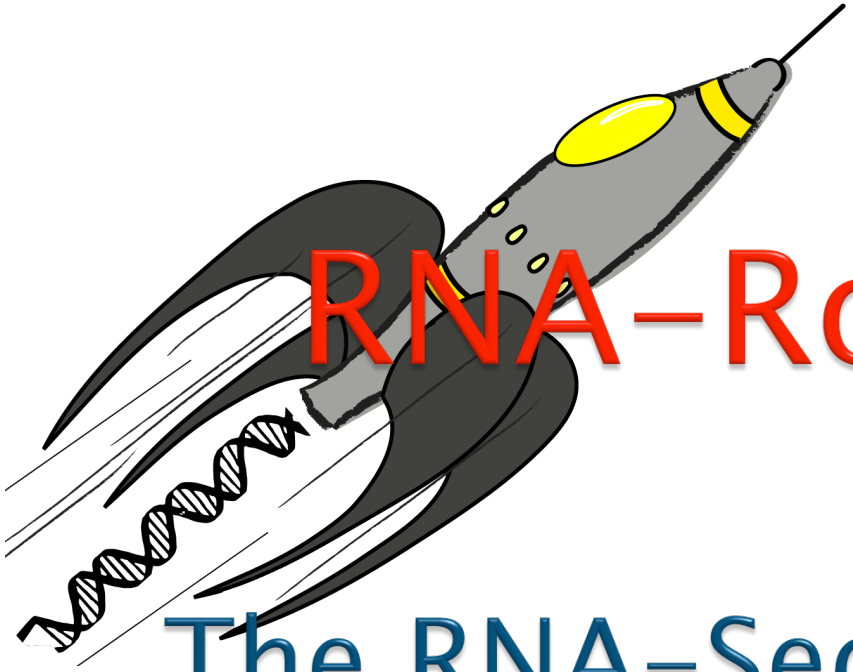


RNA-Rocket

The RNA-Seq Pipeline at

PATHOGENPORTAL
THE BIOINFORMATICS RESOURCE CENTERS PORTAL

CyberInfrastructure
DIVISION

Thank you.

Pathogen Portal

PATHOGENPORTAL
THE BIOINFORMATICS RESOURCE CENTERS PORTAL

Home | Data | Analyze | About | News and Announcements

Featuring...

RNA Rocket

Align your Illumina fastQ reads against supported genomes, view supported genomes, and estimate gene expression values using an **RNA-Seq Pipeline** running on Galaxy.

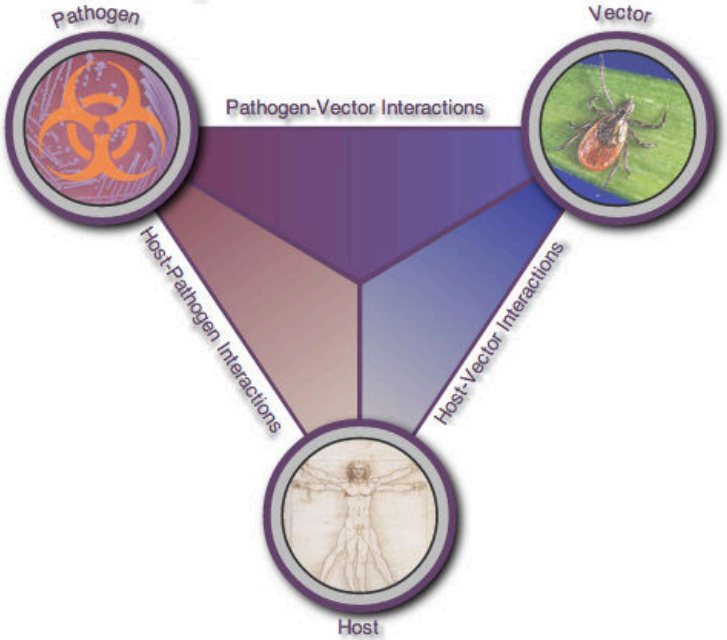
Improvements include:

