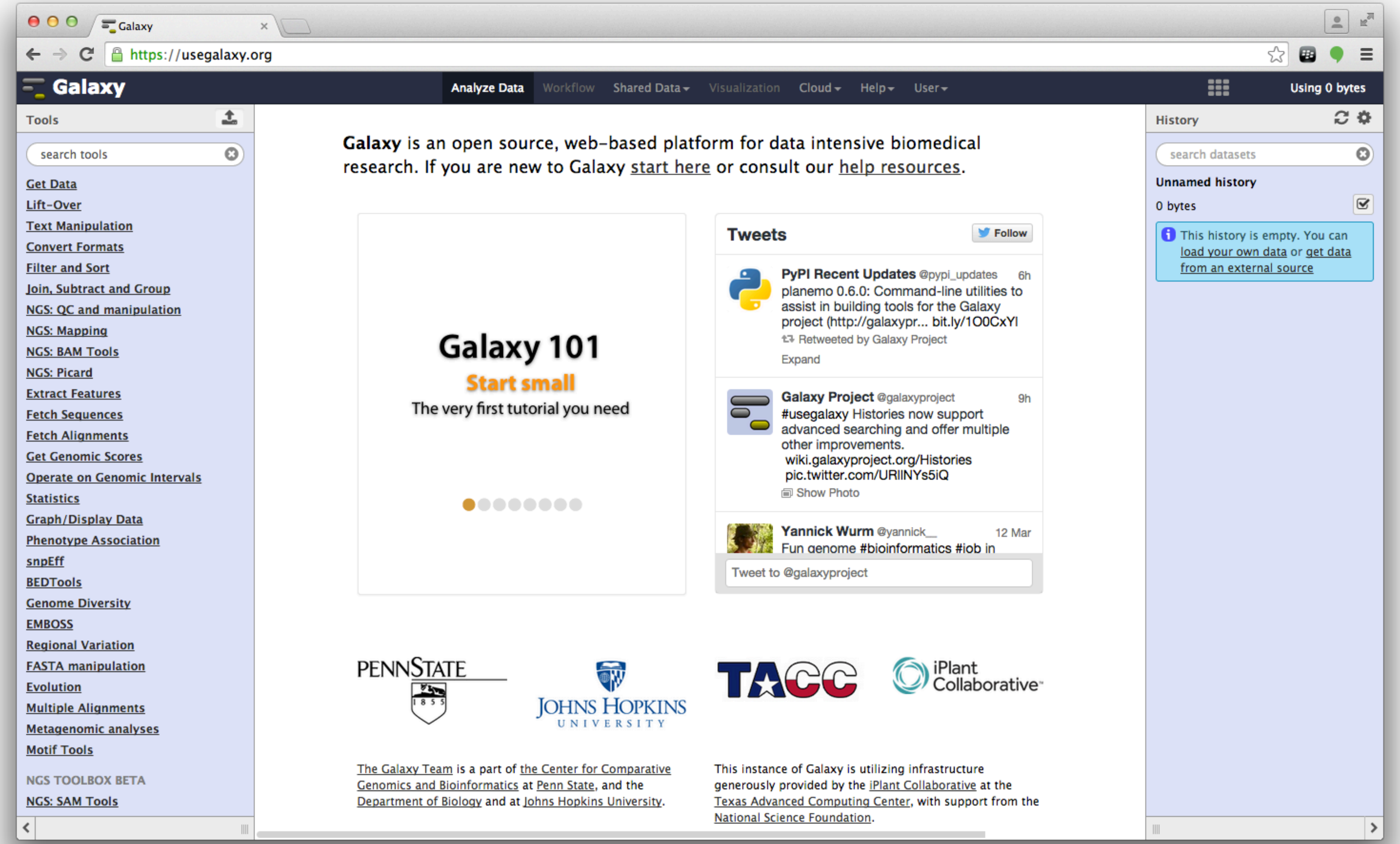
Bin reads
GC bias correction
Segment bins
Call integer copy number state
Interactive visualization

Ginkgo

Galaxy intro

Galaxy

Demo – Galaxy Pipeline



Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources.

Galaxy 101

Start small

The very first tutorial you need

PyPI Recent Updates @pypi_updates 6h
planemo 0.6.0: Command-line utilities to assist in building tools for the Galaxy project (http://galaxypr... bit.ly/1O0CxYl
Retweeted by Galaxy Project

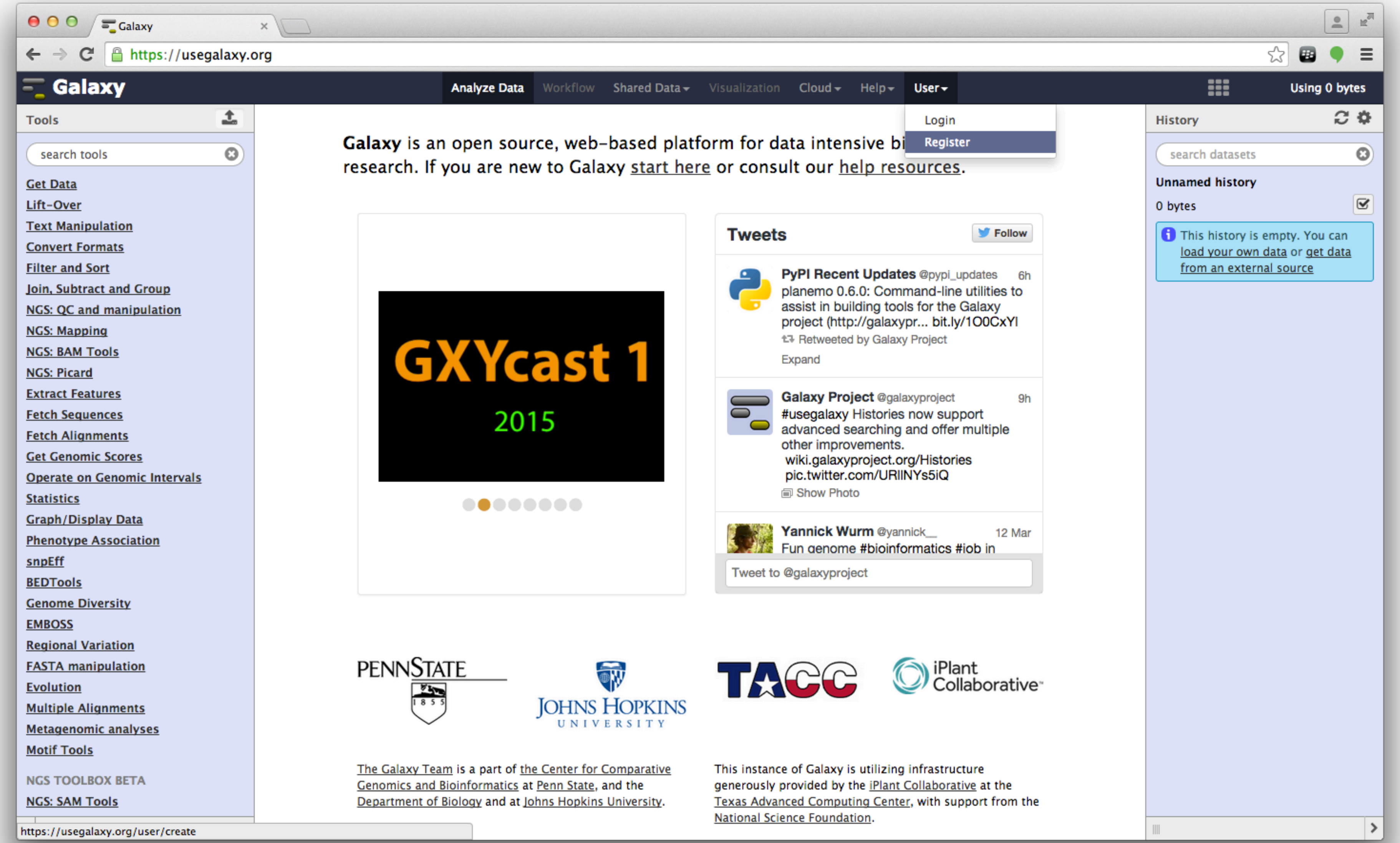
Galaxy Project @galaxyproject 9h
#usegalaxy Histories now support advanced searching and offer multiple other improvements.
wiki.galaxyproject.org/Histories
pic.twitter.com/URlINYs5iQ

Yannick Wurm @yannick__ 12 Mar
Fun genome #bioinformatics #iob in

This history is empty. You can load your own data or get data from an external source

The Galaxy Team is a part of the Center for Comparative Genomics and Bioinformatics at Penn State, and the Department of Biology at Johns Hopkins University.

This instance of Galaxy is utilizing infrastructure generously provided by the iPlant Collaborative at the Texas Advanced Computing Center, with support from the National Science Foundation.



Galaxy is an open source, web-based platform for data intensive bi… research. If you are new to Galaxy start here or consult our help resources.

This history is empty. You can load your own data or get data from an external source

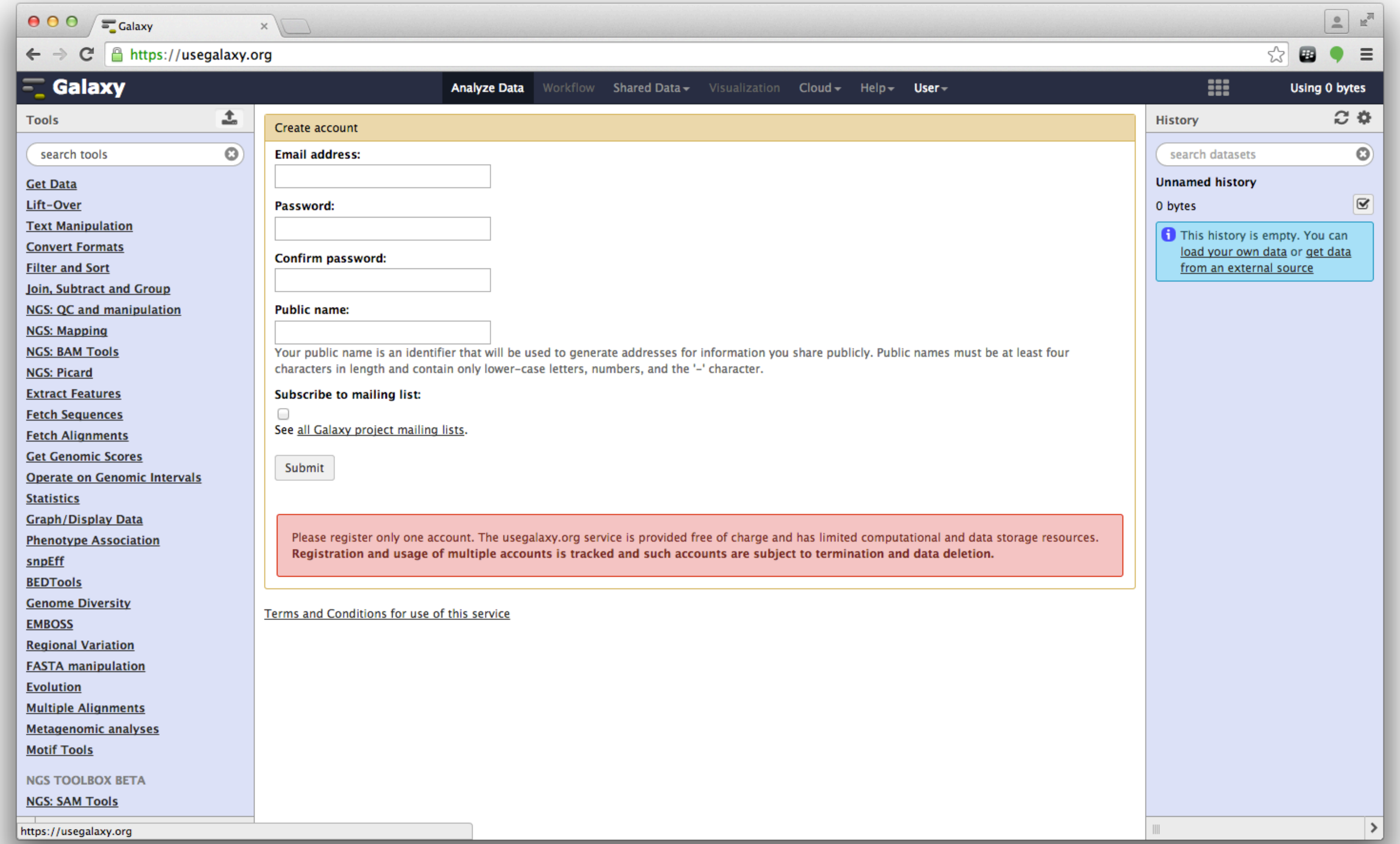
PyPI Recent Updates @pypi_updates 6h
planemo 0.6.0: Command-line utilities to assist in building tools for the Galaxy project (http://galaxypr... bit.ly/1O0CxYI
Retweeted by Galaxy Project
Expand

Galaxy Project @galaxyproject 9h
#usegalaxy Histories now support advanced searching and offer multiple other improvements.
wiki.galaxyproject.org/Histories
pic.twitter.com/URlINYs5iQ
Show Photo

Yannick Wurm @yannick__ 12 Mar
Fun genome #bioinformatics #job in

The Galaxy Team is a part of the Center for Comparative Genomics and Bioinformatics at Penn State, and the Department of Biology and at Johns Hopkins University.

This instance of Galaxy is utilizing infrastructure generously provided by the iPlant Collaborative at the Texas Advanced Computing Center, with support from the National Science Foundation.



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.

Subscribe to mailing list:

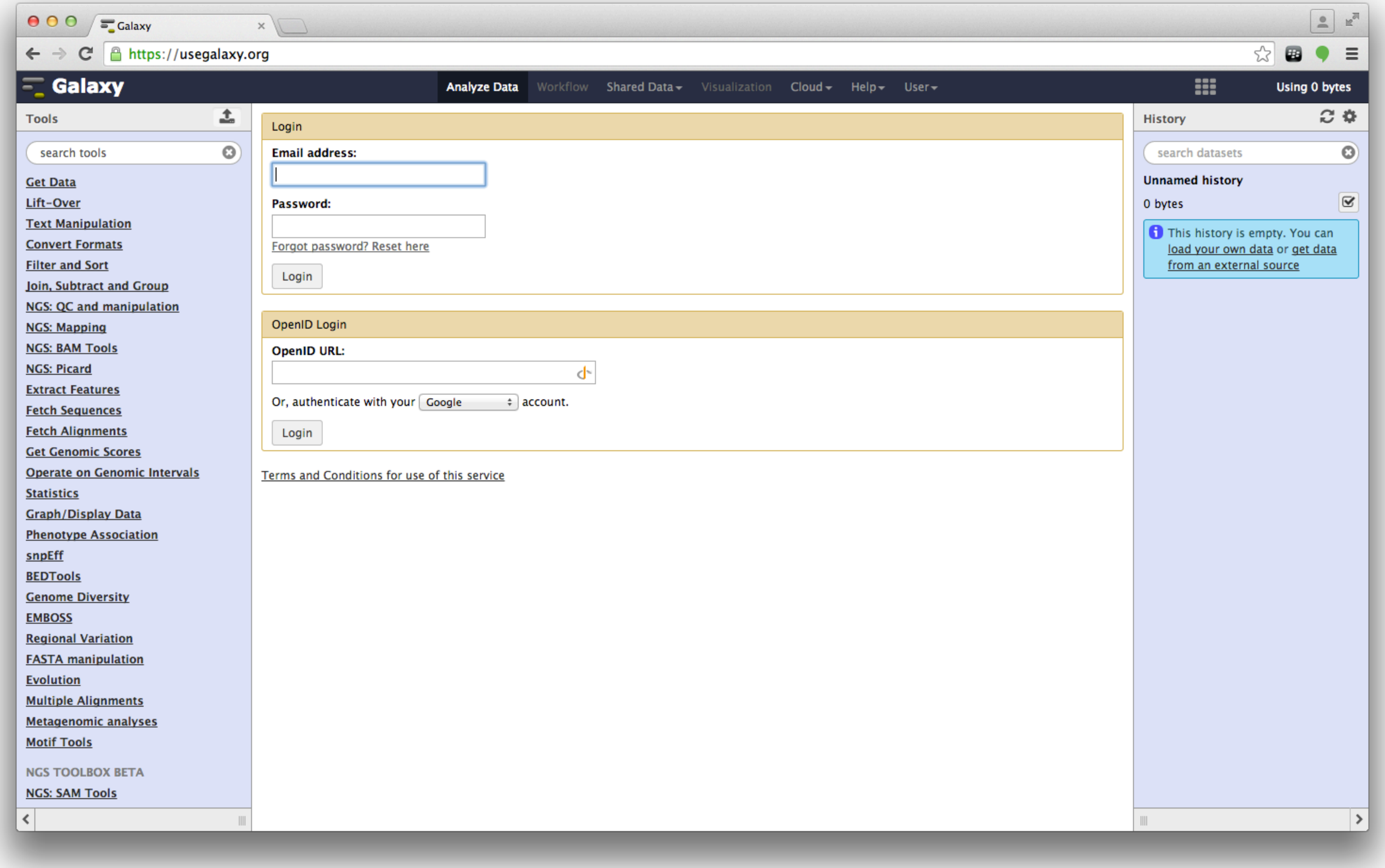
☐

See all Galaxy project mailing lists.

Please register only one account. The usegalaxy.org service is provided free of charge and has limited computational and data storage resources. **Registration and usage of multiple accounts is tracked and such accounts are subject to termination and data deletion.**

Terms and Conditions for use of this service

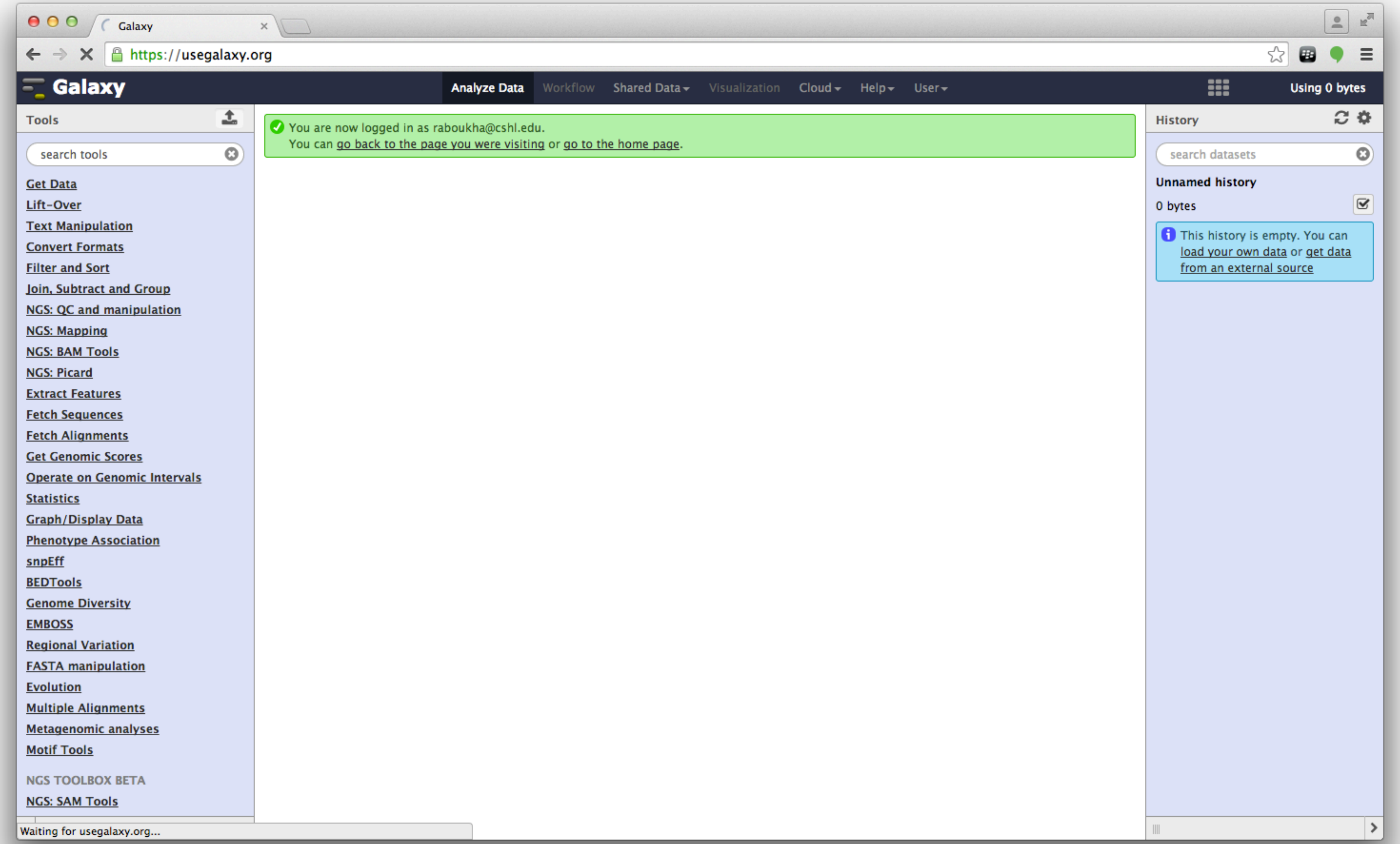
This history is empty. You can load your own data or get data from an external source



Or, authenticate with your Google account.

Terms and Conditions for use of this service

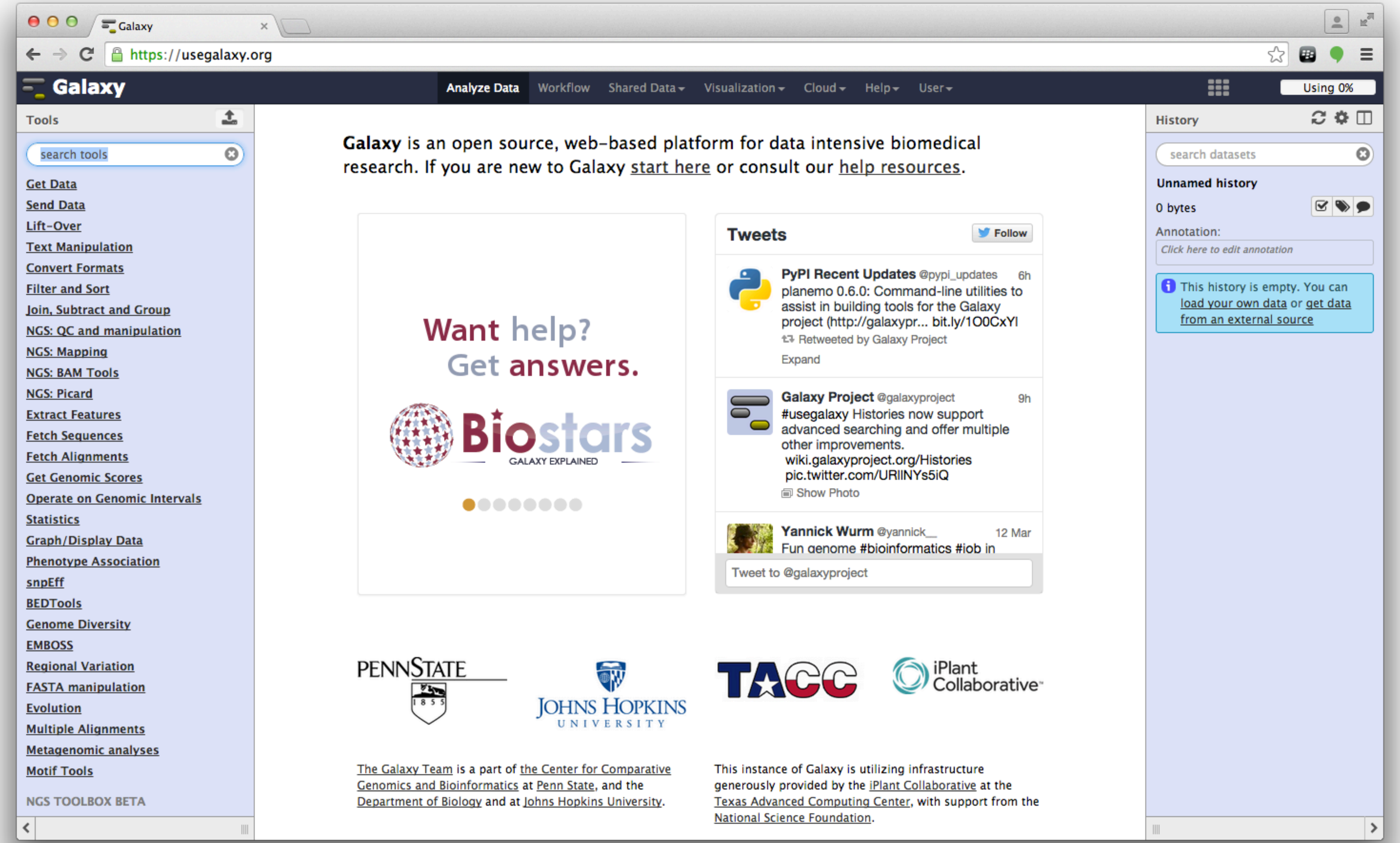
This history is empty. You can load your own data or get data from an external source



You are now logged in as raboukha@cshl.edu.
You can go back to the page you were visiting or go to the home page.

This history is empty. You can load your own data or get data from an external source

Waiting for usegalaxy.org...



Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources.

PyPI Recent Updates @pypi_updates 6h
planemo 0.6.0: Command-line utilities to assist in building tools for the Galaxy project (http://galaxypr... bit.ly/1O0CxYl
Retweeted by Galaxy Project
Expand

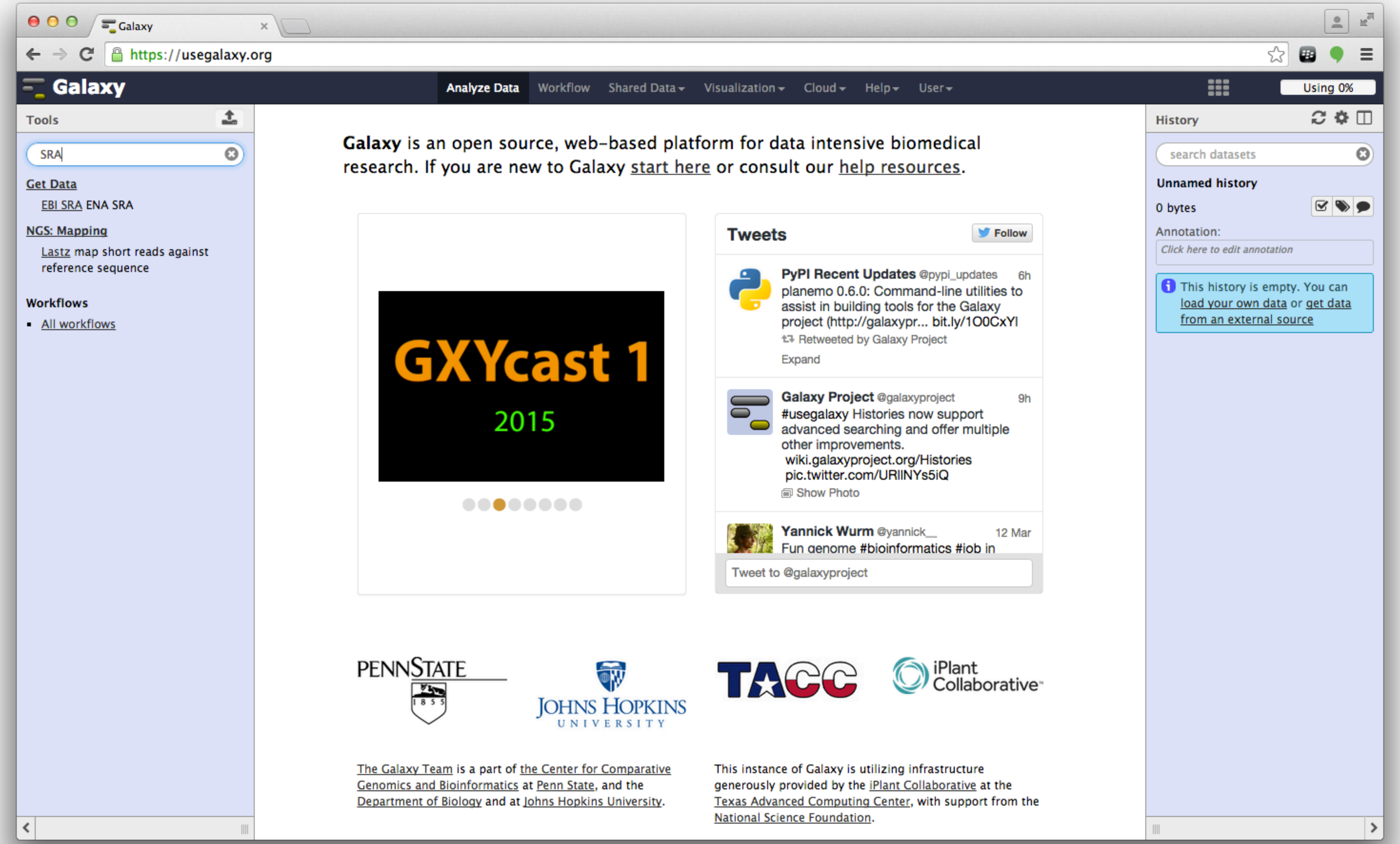
Galaxy Project @galaxyproject 9h
#usegalaxy Histories now support advanced searching and offer multiple other improvements.
wiki.galaxyproject.org/Histories
pic.twitter.com/URIlNYs5iQ
Show Photo

Yannick Wurm @yannick__ 12 Mar
Fun genome #bioinformatics #iob in

This history is empty. You can load your own data or get data from an external source

The Galaxy Team is a part of the Center for Comparative Genomics and Bioinformatics at Penn State, and the Department of Biology and at Johns Hopkins University.

This instance of Galaxy is utilizing infrastructure generously provided by the iPlant Collaborative at the Texas Advanced Computing Center, with support from the National Science Foundation.



Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources.

This history is empty. You can load your own data or get data from an external source

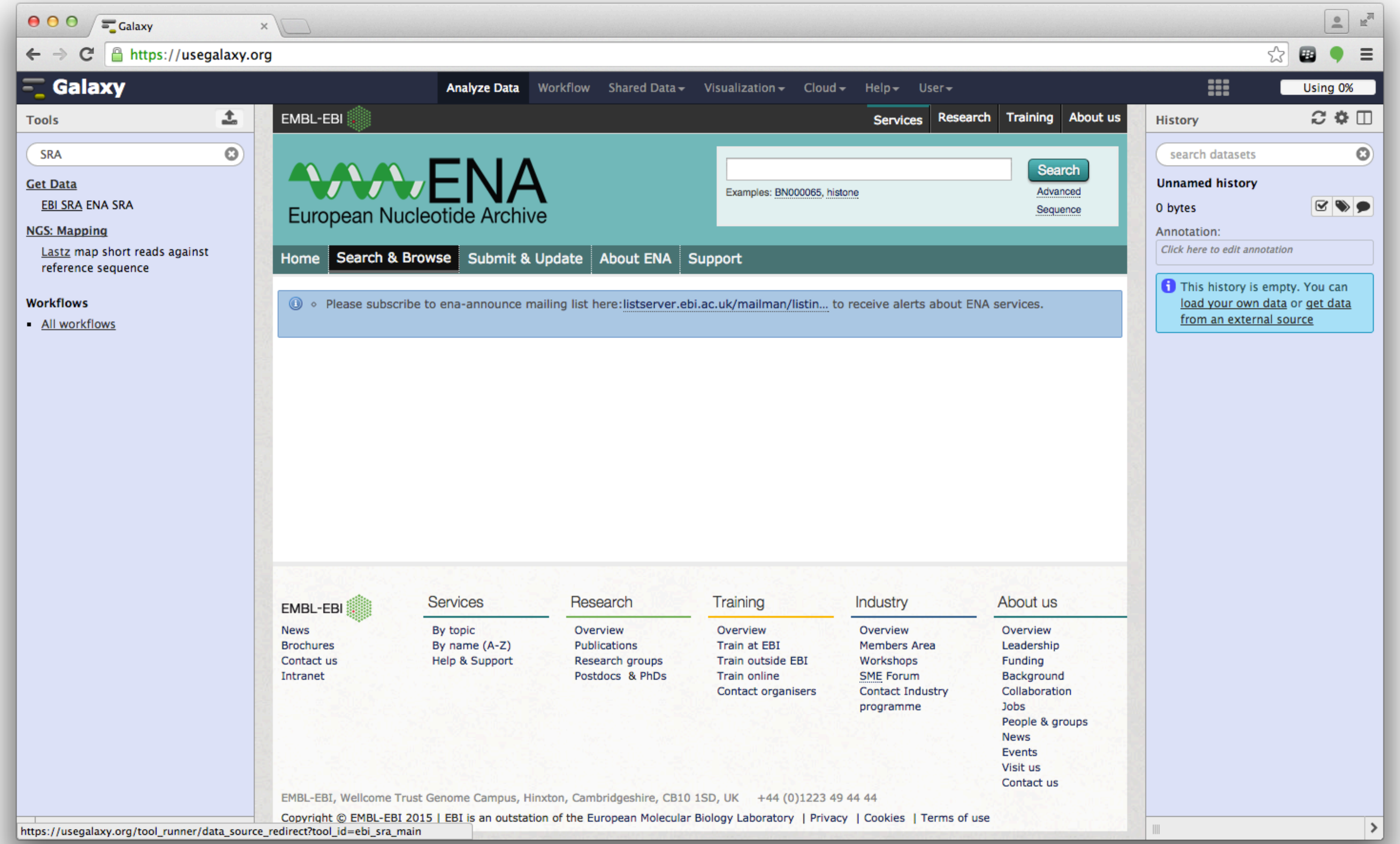
PyPI Recent Updates @pypi_updates 6h
planemo 0.6.0: Command-line utilities to assist in building tools for the Galaxy project (http://galaxypr... bit.ly/1O0CxYI
Retweeted by Galaxy Project

Galaxy Project @galaxyproject 9h
#usegalaxy Histories now support advanced searching and offer multiple other improvements.
wiki.galaxyproject.org/Histories
pic.twitter.com/URlINYs5iQ

Yannick Wurm @yannick__ 12 Mar
Fun genome #bioinformatics #iob in

The Galaxy Team is a part of the Center for Comparative Genomics and Bioinformatics at Penn State, and the Department of Biology and at Johns Hopkins University.

This instance of Galaxy is utilizing infrastructure generously provided by the iPlant Collaborative at the Texas Advanced Computing Center, with support from the National Science Foundation.



Please subscribe to ena-announce mailing list here:listserver.ebi.ac.uk/mailman/listin... to receive alerts about ENA services.

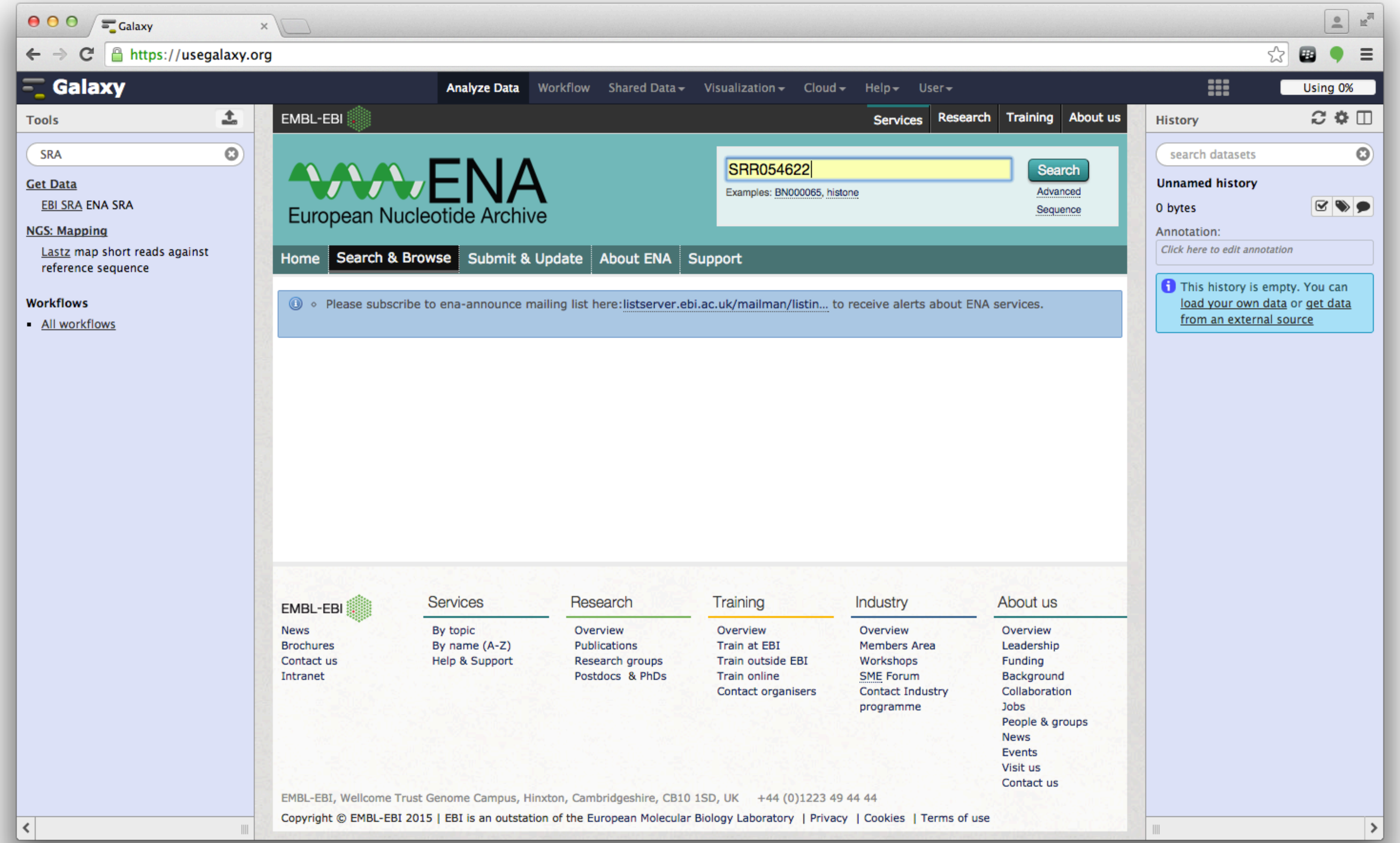
This history is empty. You can load your own data or get data from an external source

EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK +44 (0)1223 49 44 44

Copyright © EMBL-EBI 2015 | EBI is an outstation of the European Molecular Biology Laboratory | Privacy | Cookies | Terms of use

https://usegalaxy.org/tool_runner/data_source_redirect?tool_id=ebi_sra_main

SRR054622

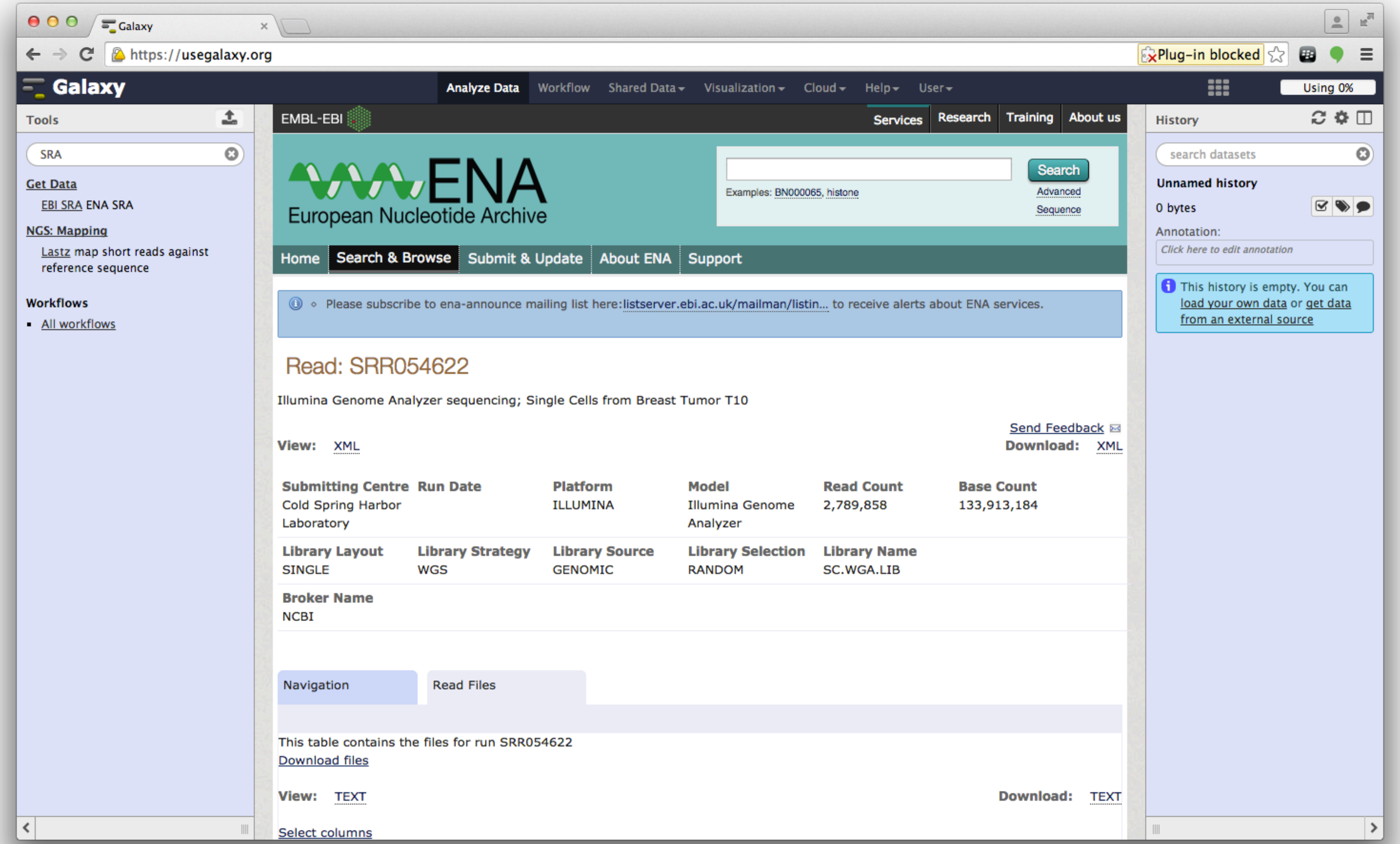


Tools search: SRA

ENA search: SRR054622

Please subscribe to ena-announce mailing list here:listserver.ebi.ac.uk/mailman/listin... to receive alerts about ENA services.

This history is empty. You can load your own data or get data from an external source



This history is empty. You can load your own data or get data from an external source

Please subscribe to ena-announce mailing list here:listserver.ebi.ac.uk/mailman/listin... to receive alerts about ENA services.

Read: SRR054622

Illumina Genome Analyzer sequencing; Single Cells from Breast Tumor T10

View: XML

Download: XML

Submitting Centre	Run Date	Platform	Model	Read Count	Base Count
Cold Spring Harbor Laboratory		ILLUMINA	Illumina Genome Analyzer	2,789,858	133,913,184

Library Layout	Library Strategy	Library Source	Library Selection	Library Name
SINGLE	WGS	GENOMIC	RANDOM	SC.WGA.LIB

Broker Name
NCBI

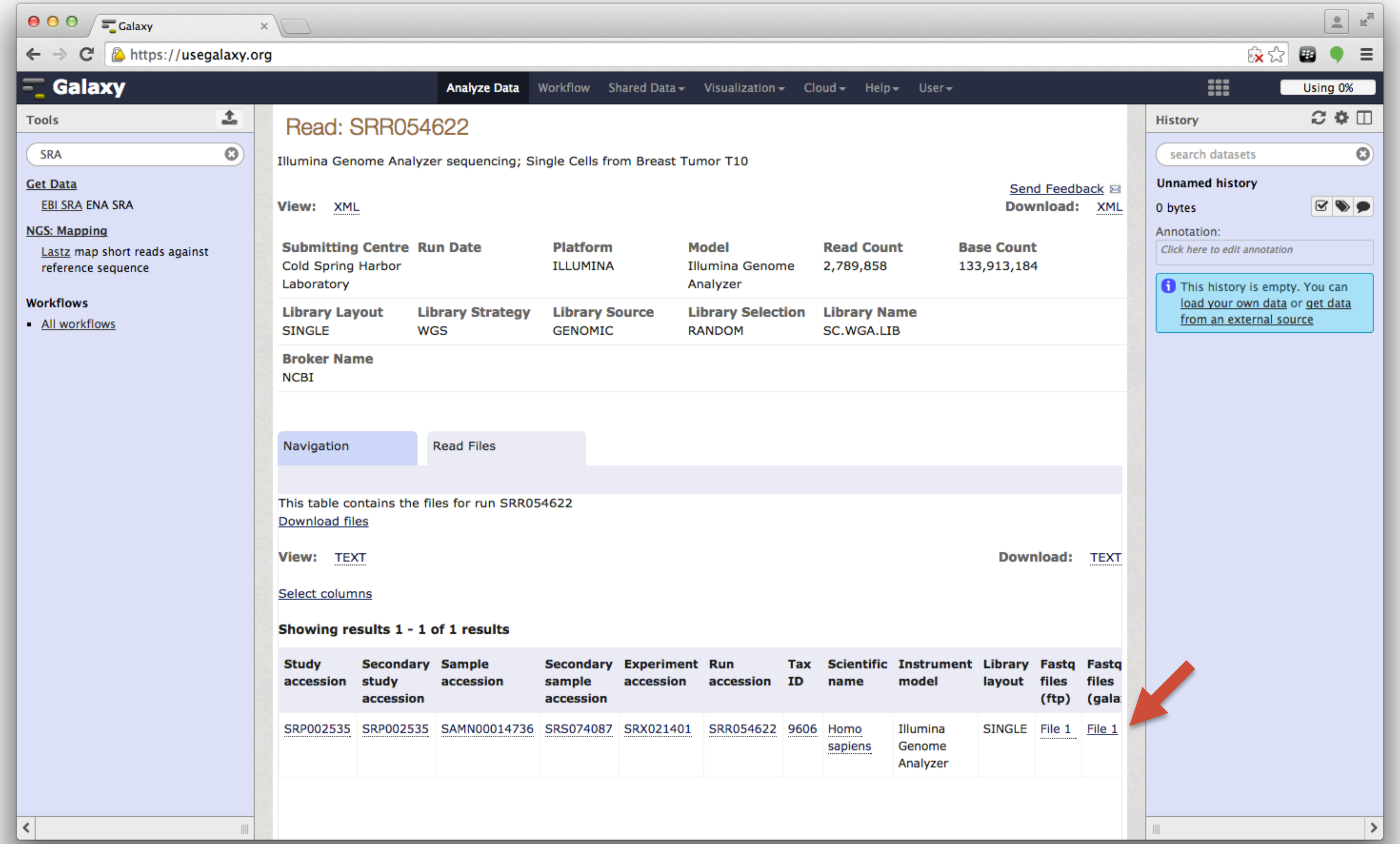
This table contains the files for run SRR054622

Download files

View: TEXT

Download: TEXT

Select columns



Read: SRR054622

Illumina Genome Analyzer sequencing; Single Cells from Breast Tumor T10

Send Feedback

View: XML

Download: XML

Submitting Centre	Run Date	Platform	Model	Read Count	Base Count
Cold Spring Harbor Laboratory		ILLUMINA	Illumina Genome Analyzer	2,789,858	133,913,184
Library Layout	**Library Strategy**	**Library Source**	**Library Selection**	**Library Name**	
SINGLE	WGS	GENOMIC	RANDOM	SC.WGA.LIB	
Broker Name					
NCBI					