- New user interface
- New reference genomes
- New tools

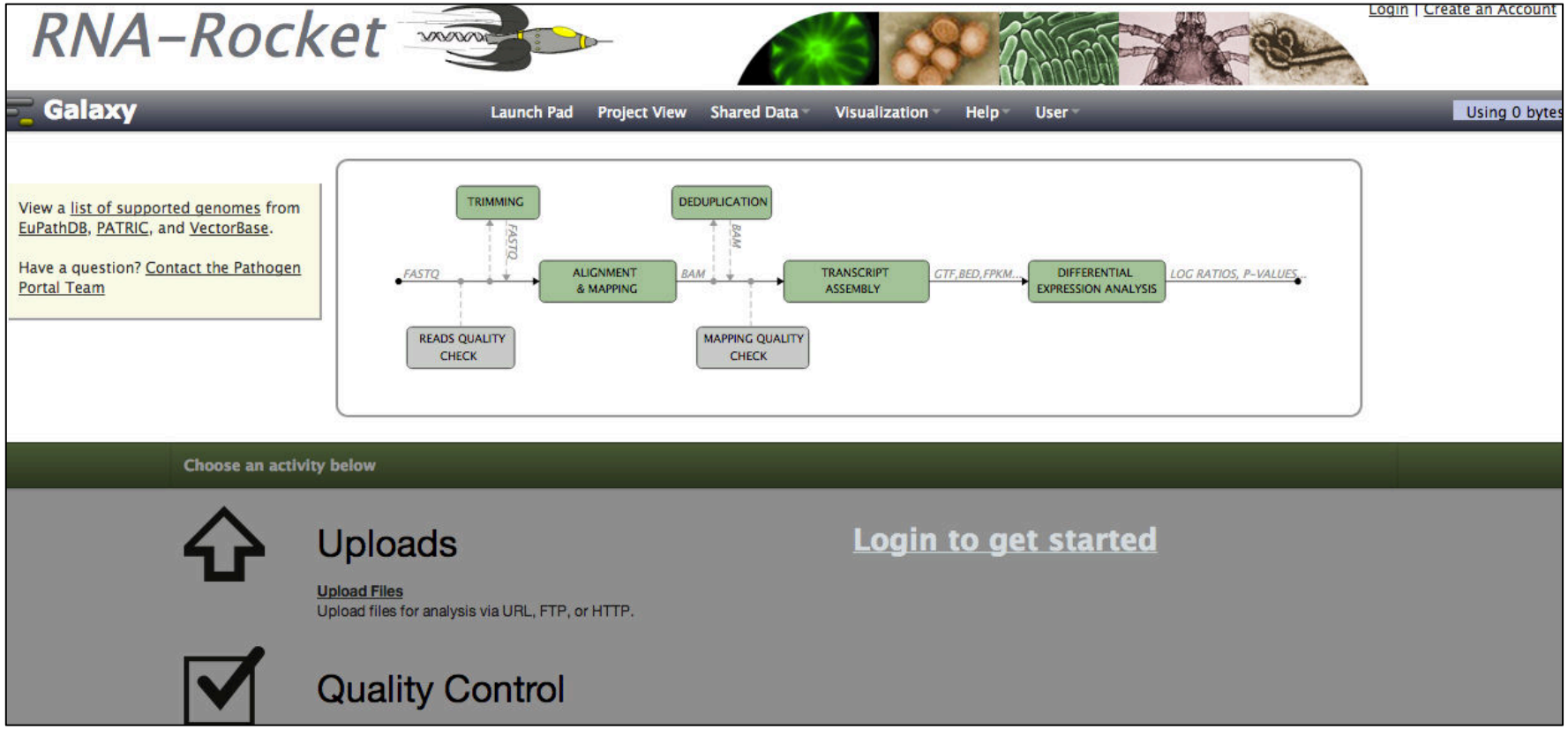
Pathogen Interaction Gateway

Generate a network graph of **Protein-Protein Interactions**, including **Host-Pathogen Interactions**, from your custom selection of hosts/vectors, bacteria, viruses, and eukaryotic pathogens.