This table contains the files for run SRR054622

Download files

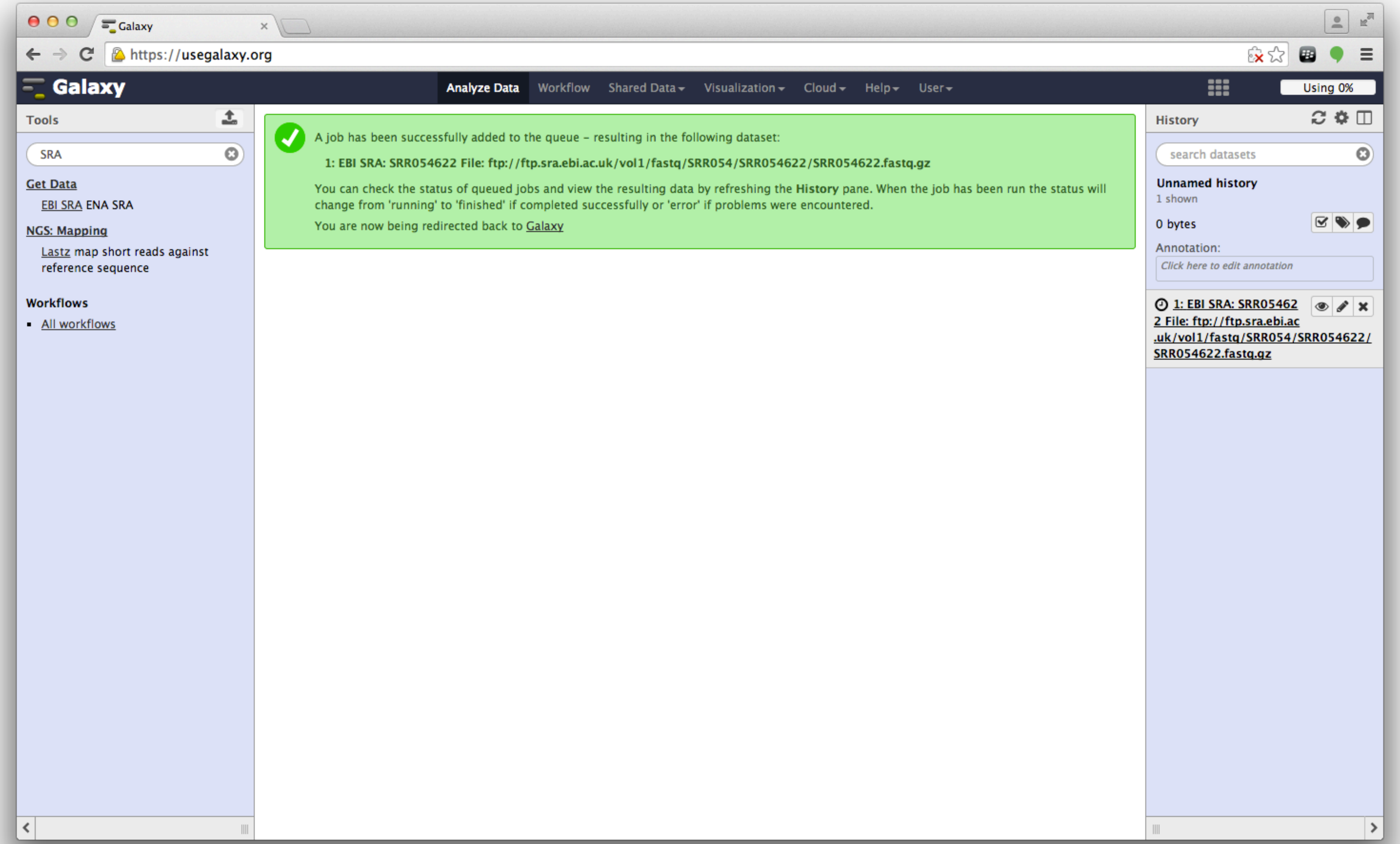
View: TEXT

Download: TEXT

Select columns

Showing results 1 - 1 of 1 results

Study accession	Secondary study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	Fastq files (ftp)	Fastq files (gala
SRP002535	SRP002535	SAMN00014736	SRS074087	SRX021401	SRR054622	9606	Homo sapiens	Illumina Genome Analyzer	SINGLE	File 1	File 1



Get Data

EBI SRA ENA SRA

NGS: Mapping

Lastz map short reads against reference sequence

Workflows

- All workflows

A job has been successfully added to the queue – resulting in the following dataset:

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

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.

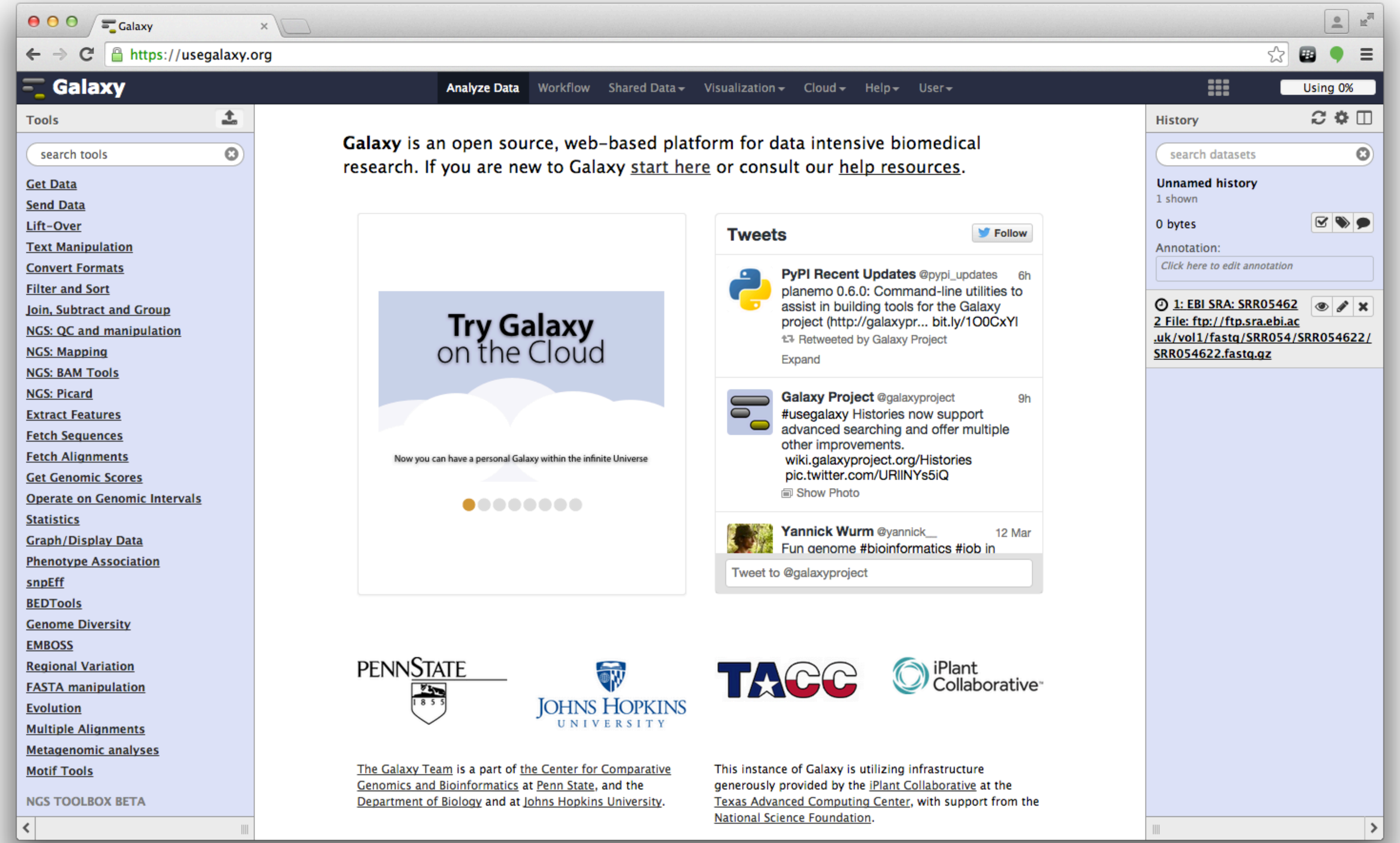
You are now being redirected back to Galaxy

Unnamed history
1 shown

0 bytes

Annotation:
Click here to edit annotation

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz



Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources.

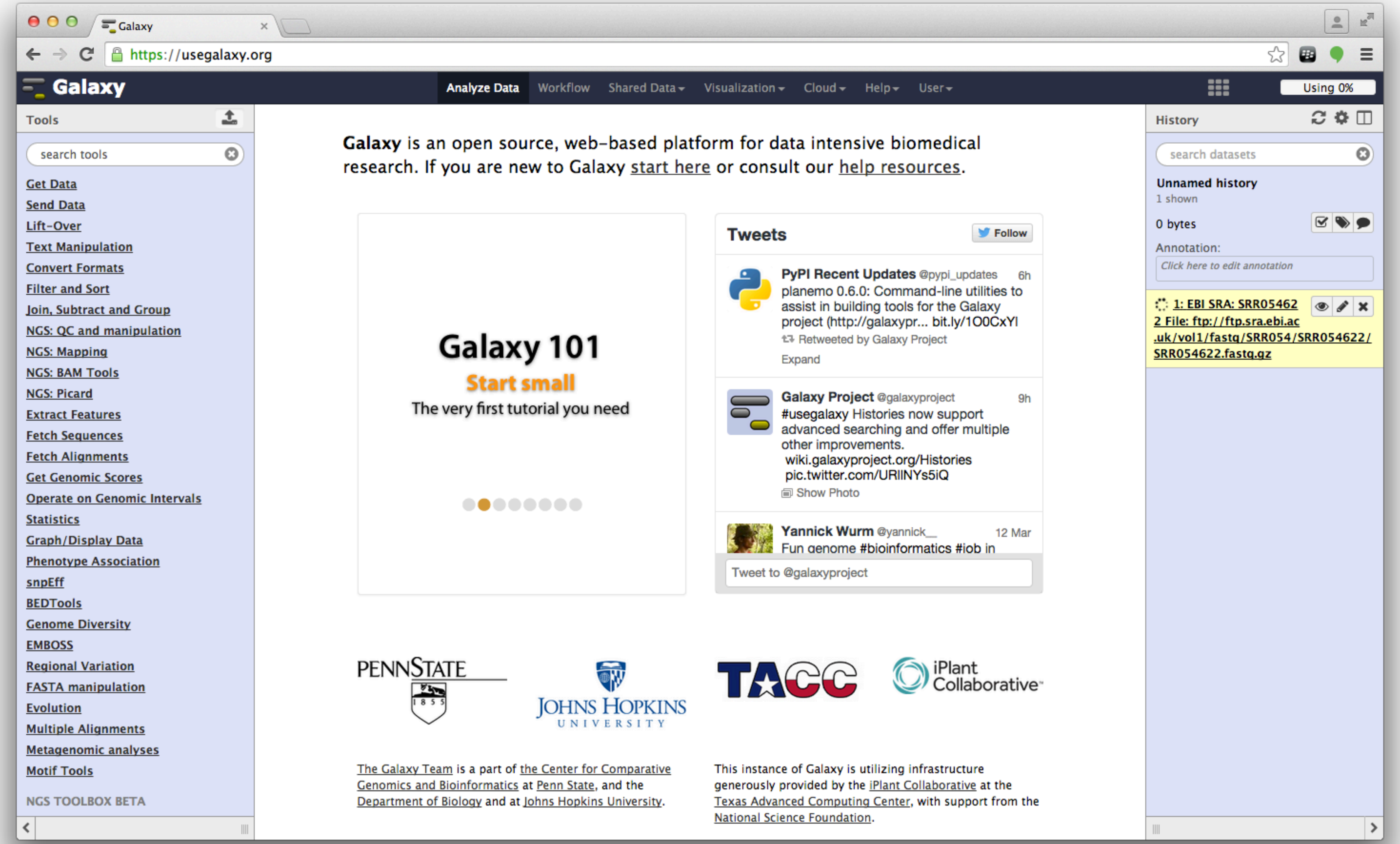
PyPI Recent Updates @pypi_updates 6h
planemo 0.6.0: Command-line utilities to assist in building tools for the Galaxy project (http://galaxypr... bit.ly/1O0CxYI
Retweeted by Galaxy Project
Expand

Galaxy Project @galaxyproject 9h
#usegalaxy Histories now support advanced searching and offer multiple other improvements.
wiki.galaxyproject.org/Histories
pic.twitter.com/URlINYs5iQ
Show Photo

Yannick Wurm @yannick__ 12 Mar
Fun genome #bioinformatics #iob in

The Galaxy Team is a part of the Center for Comparative Genomics and Bioinformatics at Penn State, and the Department of Biology and at Johns Hopkins University.

This instance of Galaxy is utilizing infrastructure generously provided by the iPlant Collaborative at the Texas Advanced Computing Center, with support from the National Science Foundation.



Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources.

PyPI Recent Updates @pypi_updates 6h
planemo 0.6.0: Command-line utilities to assist in building tools for the Galaxy project (http://galaxypr... bit.ly/1O0CxYl
Retweeted by Galaxy Project

Galaxy Project @galaxyproject 9h
#usegalaxy Histories now support advanced searching and offer multiple other improvements.
wiki.galaxyproject.org/Histories
pic.twitter.com/URIlNYs5iQ

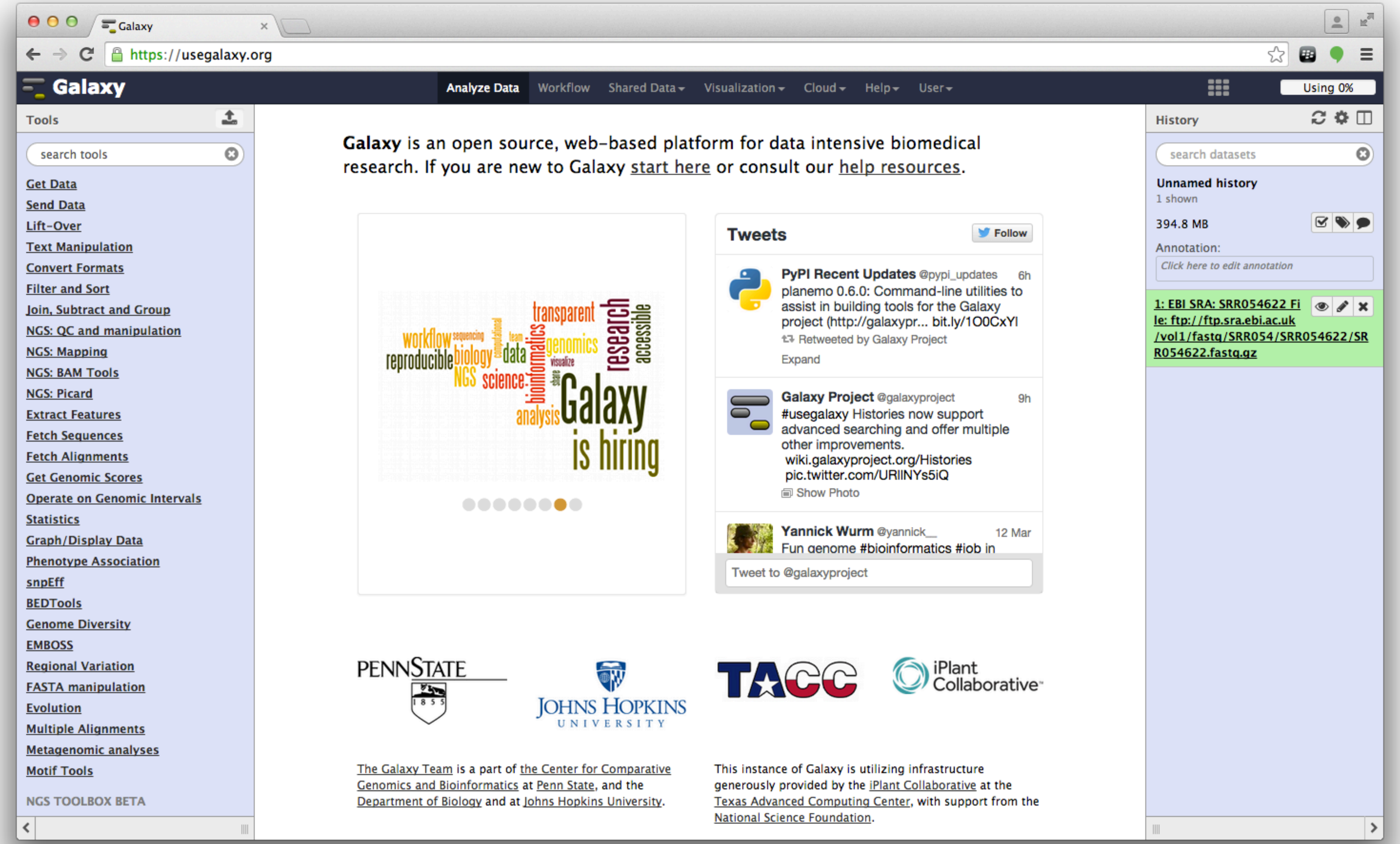
Yannick Wurm @yannick__ 12 Mar
Fun genome #bioinformatics #iob in

The Galaxy Team is a part of the Center for Comparative Genomics and Bioinformatics at Penn State, and the Department of Biology and at Johns Hopkins University.

This instance of Galaxy is utilizing infrastructure generously provided by the iPlant Collaborative at the Texas Advanced Computing Center, with support from the National Science Foundation.

Unnamed history
1 shown
0 bytes
Annotation:
Click here to edit annotation

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz



Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources.

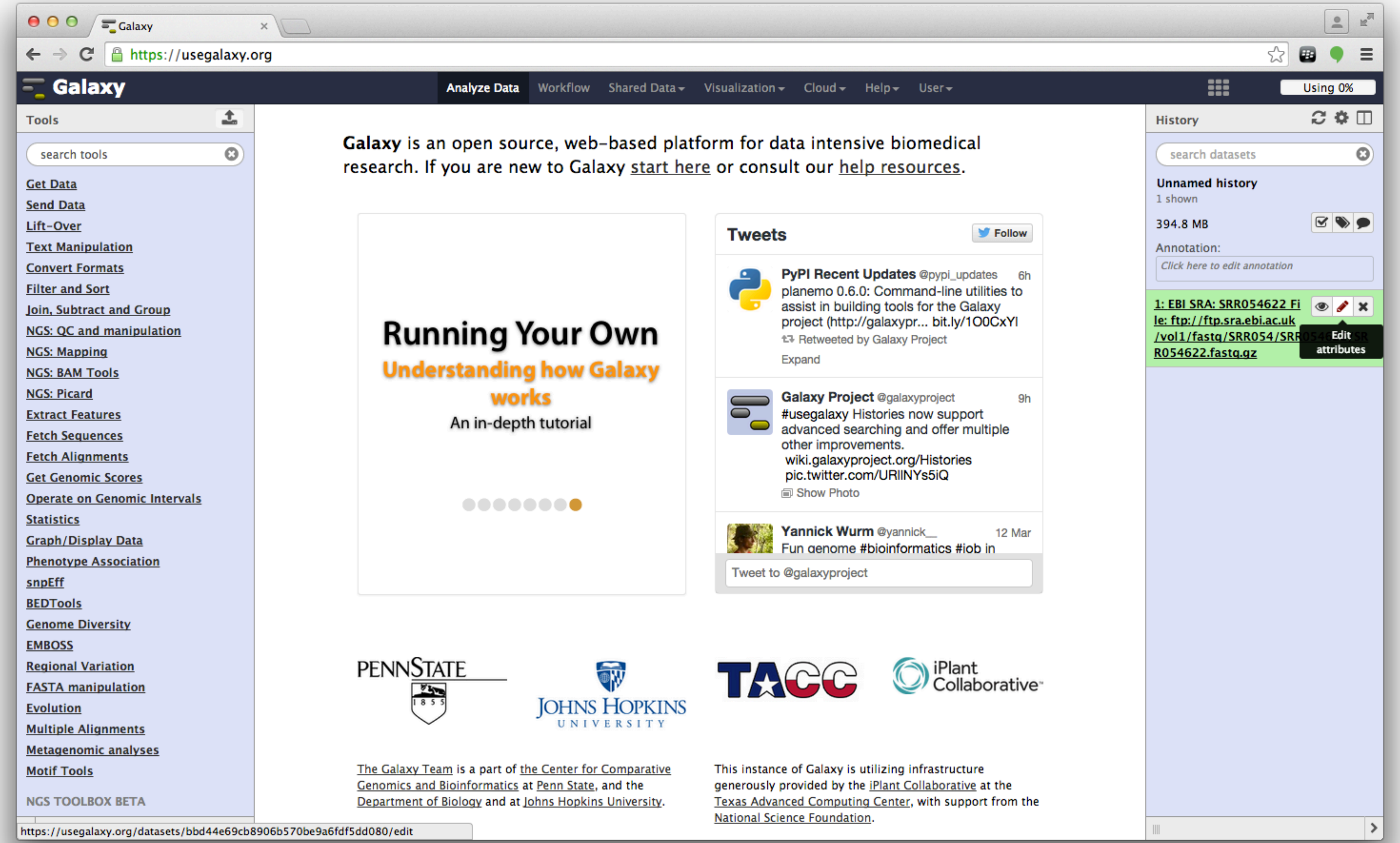
PyPI Recent Updates @pypi_updates 6h
planemo 0.6.0: Command-line utilities to assist in building tools for the Galaxy project (http://galaxypr... bit.ly/1O0CxYl
Retweeted by Galaxy Project
Expand

Galaxy Project @galaxyproject 9h
#usegalaxy Histories now support advanced searching and offer multiple other improvements.
wiki.galaxyproject.org/Histories
pic.twitter.com/URlINYs5iQ
Show Photo

Yannick Wurm @yannick__ 12 Mar
Fun genome #bioinformatics #iob in

The Galaxy Team is a part of the Center for Comparative Genomics and Bioinformatics at Penn State, and the Department of Biology and at Johns Hopkins University.

This instance of Galaxy is utilizing infrastructure generously provided by the iPlant Collaborative at the Texas Advanced Computing Center, with support from the National Science Foundation.



Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources.

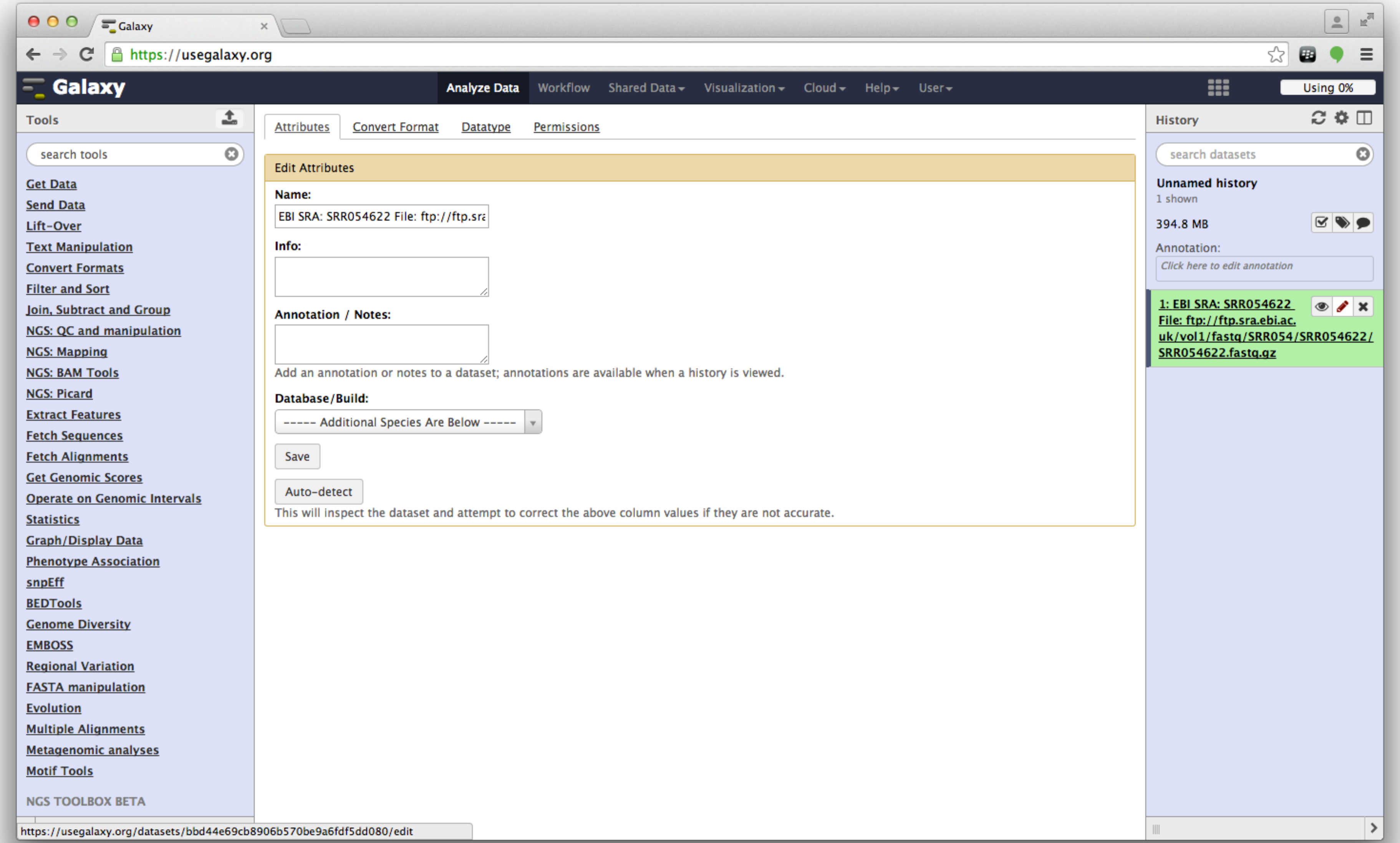
PyPI Recent Updates @pypi_updates 6h
planemo 0.6.0: Command-line utilities to assist in building tools for the Galaxy project (http://galaxypr... bit.ly/1O0CxYl
Retweeted by Galaxy Project
Expand

Galaxy Project @galaxyproject 9h
#usegalaxy Histories now support advanced searching and offer multiple other improvements.
wiki.galaxyproject.org/Histories
pic.twitter.com/URIlNYs5iQ
Show Photo

Yannick Wurm @yannick__ 12 Mar
Fun genome #bioinformatics #iob in

The Galaxy Team is a part of the Center for Comparative Genomics and Bioinformatics at Penn State, and the Department of Biology and at Johns Hopkins University.

This instance of Galaxy is utilizing infrastructure generously provided by the iPlant Collaborative at the Texas Advanced Computing Center, with support from the National Science Foundation.



Edit Attributes

Name: EBI SRA: SRR054622 File: ftp://ftp.sra

Info:

Annotation / Notes:

Add an annotation or notes to a dataset; annotations are available when a history is viewed.

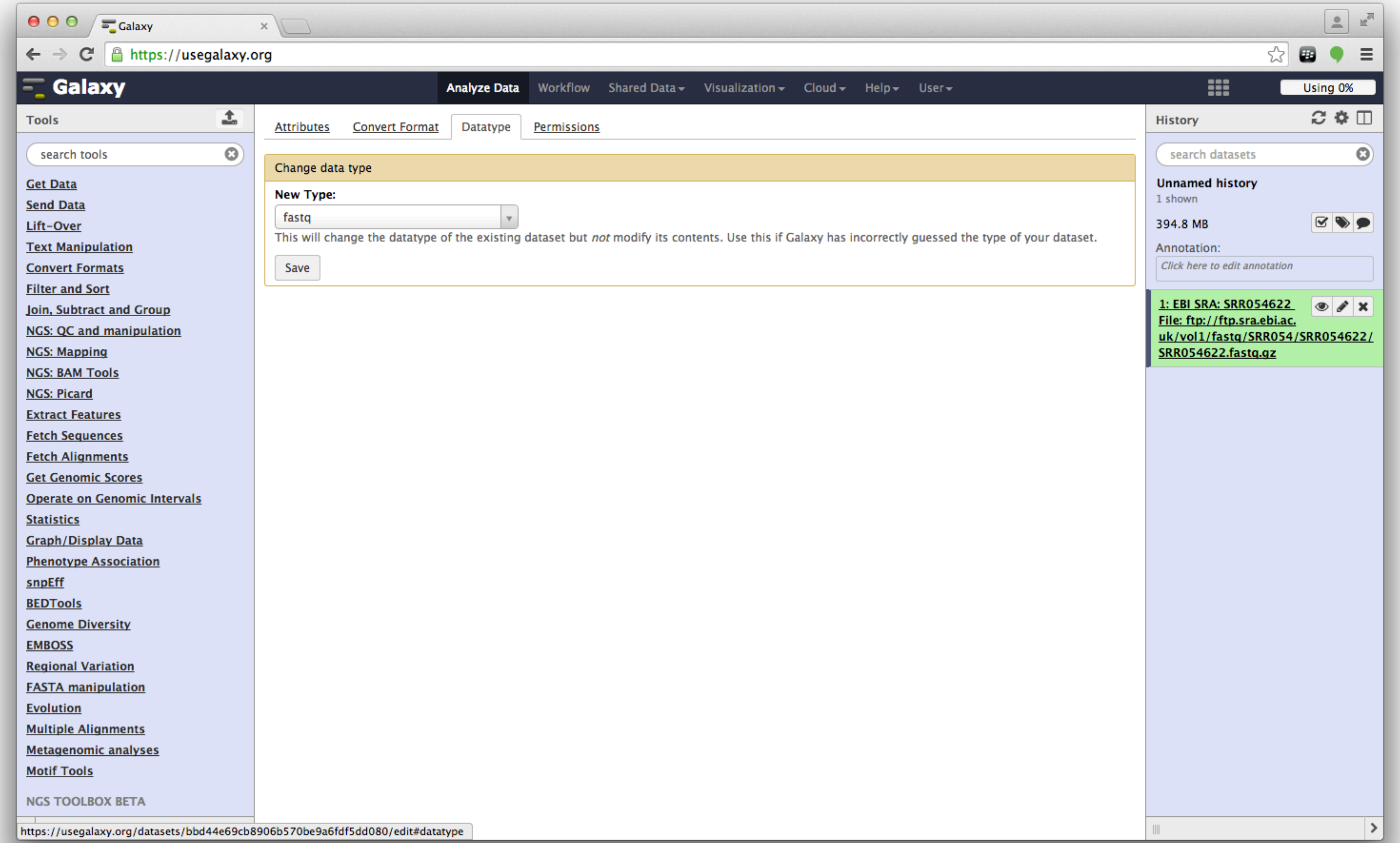
Database/Build: ----- Additional Species Are Below -----

This will inspect the dataset and attempt to correct the above column values if they are not accurate.

Unnamed history
1 shown
394.8 MB
Annotation:
Click here to edit annotation

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

https://usegalaxy.org/datasets/bbd44e69cb8906b570be9a6fdf5dd080/edit



Change data type

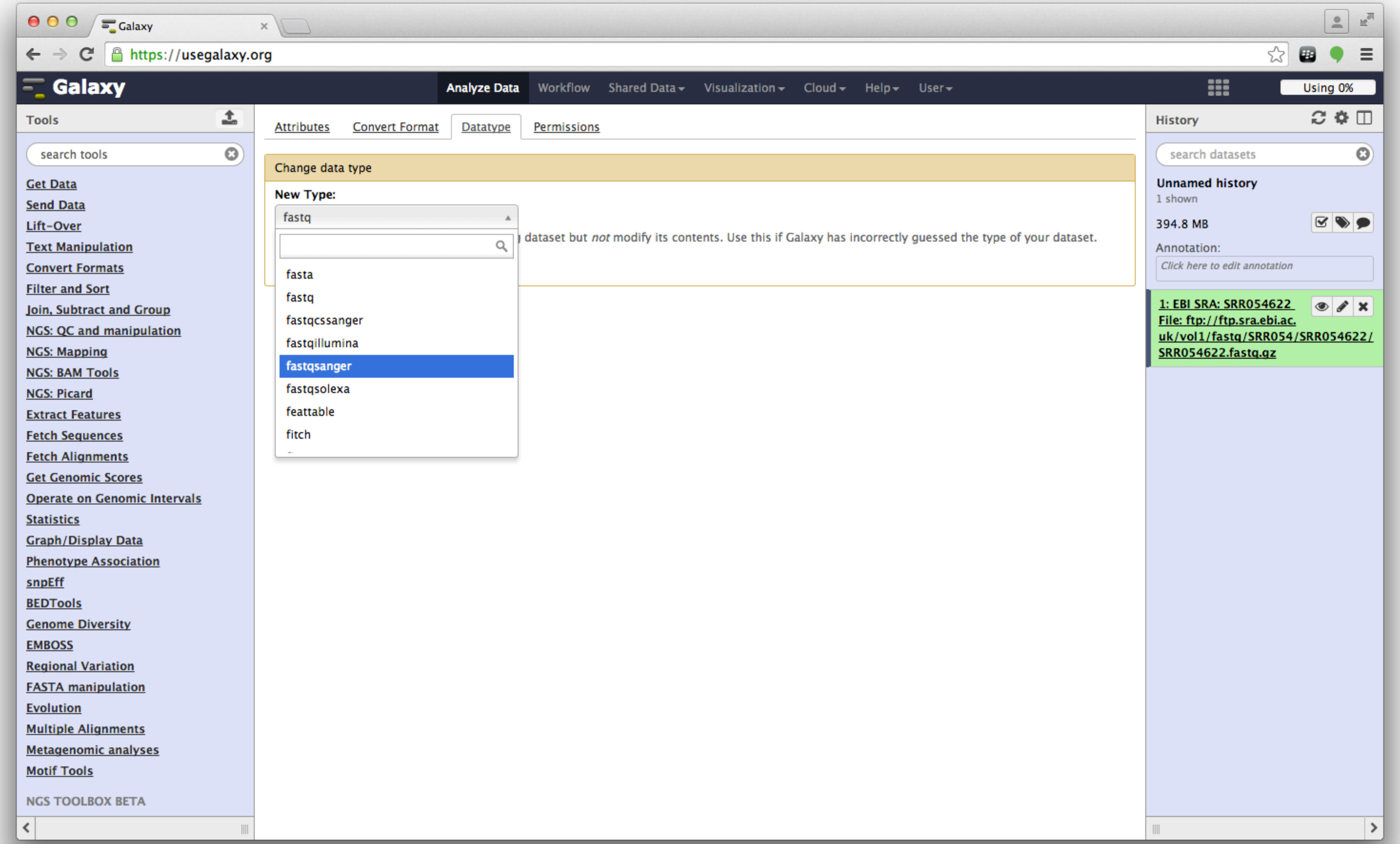
New Type: fastq

This will change the datatype of the existing dataset but *not* modify its contents. Use this if Galaxy has incorrectly guessed the type of your dataset.

Unnamed history
1 shown
394.8 MB
Annotation:
Click here to edit annotation

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

https://usegalaxy.org/datasets/bbd44e69cb8906b570be9a6fdf5dd080/edit#datatype



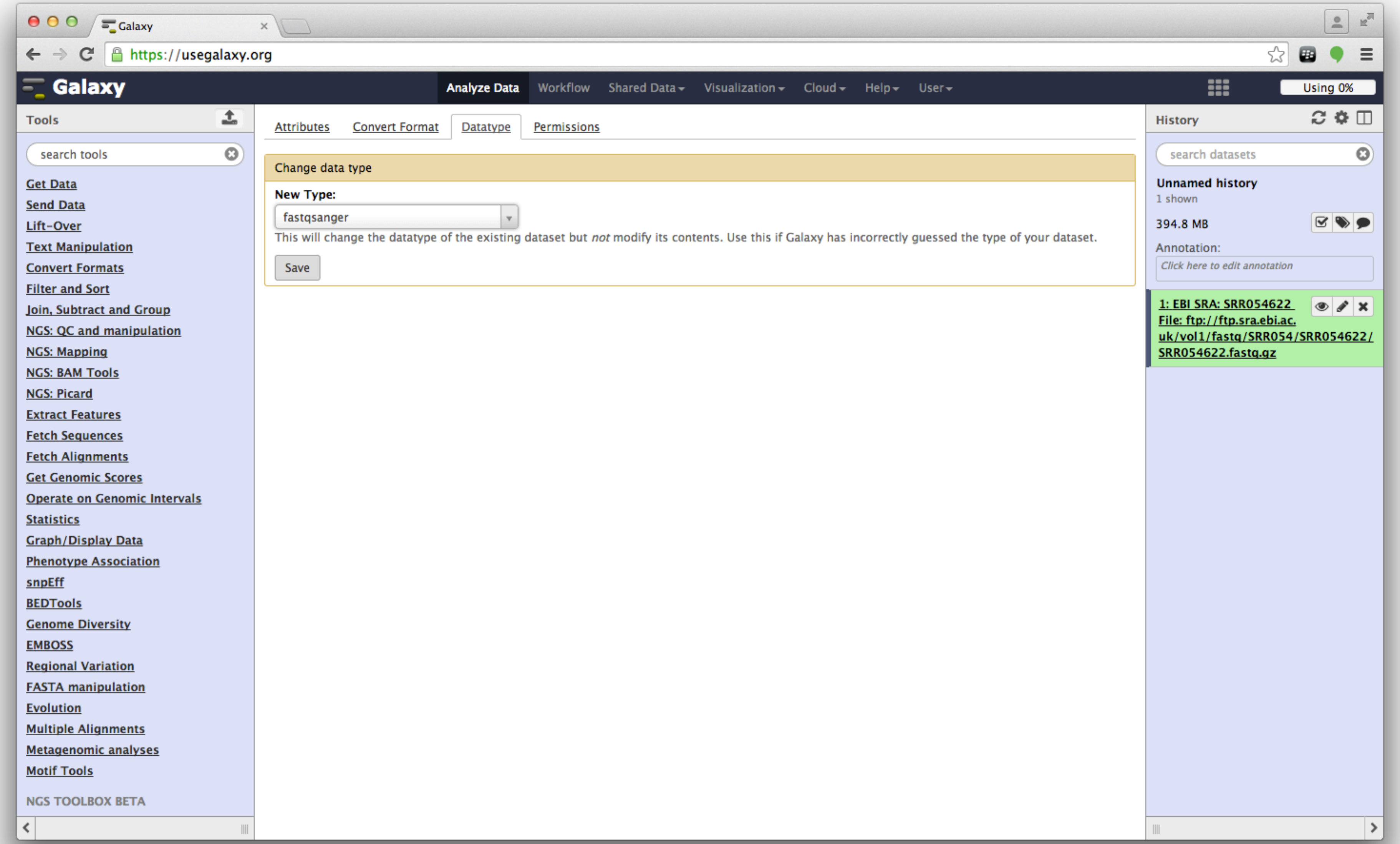
Change data type

New Type: fastqsanger

dataset but *not* modify its contents. Use this if Galaxy has incorrectly guessed the type of your dataset.

Unnamed history
1 shown
394.8 MB
Annotation:

1: EBI SRA: SRR054622
File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz



Change data type

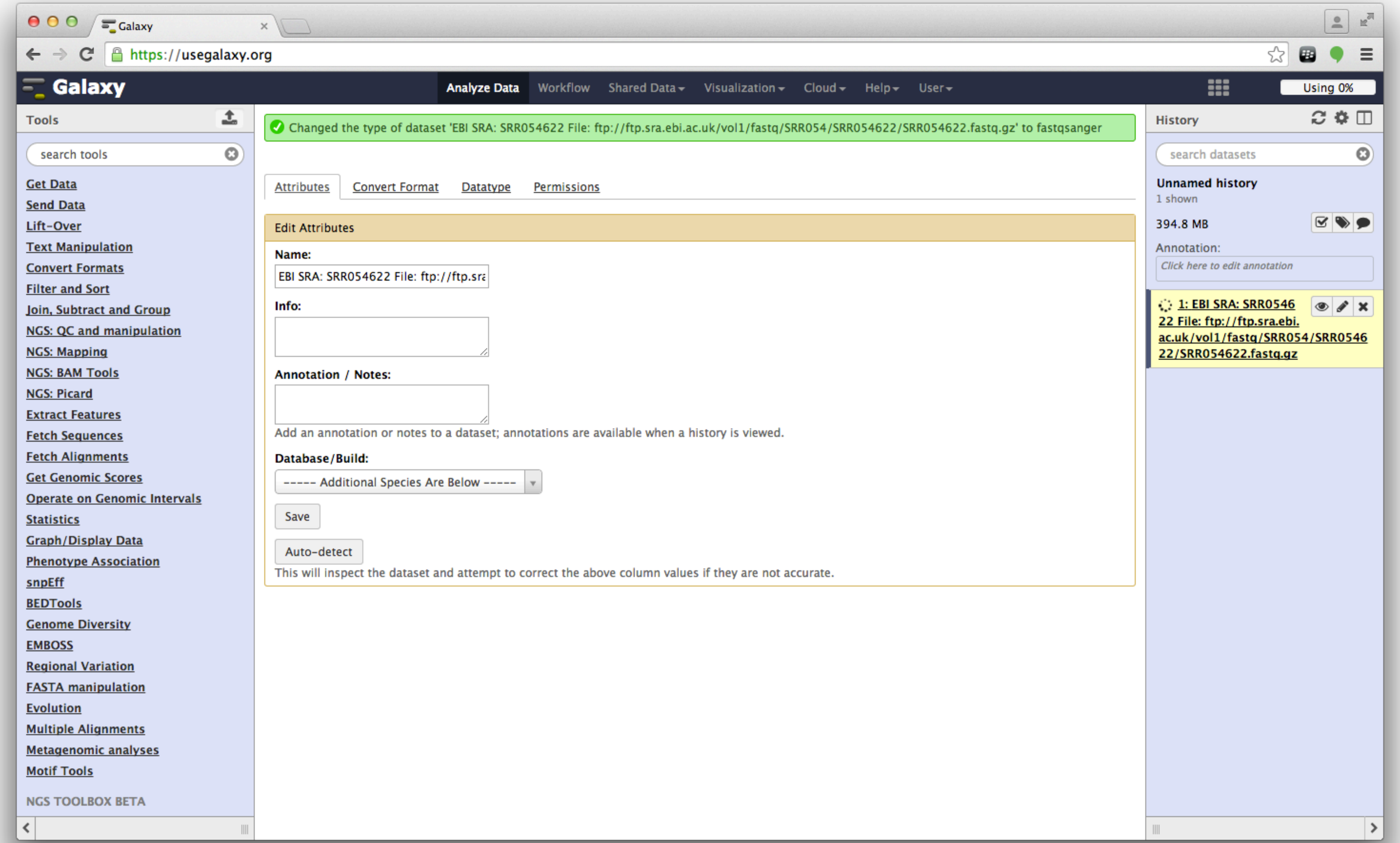
New Type: fastqsanger

This will change the datatype of the existing dataset but *not* modify its contents. Use this if Galaxy has incorrectly guessed the type of your dataset.

Save

Unnamed history
1 shown
394.8 MB
Annotation:

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz



Changed the type of dataset 'EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz' to fastqsanger

Name: EBI SRA: SRR054622 File: ftp://ftp.sra

Info:

Annotation / Notes:

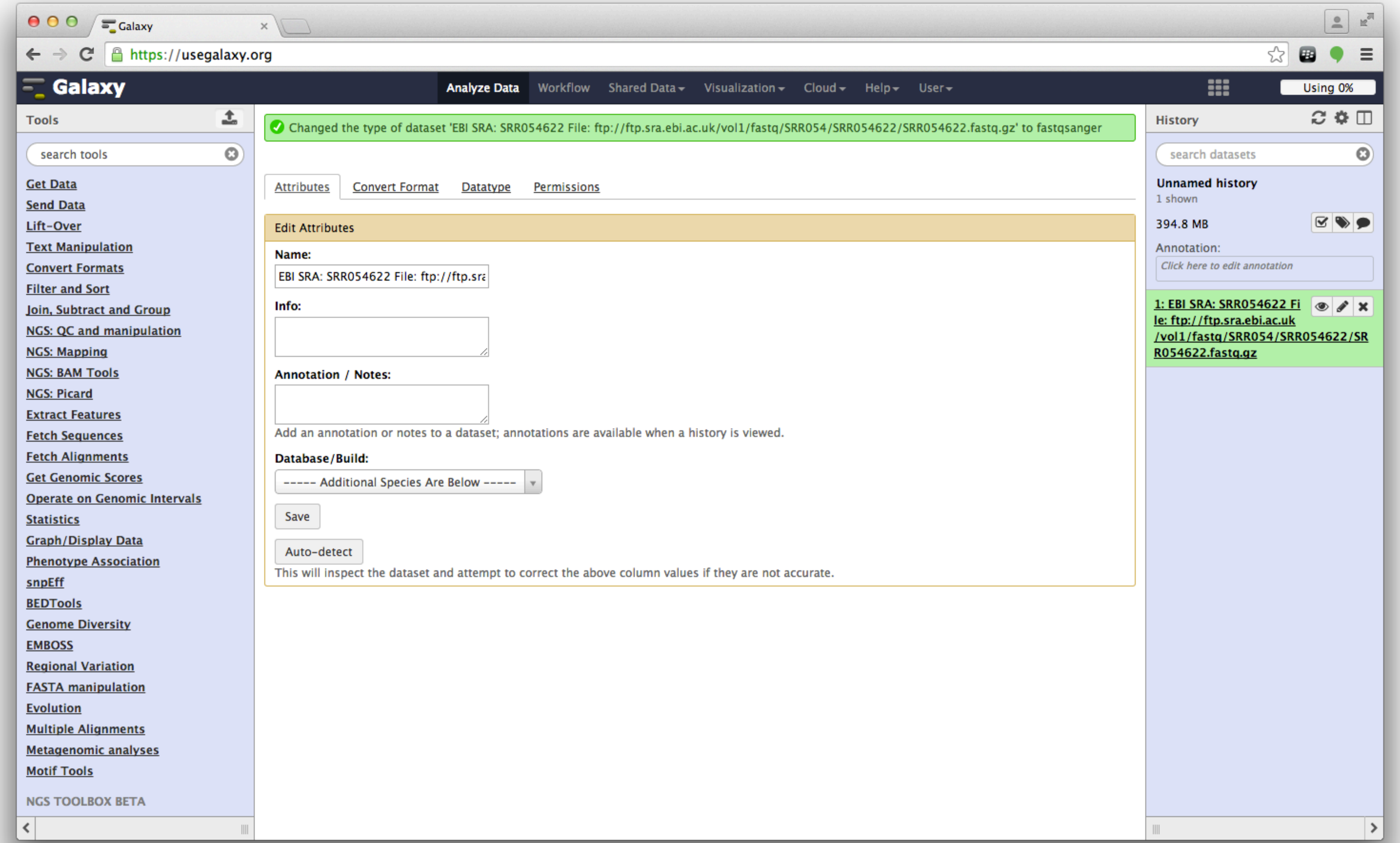
Add an annotation or notes to a dataset; annotations are available when a history is viewed.

Database/Build: ----- Additional Species Are Below -----

This will inspect the dataset and attempt to correct the above column values if they are not accurate.

Unnamed history
1 shown
394.8 MB
Annotation:

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz



✓ Changed the type of dataset 'EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz' to fastqsanger

Name: EBI SRA: SRR054622 File: ftp://ftp.sra

Info:

Annotation / Notes:

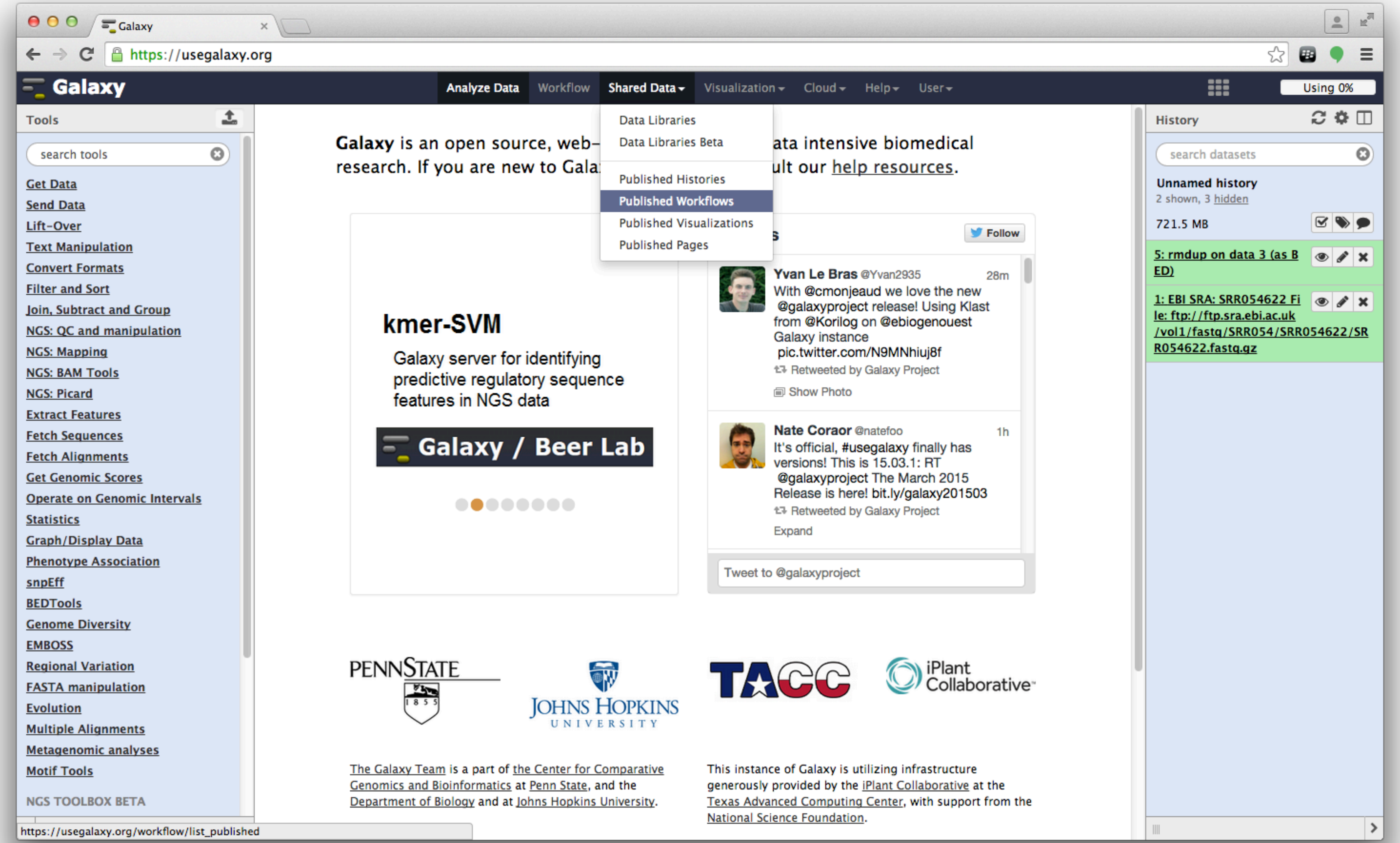
Add an annotation or notes to a dataset; annotations are available when a history is viewed.

Database/Build: ----- Additional Species Are Below -----

This will inspect the dataset and attempt to correct the above column values if they are not accurate.

Unnamed history
1 shown
394.8 MB
Annotation:

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz



Galaxy is an open source, web-… …ata intensive biomedical research. If you are new to Gala… …ult our help resources.

kmer-SVM

Galaxy server for identifying predictive regulatory sequence features in NGS data

Galaxy / Beer Lab

Yvan Le Bras @Yvan2935 28m
With @cmonjeaud we love the new @galaxyproject release! Using Klast from @Korilog on @ebiogenouest Galaxy instance pic.twitter.com/N9MNhiuj8f
Retweeted by Galaxy Project
Show Photo

Nate Coraor @natefoo 1h
It's official, #usegalaxy finally has versions! This is 15.03.1: RT @galaxyproject The March 2015 Release is here! bit.ly/galaxy201503
Retweeted by Galaxy Project
Expand

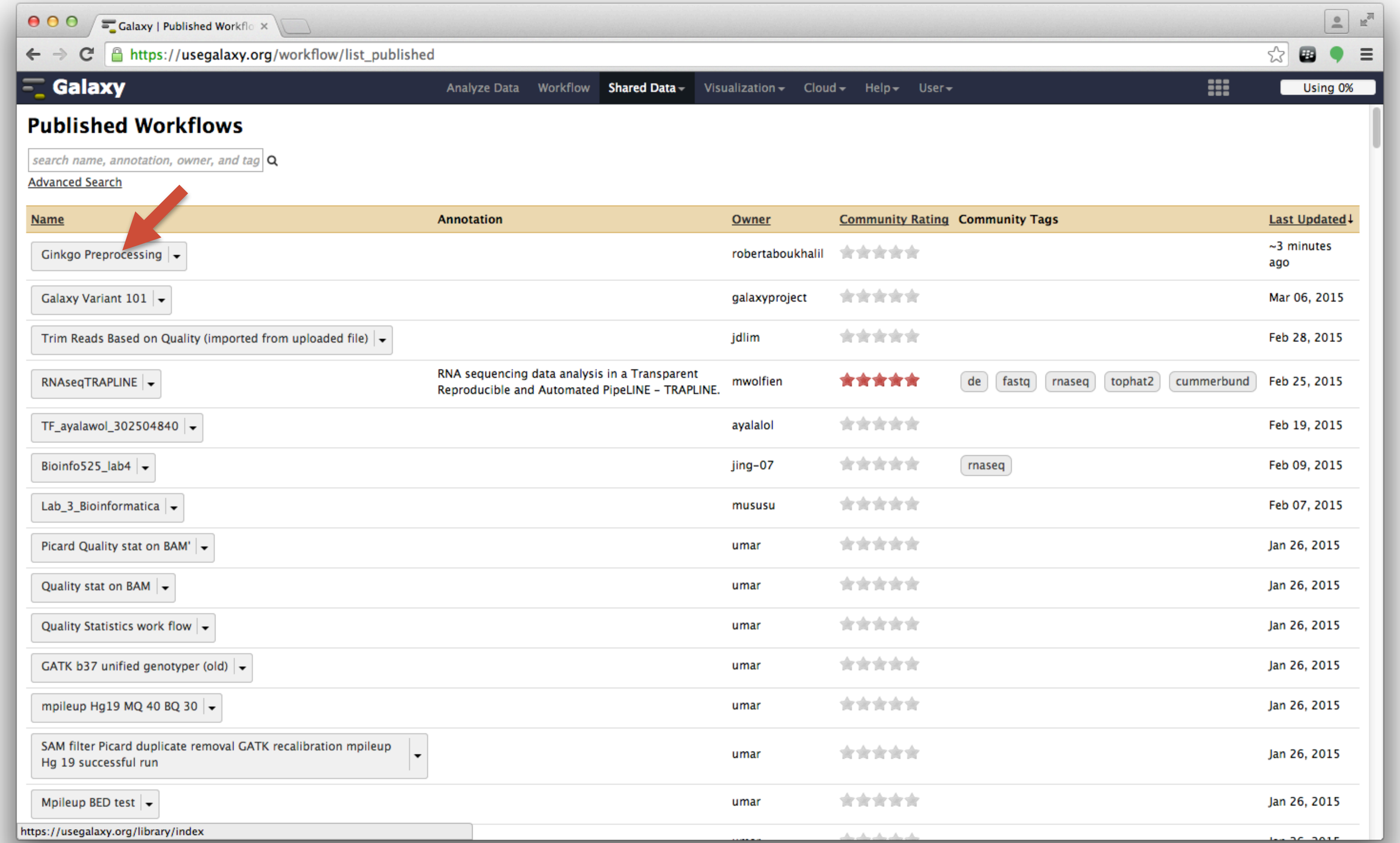
The Galaxy Team is a part of the Center for Comparative Genomics and Bioinformatics at Penn State, and the Department of Biology and at Johns Hopkins University.

This instance of Galaxy is utilizing infrastructure generously provided by the iPlant Collaborative at the Texas Advanced Computing Center, with support from the National Science Foundation.

Unnamed history
2 shown, 3 hidden
721.5 MB

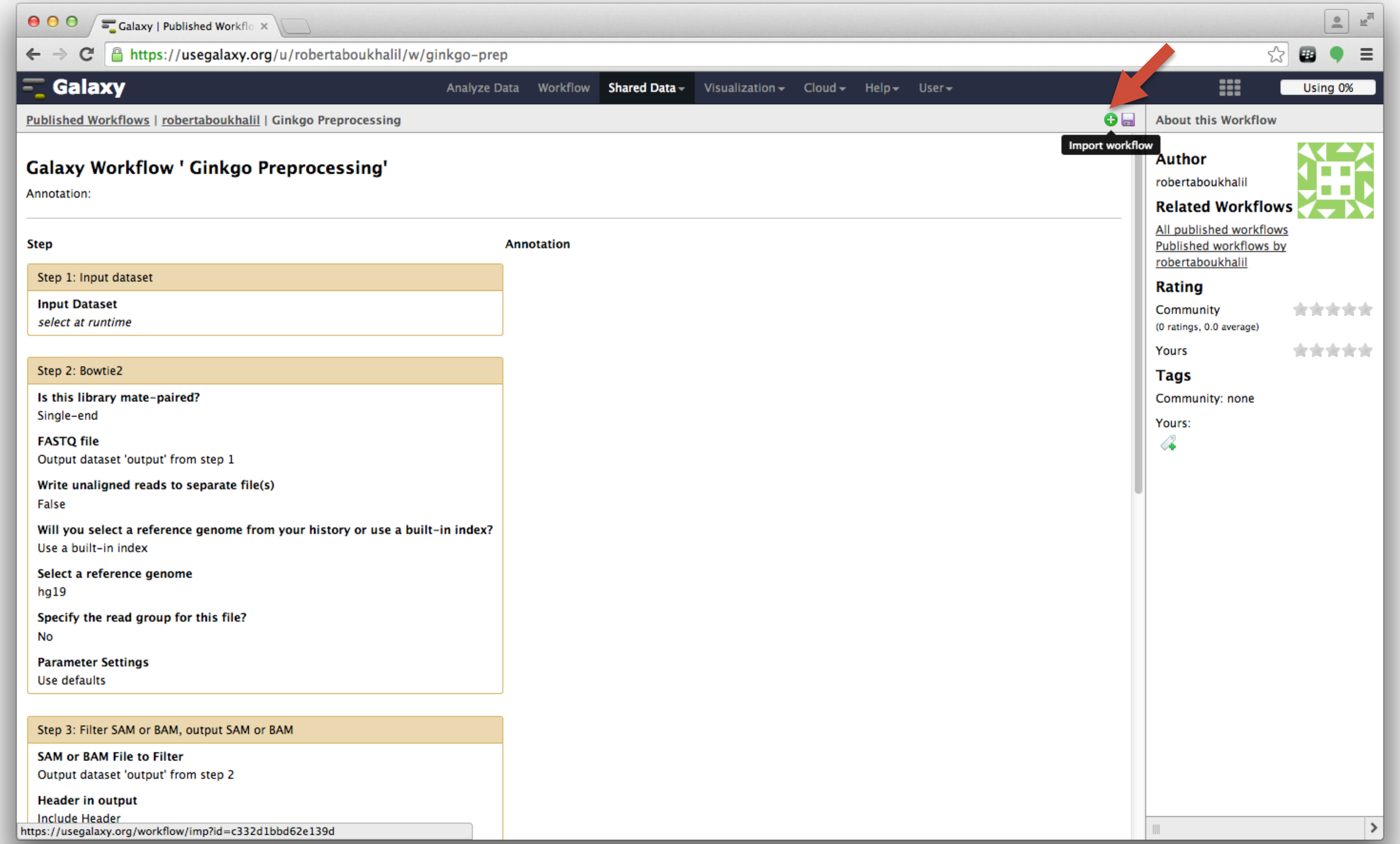
5: rmdup on data 3 (as BED)

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz



Published Workflows

Name	Annotation	Owner	Community Rating	Community Tags	Last Updated↓
Ginkgo Preprocessing		robertaboukhalil	☆☆☆☆☆		~3 minutes ago
Galaxy Variant 101		galaxyproject	☆☆☆☆☆		Mar 06, 2015
Trim Reads Based on Quality (imported from uploaded file)		jdlim	☆☆☆☆☆		Feb 28, 2015
RNAseqTRAPLINE	RNA sequencing data analysis in a Transparent Reproducible and Automated PipeLINE - TRAPLINE.	mwolfien	★★★★★	de fastq rnaseq tophat2 cummerbund	Feb 25, 2015
TF_ayalawol_302504840		ayalalol	☆☆☆☆☆		Feb 19, 2015
Bioinfo525_lab4		jing-07	☆☆☆☆☆	rnaseq	Feb 09, 2015
Lab_3_Bioinformatica		mususu	☆☆☆☆☆		Feb 07, 2015
Picard Quality stat on BAM'		umar	☆☆☆☆☆		Jan 26, 2015
Quality stat on BAM		umar	☆☆☆☆☆		Jan 26, 2015
Quality Statistics work flow		umar	☆☆☆☆☆		Jan 26, 2015
GATK b37 unified genotyper (old)		umar	☆☆☆☆☆		Jan 26, 2015
mpileup Hg19 MQ 40 BQ 30		umar	☆☆☆☆☆		Jan 26, 2015
SAM filter Picard duplicate removal GATK recalibration mpileup Hg 19 successful run		umar	☆☆☆☆☆		Jan 26, 2015
Mpileup BED test		umar	☆☆☆☆☆		Jan 26, 2015



Published Workflows | robertaboukhalil | Ginkgo Preprocessing

Galaxy Workflow ' Ginkgo Preprocessing'

Annotation:

Step	Annotation
Step 1: Input dataset **Input Dataset** *select at runtime*	
Step 2: Bowtie2 **Is this library mate-paired?** Single-end **FASTQ file** Output dataset 'output' from step 1 **Write unaligned reads to separate file(s)** False **Will you select a reference genome from your history or use a built-in index?** Use a built-in index **Select a reference genome** hg19 **Specify the read group for this file?** No **Parameter Settings** Use defaults	
Step 3: Filter SAM or BAM, output SAM or BAM **SAM or BAM File to Filter** Output dataset 'output' from step 2 **Header in output** Include Header	

https://usegalaxy.org/workflow/imp?id=c332d1bbd62e139d

Author

robertaboukhalil

Related Workflows

All published workflows

Published workflows by robertaboukhalil

Rating

Community

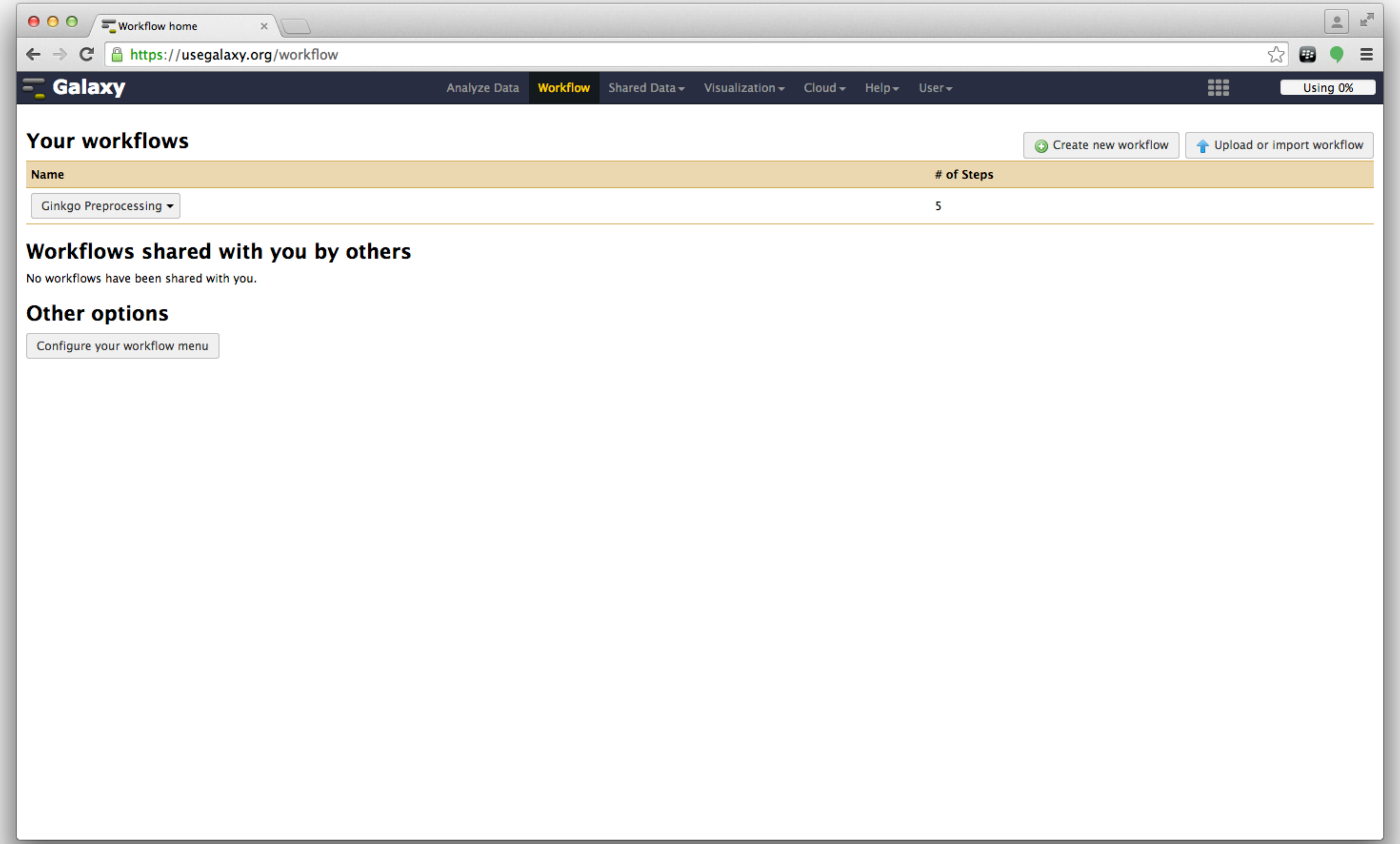
(0 ratings, 0.0 average)

Yours

Tags

Community: none

Yours:



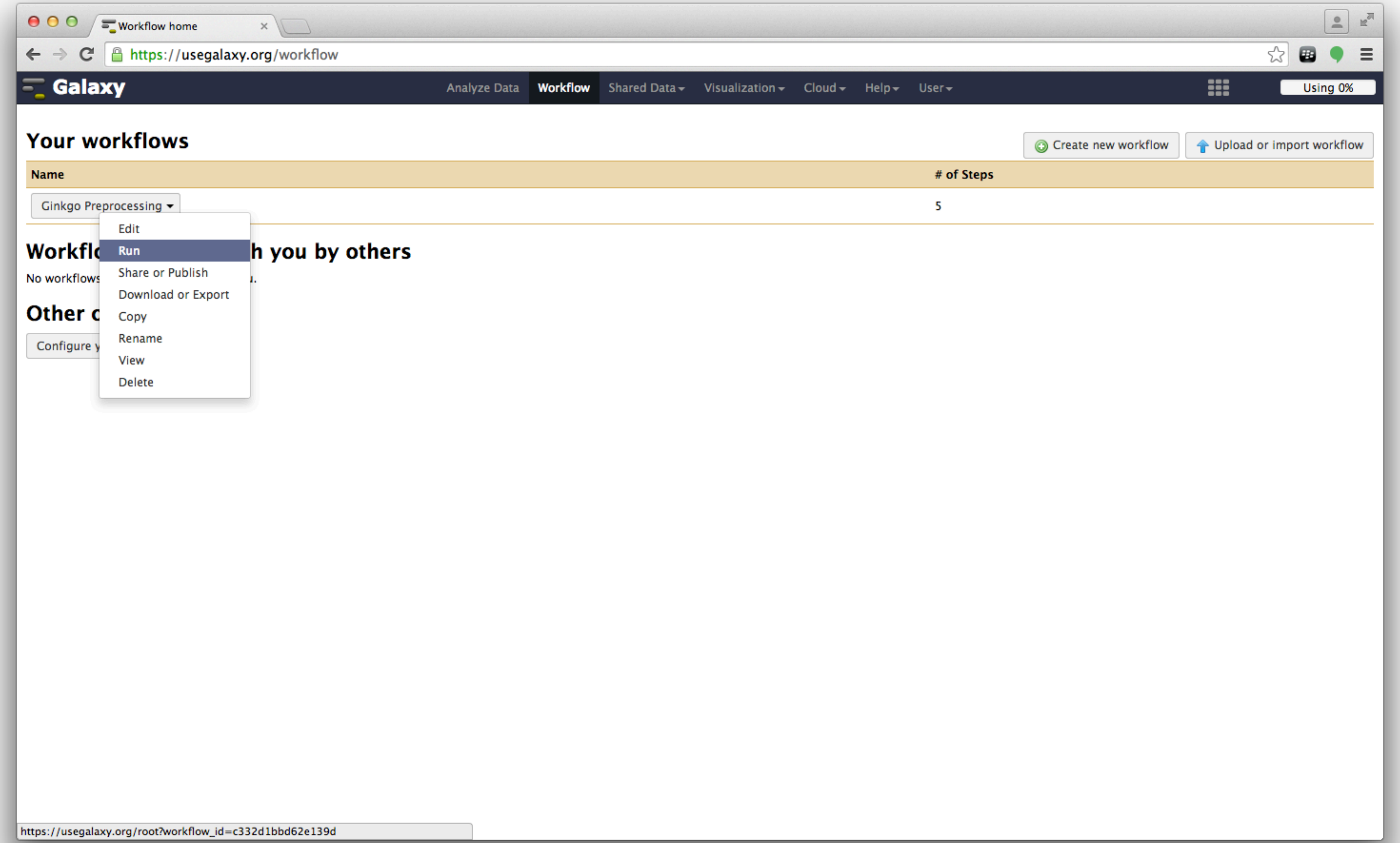
Your workflows

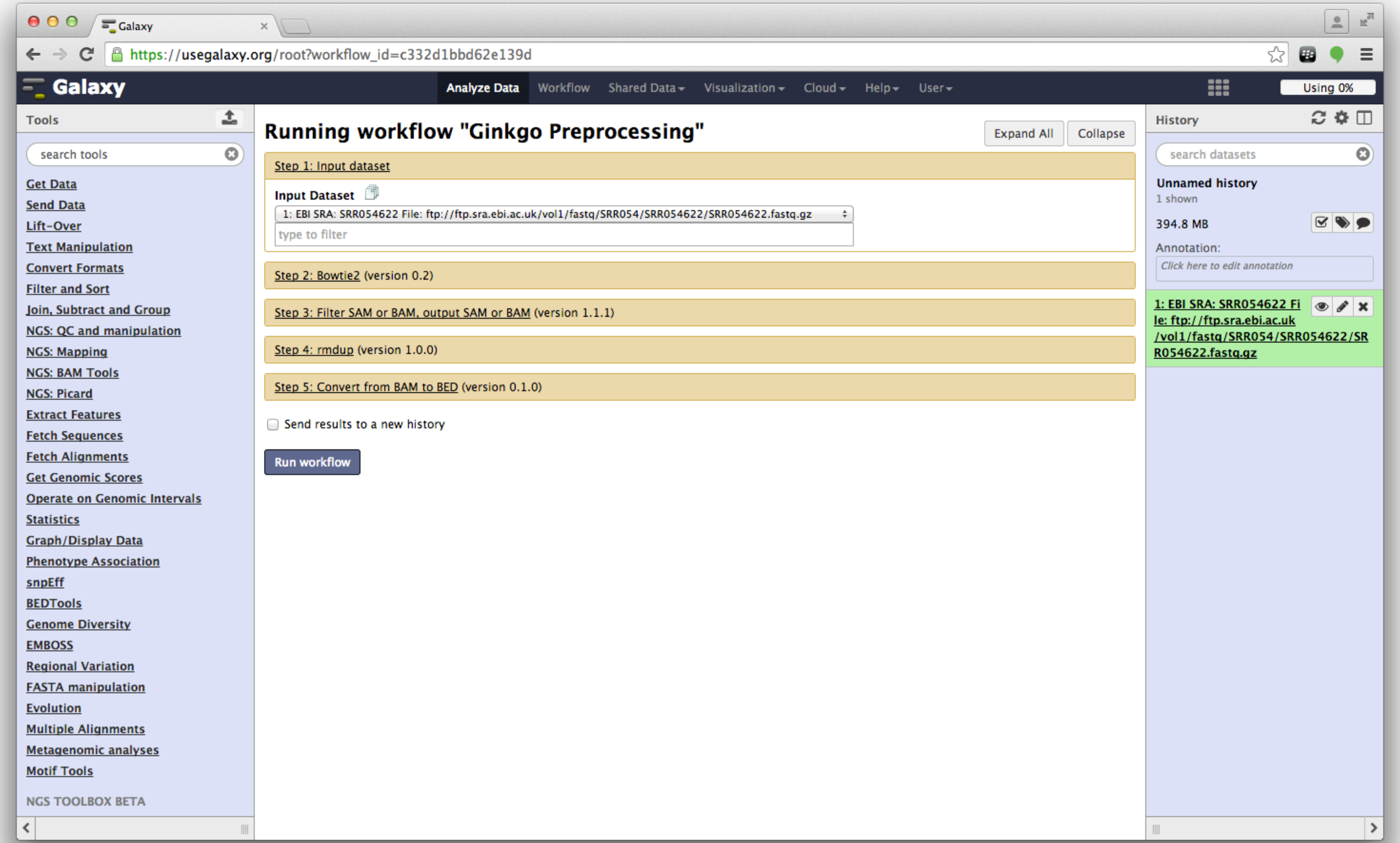
Name	# of Steps
Ginkgo Preprocessing	5

Workflows shared with you by others

No workflows have been shared with you.

Other options


Workflow home
https://usegalaxy.org/workflow
Galaxy
Analyze Data
Workflow
Shared Data
Visualization
Cloud
Help
User
Using 0%
Your workflows
Create new workflow
Upload or import workflow
Name
# of Steps
Ginkgo Preprocessing
5
Edit
Run
Share or Publish
Download or Export
Copy
Rename
View
Delete
Workfl h you by others
No workflows .
Other c
Configure y
https://usegalaxy.org/root?workflow_id=c332d1bbd62e139d



Running workflow "Ginkgo Preprocessing"

Step 1: Input dataset

Input Dataset: 1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

type to filter

Step 2: Bowtie2 (version 0.2)

Step 3: Filter SAM or BAM, output SAM or BAM (version 1.1.1)

Step 4: rmdup (version 1.0.0)

Step 5: Convert from BAM to BED (version 0.1.0)

☐ Send results to a new history

Unnamed history

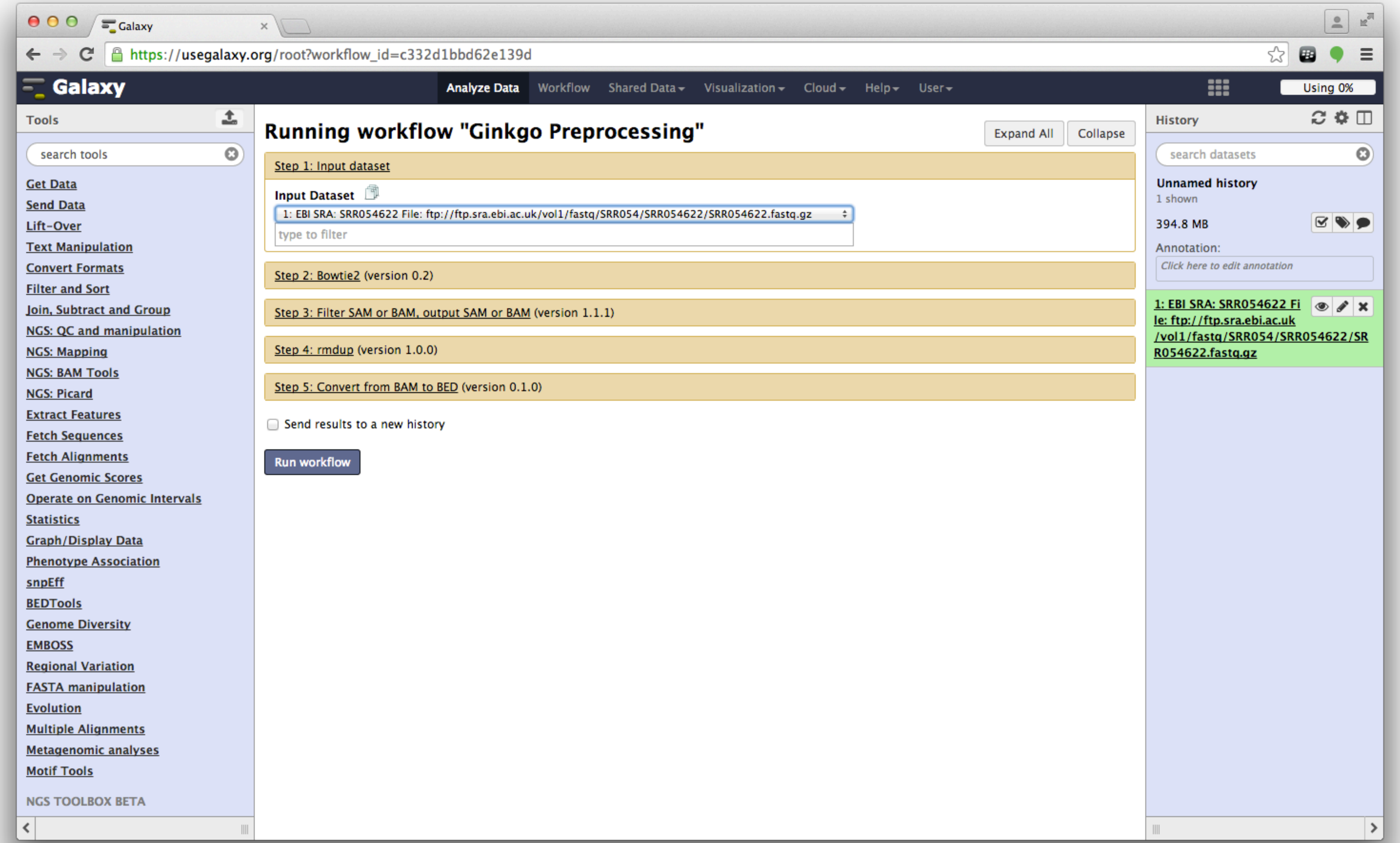
1 shown

394.8 MB

Annotation:

Click here to edit annotation

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz



Running workflow "Ginkgo Preprocessing"

Step 1: Input dataset

Input Dataset: 1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

type to filter

Step 2: Bowtie2 (version 0.2)

Step 3: Filter SAM or BAM, output SAM or BAM (version 1.1.1)

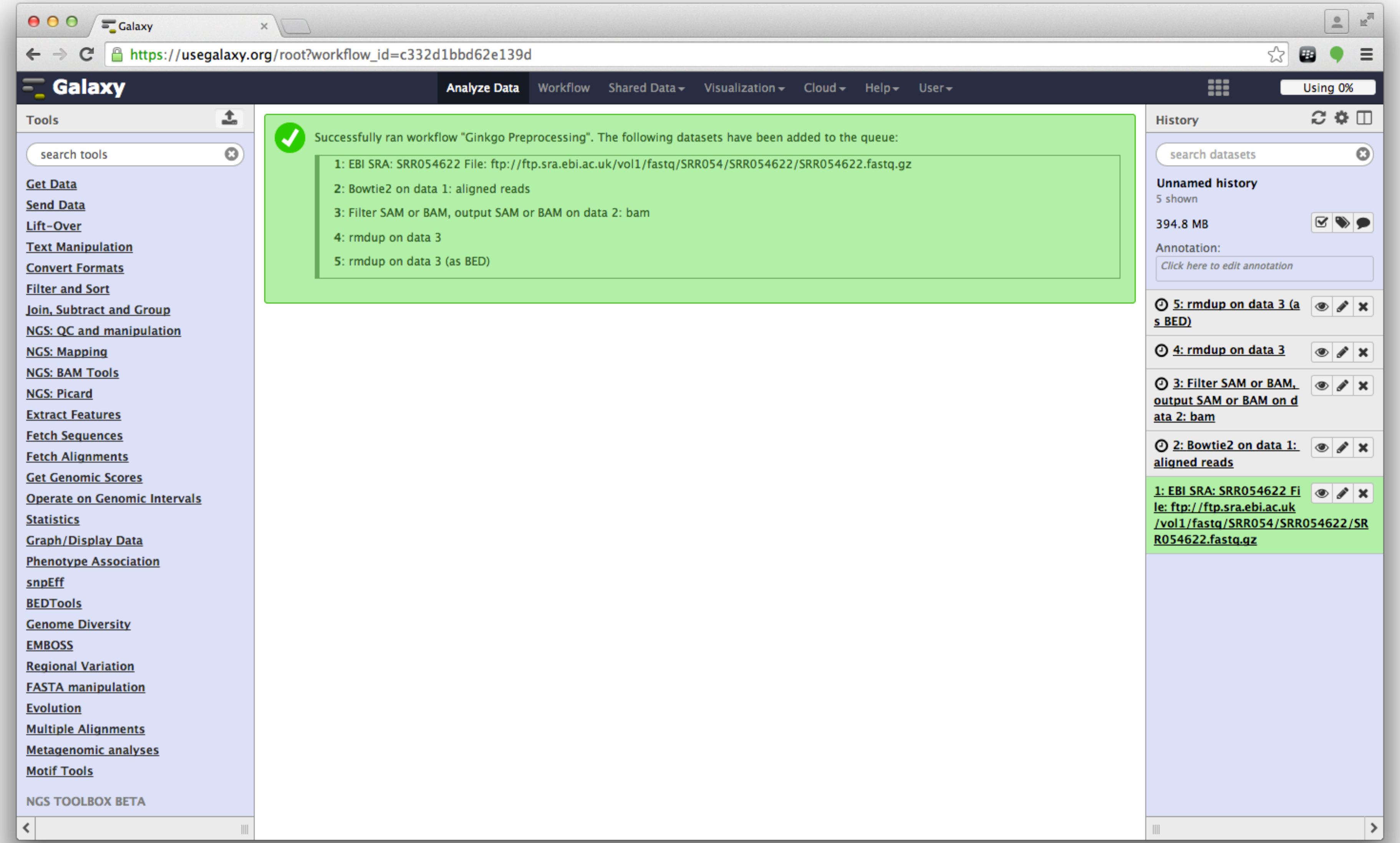
Step 4: rmdup (version 1.0.0)

Step 5: Convert from BAM to BED (version 0.1.0)

☐ Send results to a new history

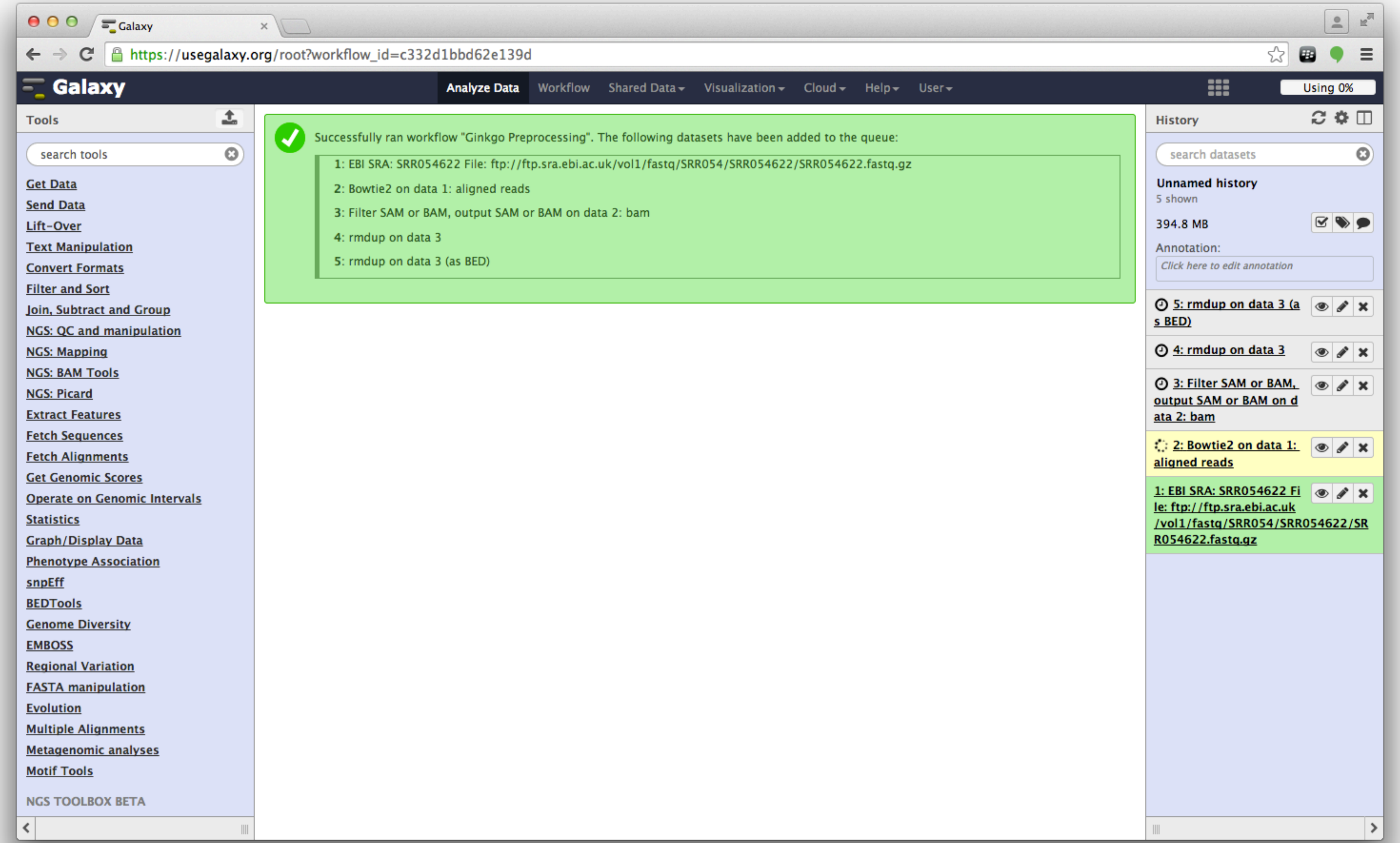
Unnamed history
1 shown
394.8 MB
Annotation:

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz



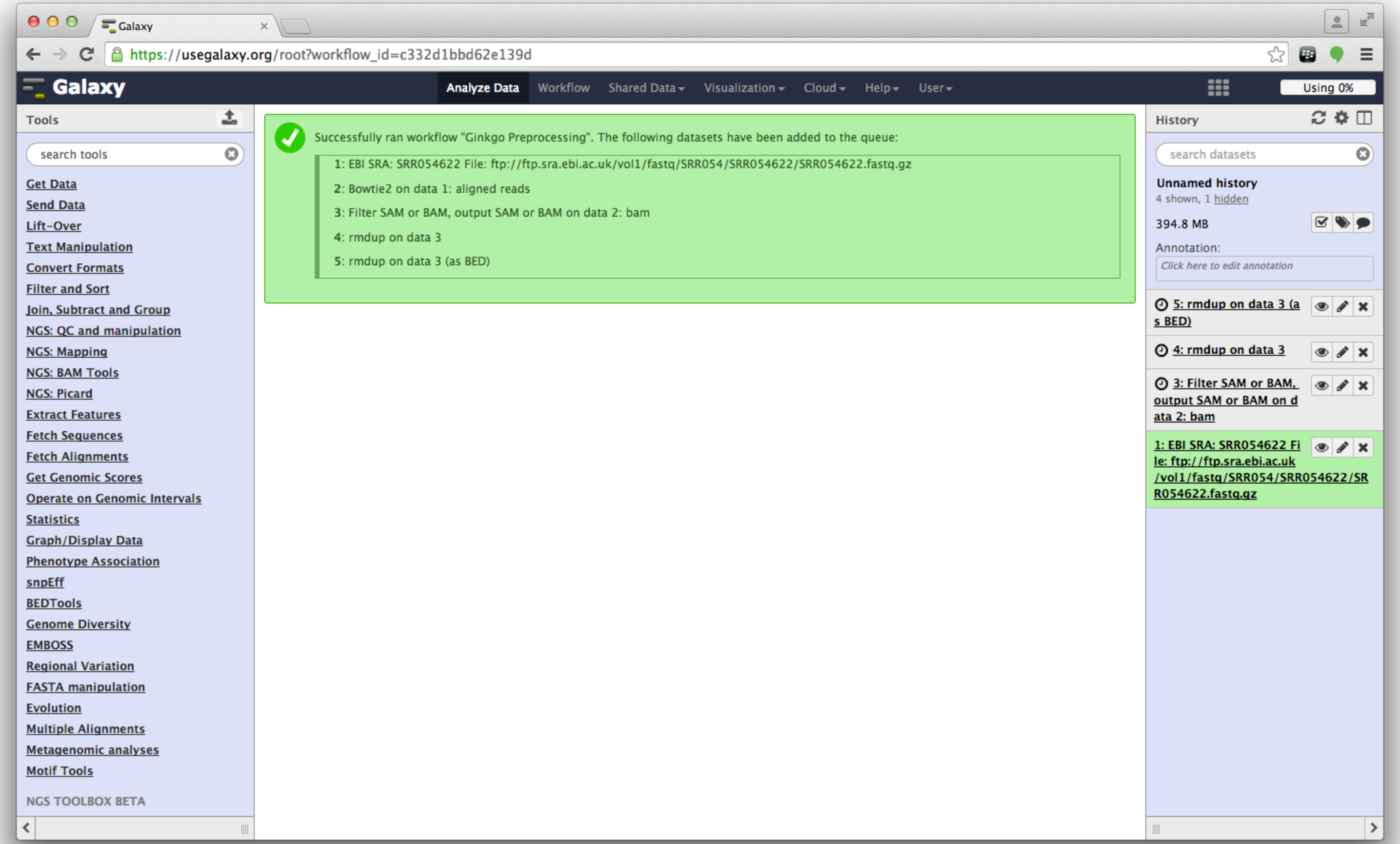
Successfully ran workflow "Ginkgo Preprocessing". The following datasets have been added to the queue:

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz
2: Bowtie2 on data 1: aligned reads
3: Filter SAM or BAM, output SAM or BAM on data 2: bam
4: rmdup on data 3
5: rmdup on data 3 (as BED)



Successfully ran workflow "Ginkgo Preprocessing". The following datasets have been added to the queue:

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz
2: Bowtie2 on data 1: aligned reads
3: Filter SAM or BAM, output SAM or BAM on data 2: bam
4: rmdup on data 3
5: rmdup on data 3 (as BED)



Successfully ran workflow "Ginkgo Preprocessing". The following datasets have been added to the queue:

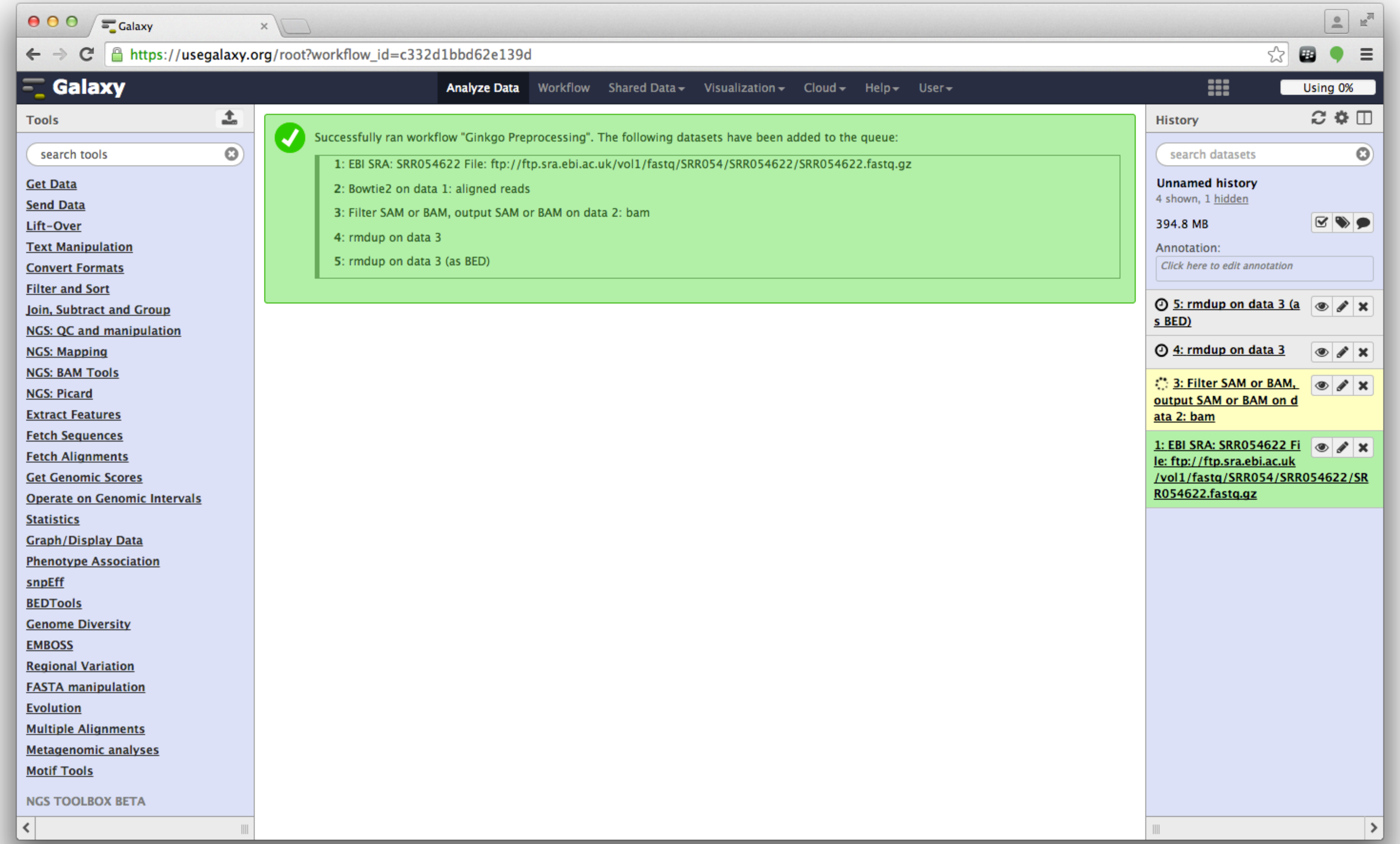
1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

2: Bowtie2 on data 1: aligned reads

3: Filter SAM or BAM, output SAM or BAM on data 2: bam

4: rmdup on data 3

5: rmdup on data 3 (as BED)



Successfully ran workflow "Ginkgo Preprocessing". The following datasets have been added to the queue:

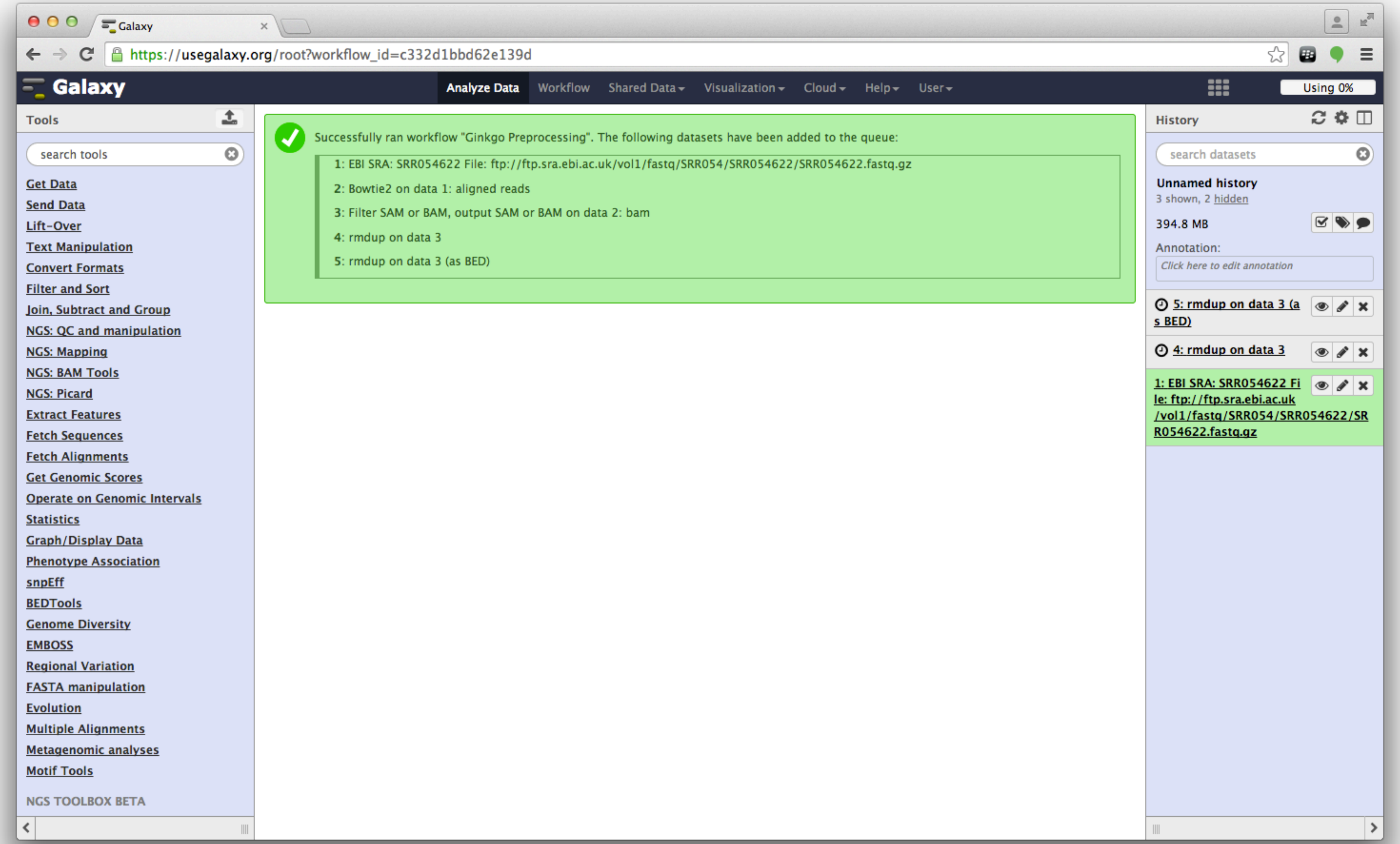
1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

2: Bowtie2 on data 1: aligned reads

3: Filter SAM or BAM, output SAM or BAM on data 2: bam

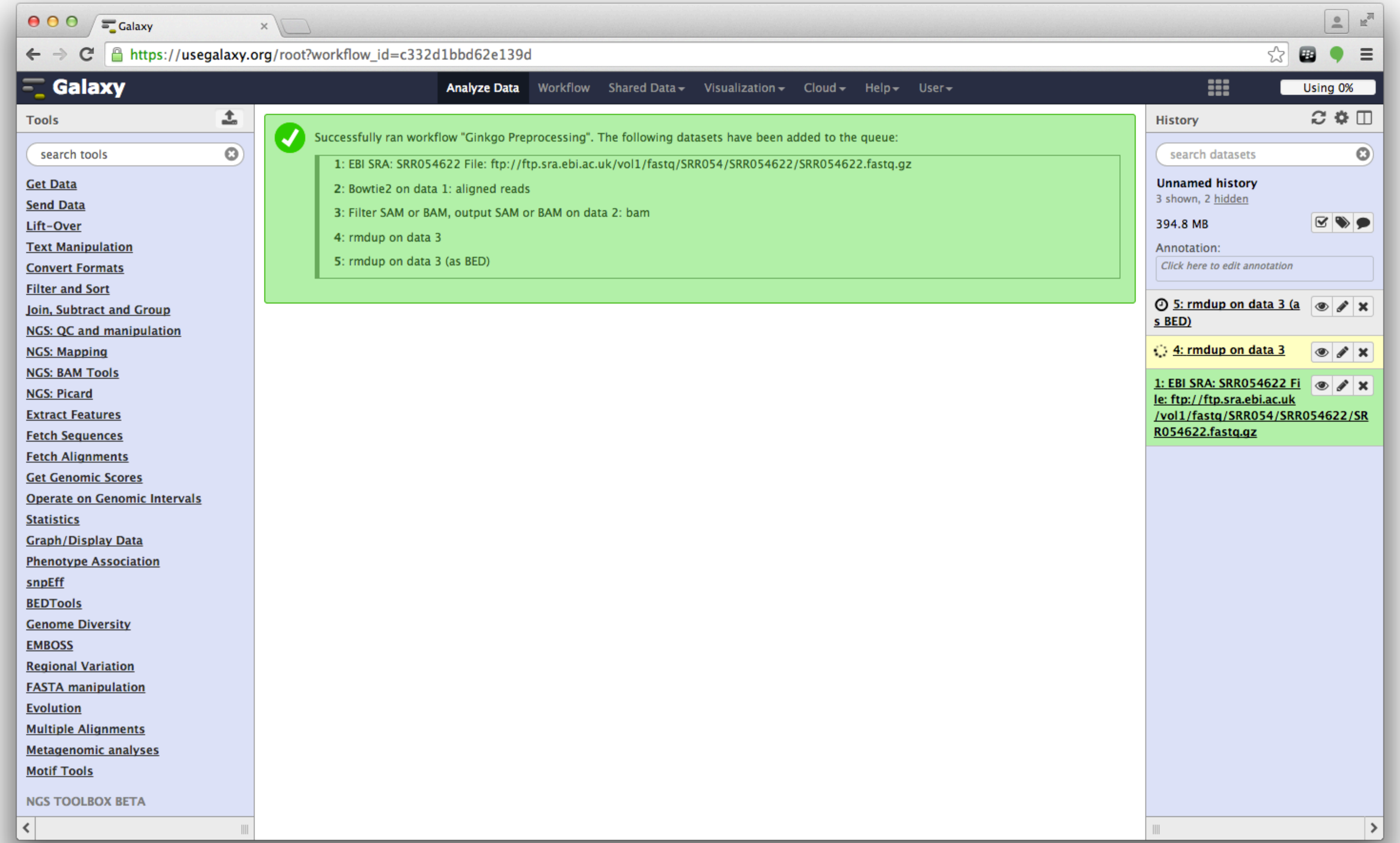
4: rmdup on data 3

5: rmdup on data 3 (as BED)



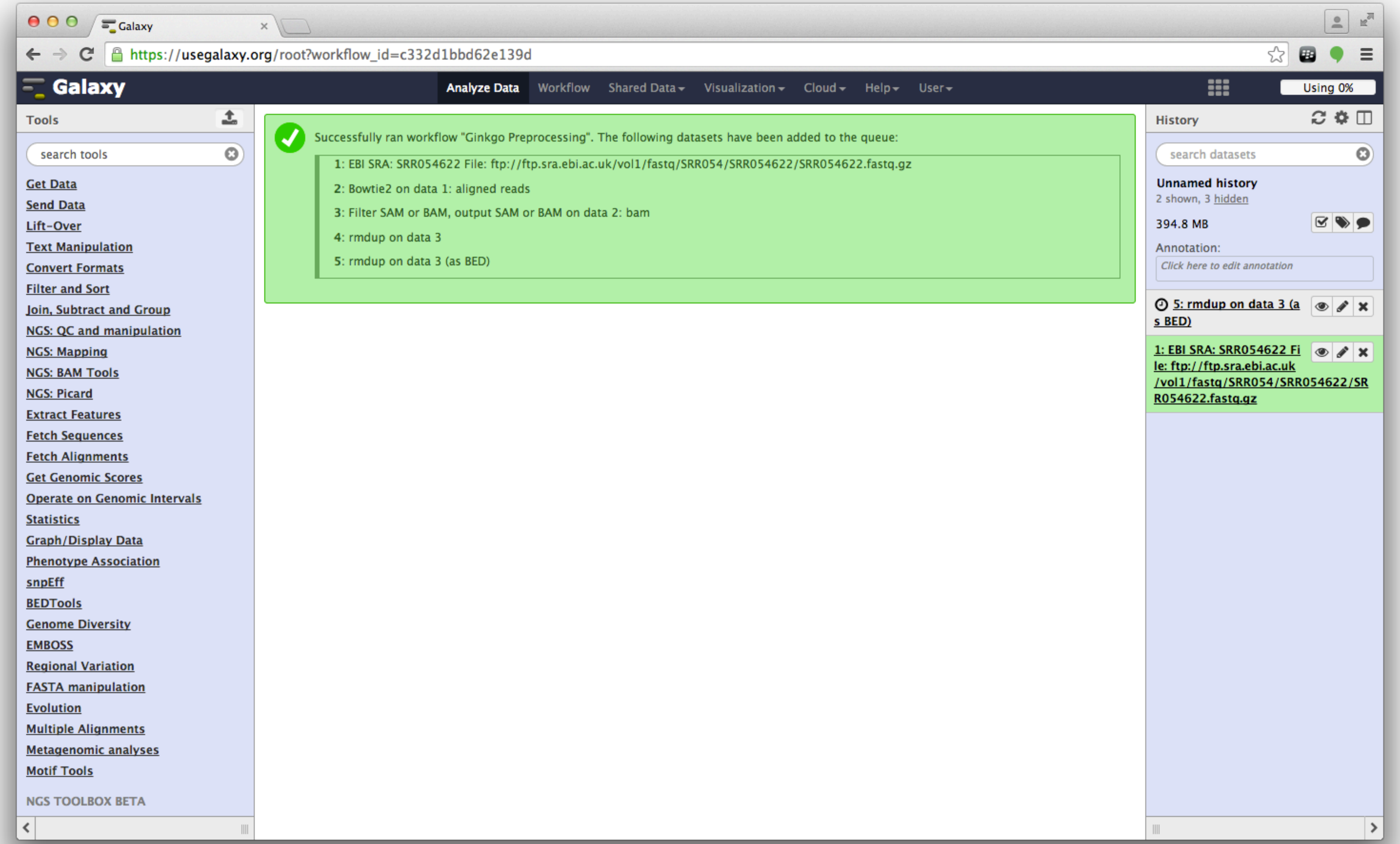
Successfully ran workflow "Ginkgo Preprocessing". The following datasets have been added to the queue:

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz
2: Bowtie2 on data 1: aligned reads
3: Filter SAM or BAM, output SAM or BAM on data 2: bam
4: rmdup on data 3
5: rmdup on data 3 (as BED)



Successfully ran workflow "Ginkgo Preprocessing". The following datasets have been added to the queue:

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz
2: Bowtie2 on data 1: aligned reads
3: Filter SAM or BAM, output SAM or BAM on data 2: bam
4: rmdup on data 3
5: rmdup on data 3 (as BED)



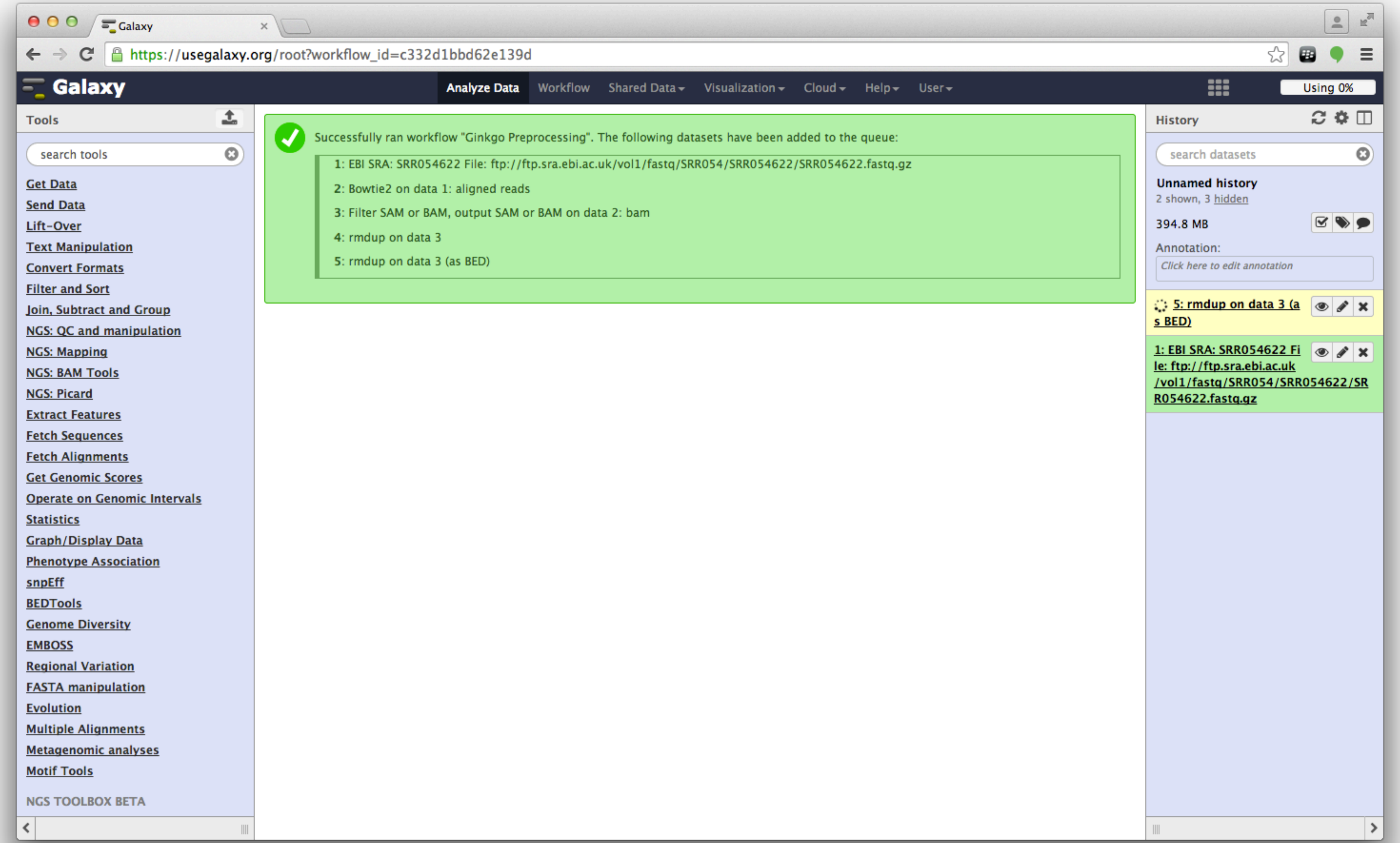
Successfully ran workflow "Ginkgo Preprocessing". The following datasets have been added to the queue:

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz
2: Bowtie2 on data 1: aligned reads
3: Filter SAM or BAM, output SAM or BAM on data 2: bam
4: rmdup on data 3
5: rmdup on data 3 (as BED)

Unnamed history
2 shown, 3 hidden
394.8 MB

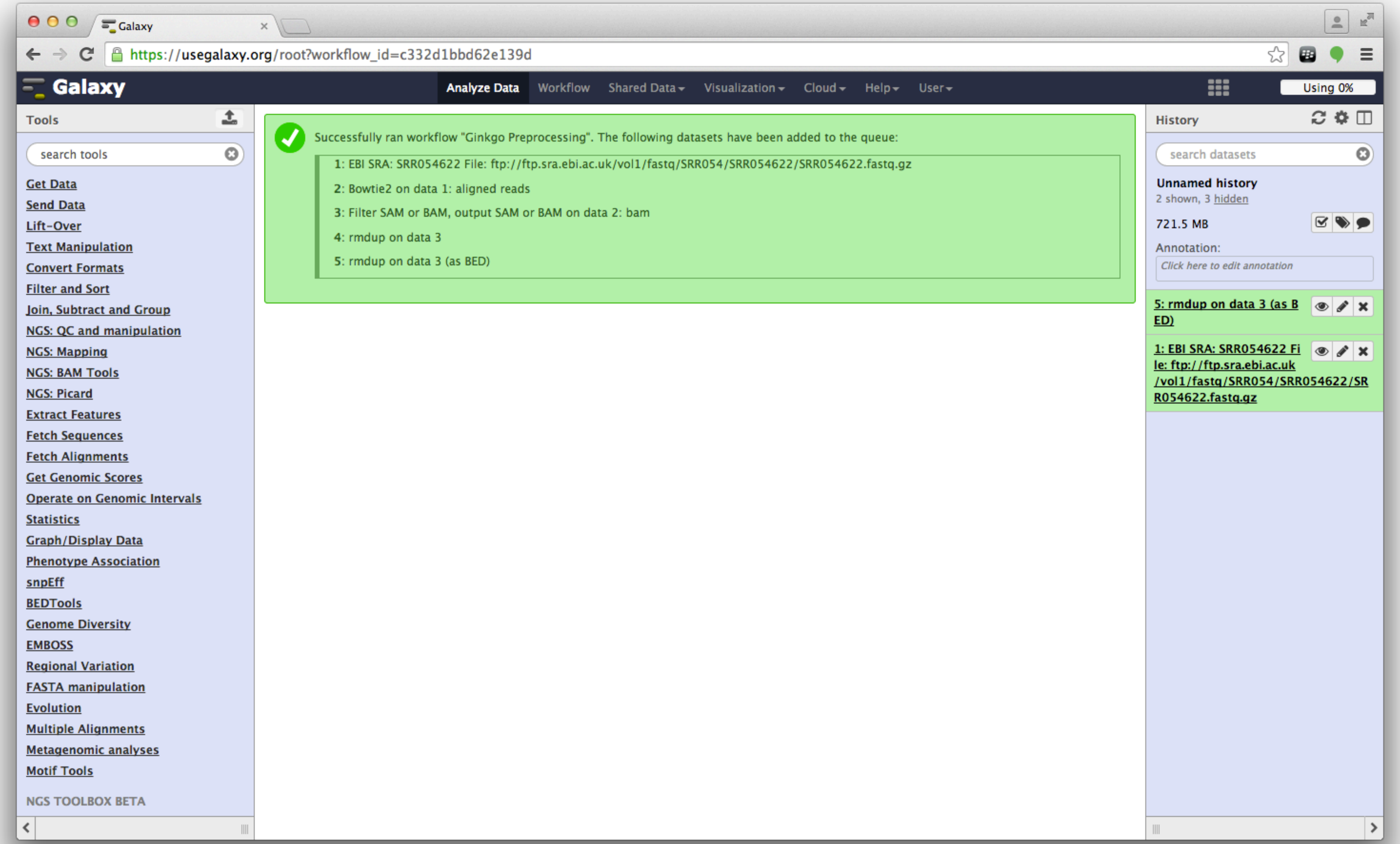
5: rmdup on data 3 (as BED)

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz



Successfully ran workflow "Ginkgo Preprocessing". The following datasets have been added to the queue:

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz
2: Bowtie2 on data 1: aligned reads
3: Filter SAM or BAM, output SAM or BAM on data 2: bam
4: rmdup on data 3
5: rmdup on data 3 (as BED)



Successfully ran workflow "Ginkgo Preprocessing". The following datasets have been added to the queue:

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

2: Bowtie2 on data 1: aligned reads

3: Filter SAM or BAM, output SAM or BAM on data 2: bam

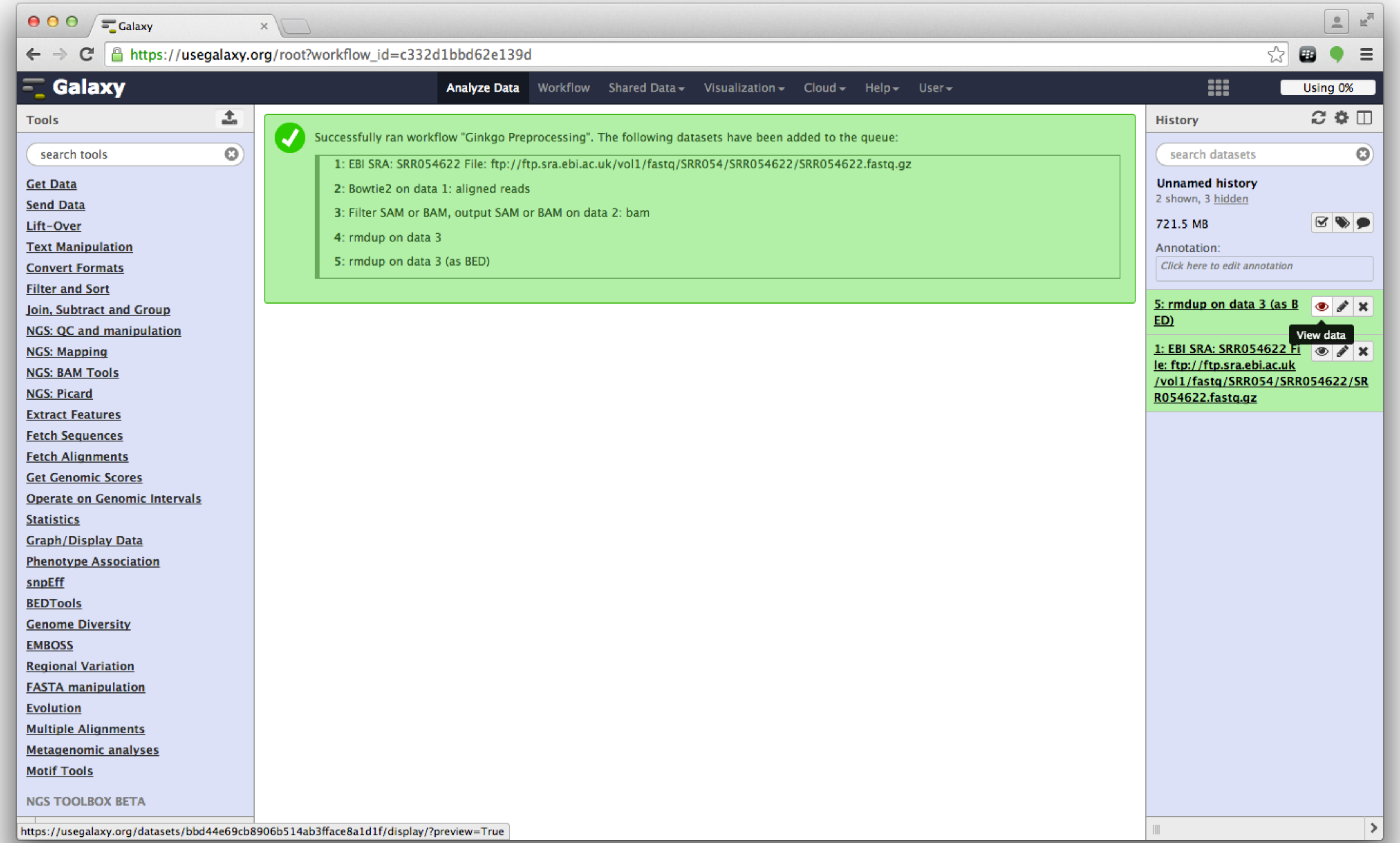
4: rmdup on data 3

5: rmdup on data 3 (as BED)

Unnamed history
2 shown, 3 hidden
721.5 MB

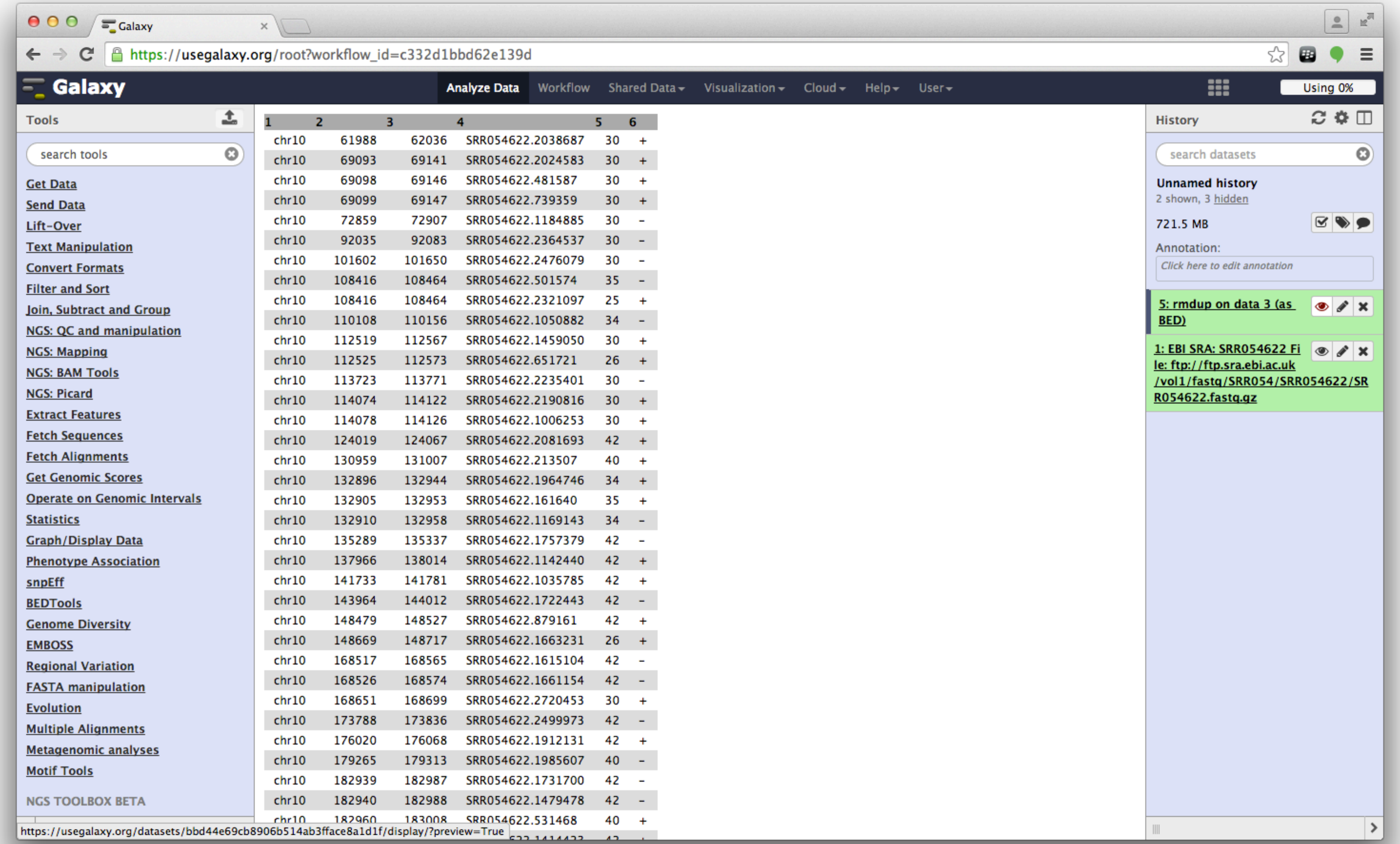
5: rmdup on data 3 (as BED)

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

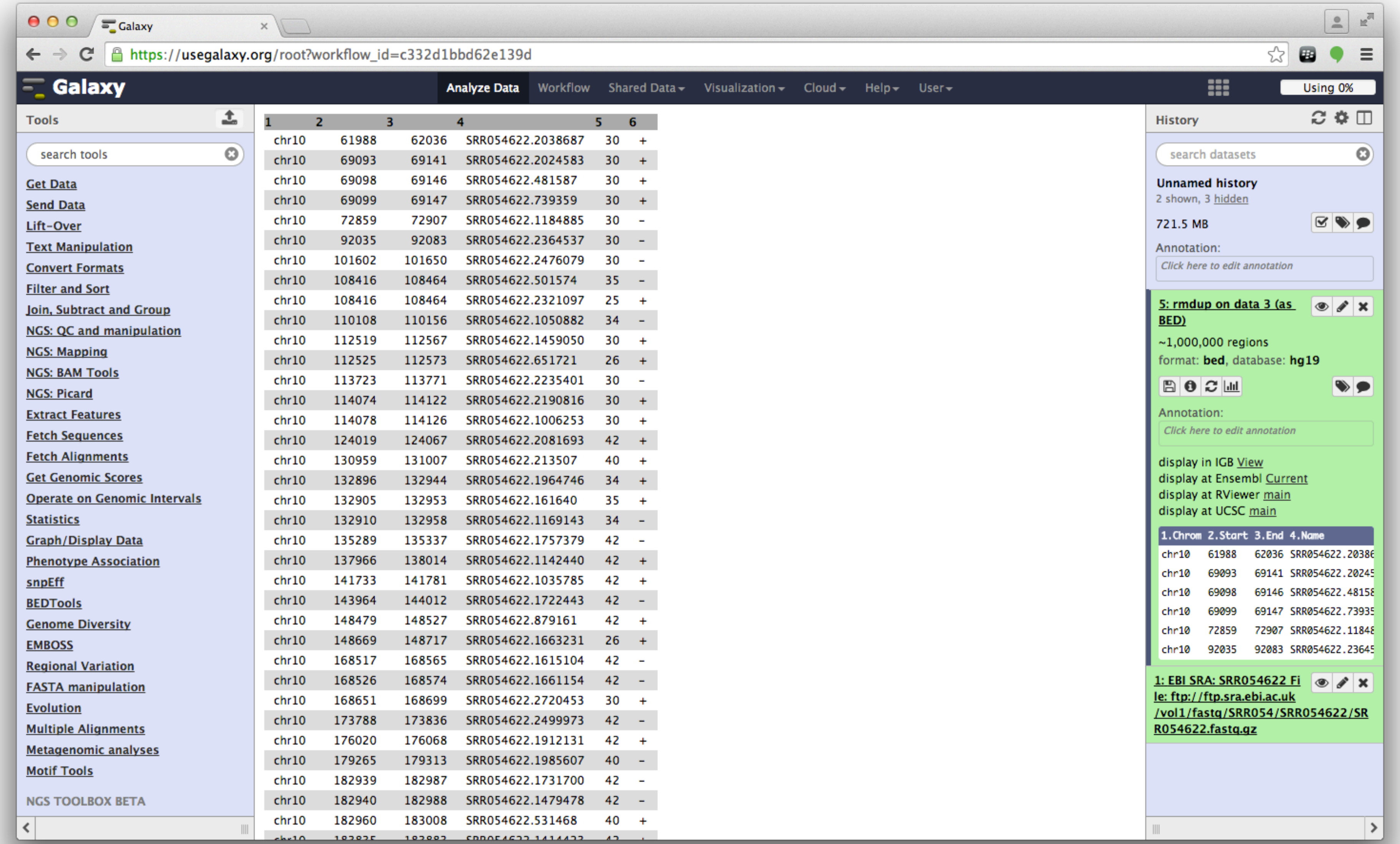


Successfully ran workflow "Ginkgo Preprocessing". The following datasets have been added to the queue:

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz
2: Bowtie2 on data 1: aligned reads
3: Filter SAM or BAM, output SAM or BAM on data 2: bam
4: rmdup on data 3
5: rmdup on data 3 (as BED)



1	2	3	4	5	6
chr10	61988	62036	SRR054622.2038687	30	+
chr10	69093	69141	SRR054622.2024583	30	+
chr10	69098	69146	SRR054622.481587	30	+
chr10	69099	69147	SRR054622.739359	30	+
chr10	72859	72907	SRR054622.1184885	30	-
chr10	92035	92083	SRR054622.2364537	30	-
chr10	101602	101650	SRR054622.2476079	30	-
chr10	108416	108464	SRR054622.501574	35	-
chr10	108416	108464	SRR054622.2321097	25	+
chr10	110108	110156	SRR054622.1050882	34	-
chr10	112519	112567	SRR054622.1459050	30	+
chr10	112525	112573	SRR054622.651721	26	+
chr10	113723	113771	SRR054622.2235401	30	-
chr10	114074	114122	SRR054622.2190816	30	+
chr10	114078	114126	SRR054622.1006253	30	+
chr10	124019	124067	SRR054622.2081693	42	+
chr10	130959	131007	SRR054622.213507	40	+
chr10	132896	132944	SRR054622.1964746	34	+
chr10	132905	132953	SRR054622.161640	35	+
chr10	132910	132958	SRR054622.1169143	34	-
chr10	135289	135337	SRR054622.1757379	42	-
chr10	137966	138014	SRR054622.1142440	42	+
chr10	141733	141781	SRR054622.1035785	42	+
chr10	143964	144012	SRR054622.1722443	42	-
chr10	148479	148527	SRR054622.879161	42	+
chr10	148669	148717	SRR054622.1663231	26	+
chr10	168517	168565	SRR054622.1615104	42	-
chr10	168526	168574	SRR054622.1661154	42	-
chr10	168651	168699	SRR054622.2720453	30	+
chr10	173788	173836	SRR054622.2499973	42	-
chr10	176020	176068	SRR054622.1912131	42	+
chr10	179265	179313	SRR054622.1985607	40	-
chr10	182939	182987	SRR054622.1731700	42	-
chr10	182940	182988	SRR054622.1479478	42	-
chr10	182960	183008	SRR054622.531468	40	+



1	2	3	4	5	6
chr10	61988	62036	SRR054622.2038687	30	+
chr10	69093	69141	SRR054622.2024583	30	+
chr10	69098	69146	SRR054622.481587	30	+
chr10	69099	69147	SRR054622.739359	30	+
chr10	72859	72907	SRR054622.1184885	30	-
chr10	92035	92083	SRR054622.2364537	30	-
chr10	101602	101650	SRR054622.2476079	30	-
chr10	108416	108464	SRR054622.501574	35	-
chr10	108416	108464	SRR054622.2321097	25	+
chr10	110108	110156	SRR054622.1050882	34	-
chr10	112519	112567	SRR054622.1459050	30	+
chr10	112525	112573	SRR054622.651721	26	+
chr10	113723	113771	SRR054622.2235401	30	-
chr10	114074	114122	SRR054622.2190816	30	+
chr10	114078	114126	SRR054622.1006253	30	+
chr10	124019	124067	SRR054622.2081693	42	+
chr10	130959	131007	SRR054622.213507	40	+
chr10	132896	132944	SRR054622.1964746	34	+
chr10	132905	132953	SRR054622.161640	35	+
chr10	132910	132958	SRR054622.1169143	34	-
chr10	135289	135337	SRR054622.1757379	42	-
chr10	137966	138014	SRR054622.1142440	42	+
chr10	141733	141781	SRR054622.1035785	42	+
chr10	143964	144012	SRR054622.1722443	42	-
chr10	148479	148527	SRR054622.879161	42	+
chr10	148669	148717	SRR054622.1663231	26	+
chr10	168517	168565	SRR054622.1615104	42	-
chr10	168526	168574	SRR054622.1661154	42	-
chr10	168651	168699	SRR054622.2720453	30	+
chr10	173788	173836	SRR054622.2499973	42	-
chr10	176020	176068	SRR054622.1912131	42	+
chr10	179265	179313	SRR054622.1985607	40	-
chr10	182939	182987	SRR054622.1731700	42	-
chr10	182940	182988	SRR054622.1479478	42	-
chr10	182960	183008	SRR054622.531468	40	+

Unnamed history
2 shown, 3 hidden

721.5 MB

Annotation:
Click here to edit annotation

5: rmdup on data 3 (as BED)

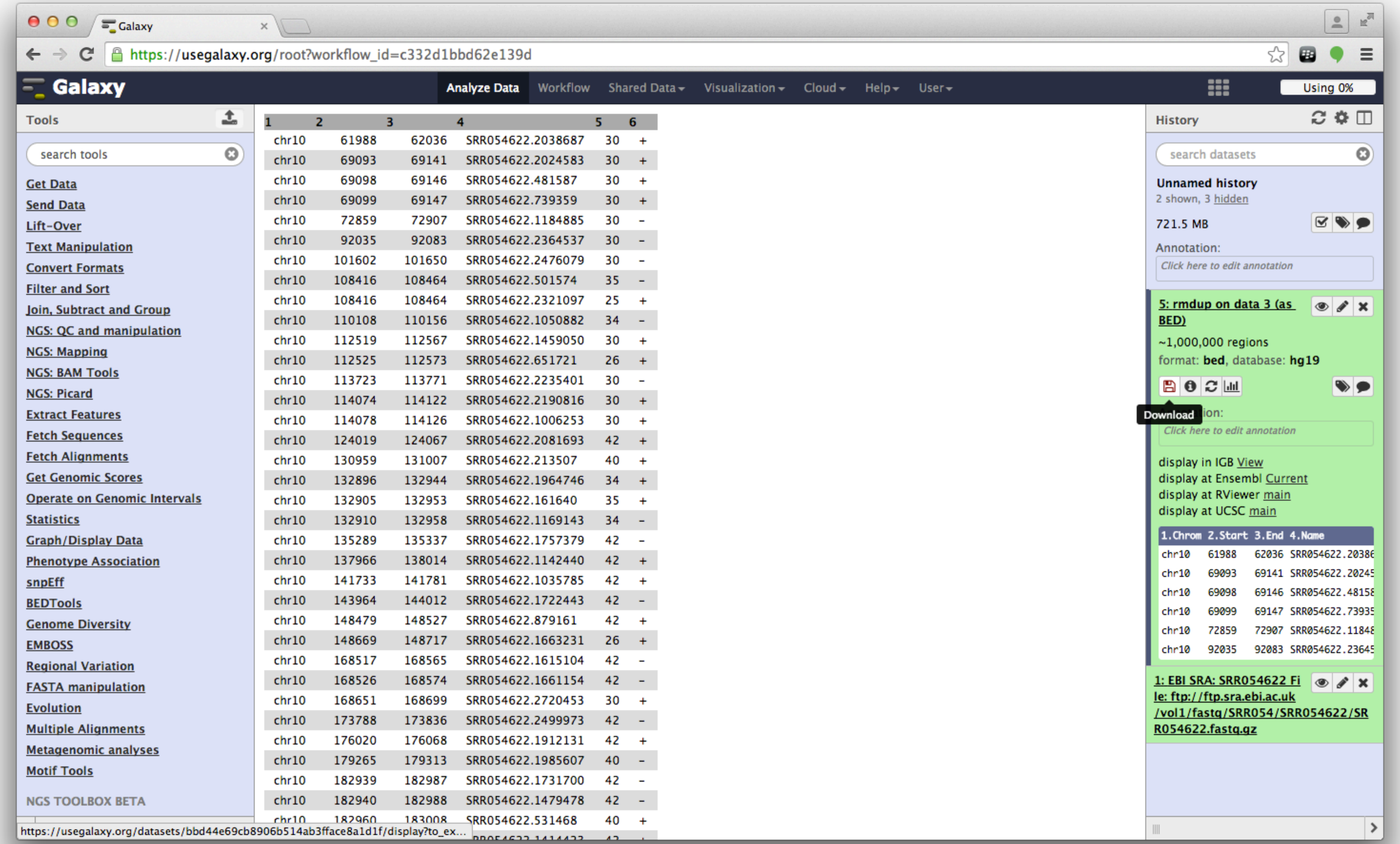
~1,000,000 regions
format: **bed**, database: **hg19**

Annotation:
Click here to edit annotation

display in IGB View
display at Ensembl Current
display at RViewer main
display at UCSC main

1.Chrom	2.Start	3.End	4.Name
chr10	61988	62036	SRR054622.2038
chr10	69093	69141	SRR054622.2024
chr10	69098	69146	SRR054622.4815
chr10	69099	69147	SRR054622.7393
chr10	72859	72907	SRR054622.1184
chr10	92035	92083	SRR054622.2364

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz



1	2	3	4	5	6
chr10	61988	62036	SRR054622.2038687	30	+
chr10	69093	69141	SRR054622.2024583	30	+
chr10	69098	69146	SRR054622.481587	30	+
chr10	69099	69147	SRR054622.739359	30	+
chr10	72859	72907	SRR054622.1184885	30	-
chr10	92035	92083	SRR054622.2364537	30	-
chr10	101602	101650	SRR054622.2476079	30	-
chr10	108416	108464	SRR054622.501574	35	-
chr10	108416	108464	SRR054622.2321097	25	+
chr10	110108	110156	SRR054622.1050882	34	-
chr10	112519	112567	SRR054622.1459050	30	+
chr10	112525	112573	SRR054622.651721	26	+
chr10	113723	113771	SRR054622.2235401	30	-
chr10	114074	114122	SRR054622.2190816	30	+
chr10	114078	114126	SRR054622.1006253	30	+
chr10	124019	124067	SRR054622.2081693	42	+
chr10	130959	131007	SRR054622.213507	40	+
chr10	132896	132944	SRR054622.1964746	34	+
chr10	132905	132953	SRR054622.161640	35	+
chr10	132910	132958	SRR054622.1169143	34	-
chr10	135289	135337	SRR054622.1757379	42	-
chr10	137966	138014	SRR054622.1142440	42	+
chr10	141733	141781	SRR054622.1035785	42	+
chr10	143964	144012	SRR054622.1722443	42	-
chr10	148479	148527	SRR054622.879161	42	+
chr10	148669	148717	SRR054622.1663231	26	+
chr10	168517	168565	SRR054622.1615104	42	-
chr10	168526	168574	SRR054622.1661154	42	-
chr10	168651	168699	SRR054622.2720453	30	+
chr10	173788	173836	SRR054622.2499973	42	-
chr10	176020	176068	SRR054622.1912131	42	+
chr10	179265	179313	SRR054622.1985607	40	-
chr10	182939	182987	SRR054622.1731700	42	-
chr10	182940	182988	SRR054622.1479478	42	-
chr10	182960	183008	SRR054622.531468	40	+

Unnamed history
2 shown, 3 hidden

721.5 MB

Annotation:
Click here to edit annotation

5: rmdup on data 3 (as BED)

~1,000,000 regions
format: **bed**, database: **hg19**

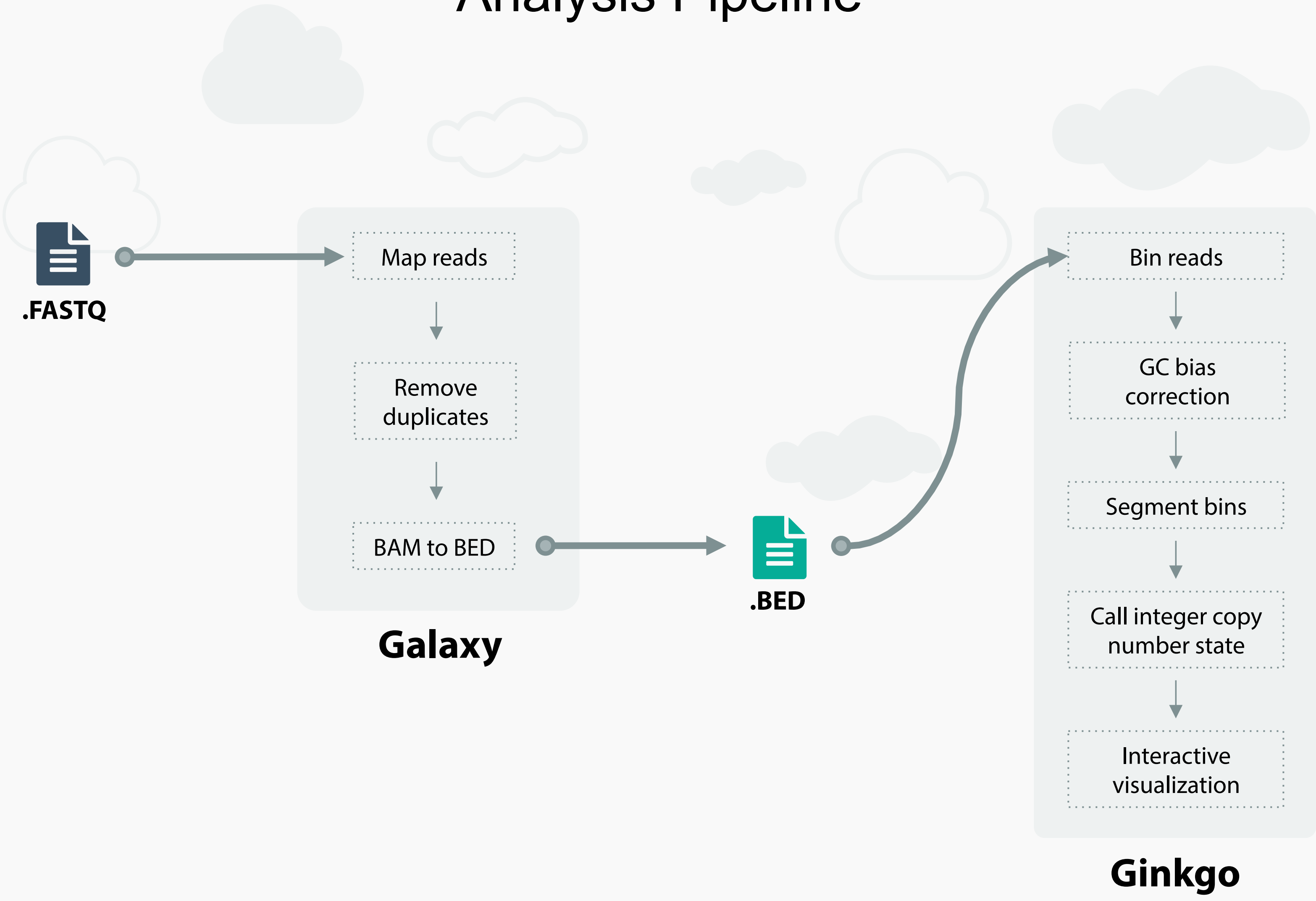
Annotation:
Click here to edit annotation

display in IGB View
display at Ensembl Current
display at RViewer main
display at UCSC main

1.Chrom	2.Start	3.End	4.Name
chr10	61988	62036	SRR054622.2038
chr10	69093	69141	SRR054622.2024
chr10	69098	69146	SRR054622.4815
chr10	69099	69147	SRR054622.7393
chr10	72859	72907	SRR054622.1184
chr10	92035	92083	SRR054622.2364

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

Analysis Pipeline



.FASTQ

Map reads

Remove duplicates

BAM to BED

Galaxy

.BED

Bin reads

GC bias correction

Segment bins

Call integer copy number state

Interactive visualization

Ginkgo

Hands-on with Ginkgo

ginkgo

Ginkgo: qb.cshl.edu/ginkgo

Sample Data: qb.cshl.edu/ginkgo/vizbi2015.tar