Explore Infectious Disease



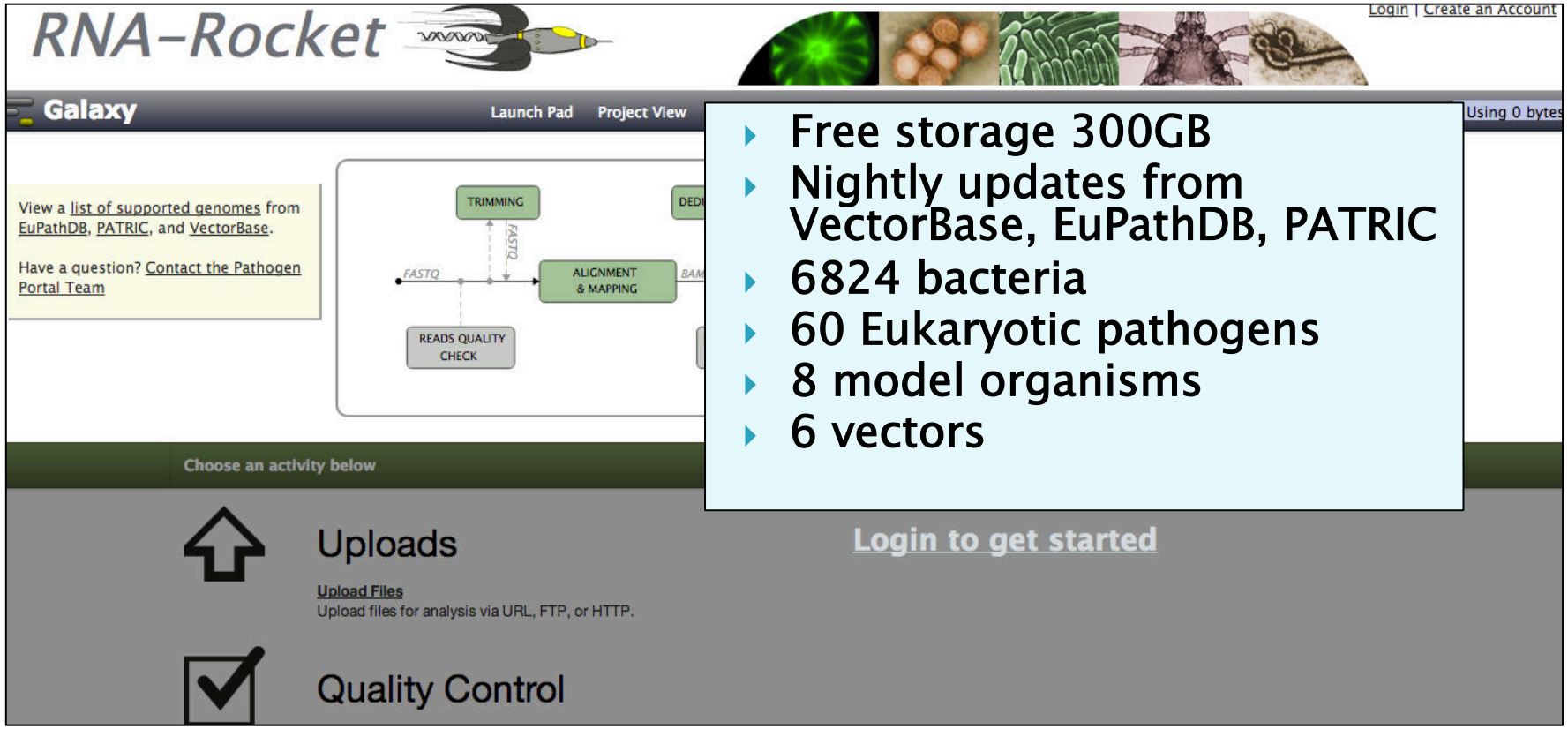
RNA-Rocket



PATHOGENPORTAL
THE BIOINFORMATICS RESOURCE CENTERS PORTAL

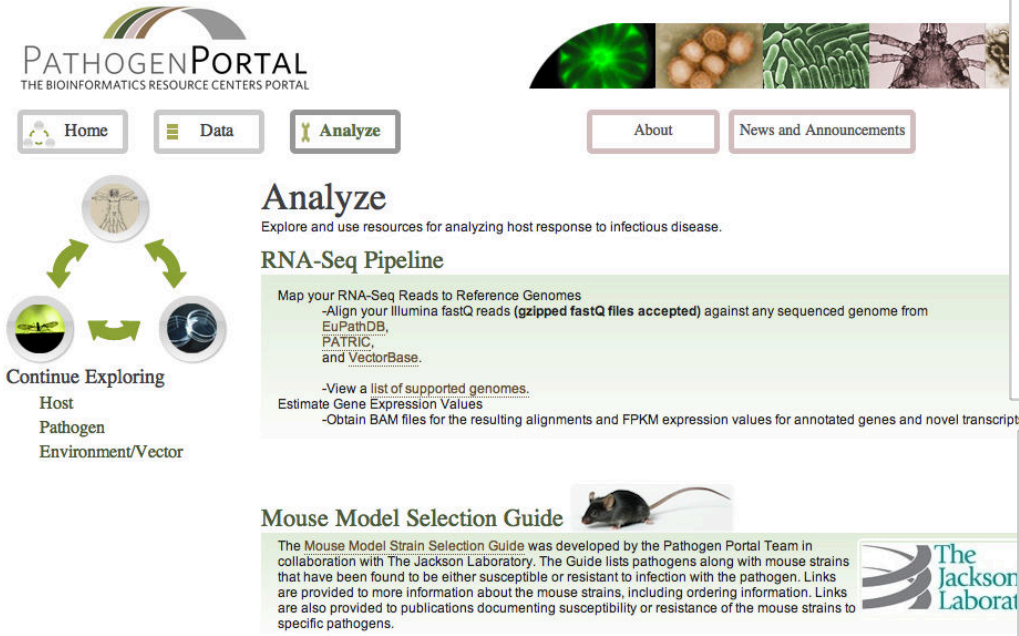
CyberInfrastructure
DIVISION

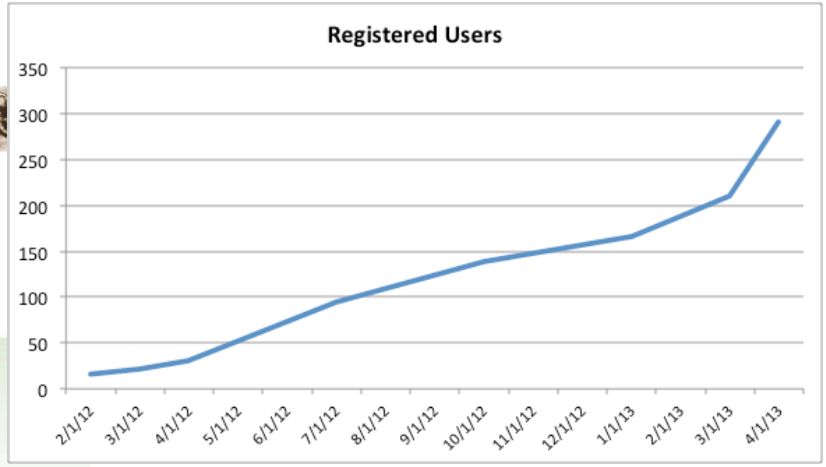
RNA-Rocket

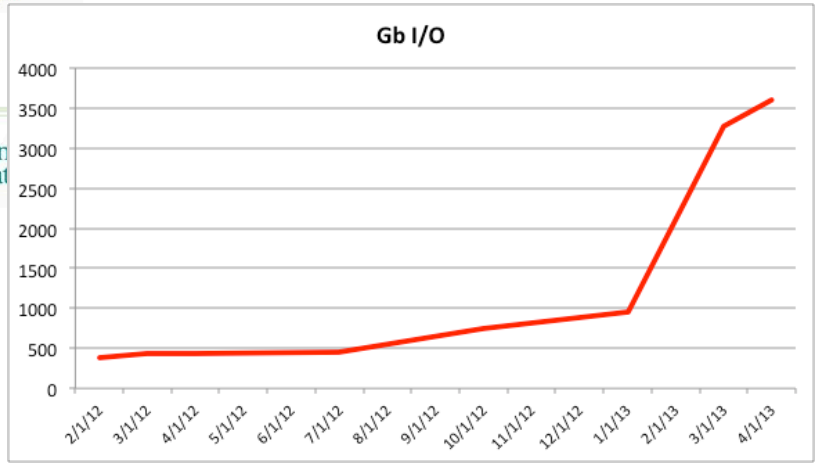


- Free storage 300GB
- Nightly updates from VectorBase, EuPathDB, PATRIC
- 6824 bacteria
- 60 Eukaryotic pathogens
- 8 model organisms
- 6 vectors

RNA-Seq Pipeline: Initial Release January 2012



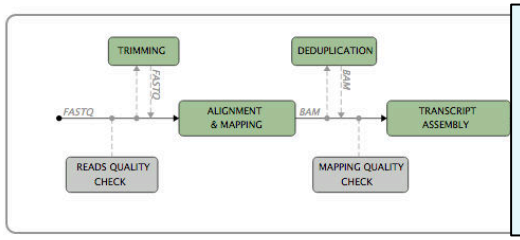




RNA-Seq Pipeline Features

View a list of supported genomes from EuPathDB, PATRIC, and VectorBase.

Have a question? Contact the Pathogen Portal Team



- Guided tour for naïve users

Choose an activity below

Uploads

Upload Files
Upload files for analysis via URL, FTP, or HTTP.

Quality Control

Check read quality
Optional: Run FastQC to get a report on the quality of base calls that could affect your read mapping.
Trim Reads
Optional: Run the sickle trimming tool to trim your reads and prepare them for alignment.
Check mapping quality
Optional: Check the number of reads mapped and the alignment quality.

RNA-Seq Analysis

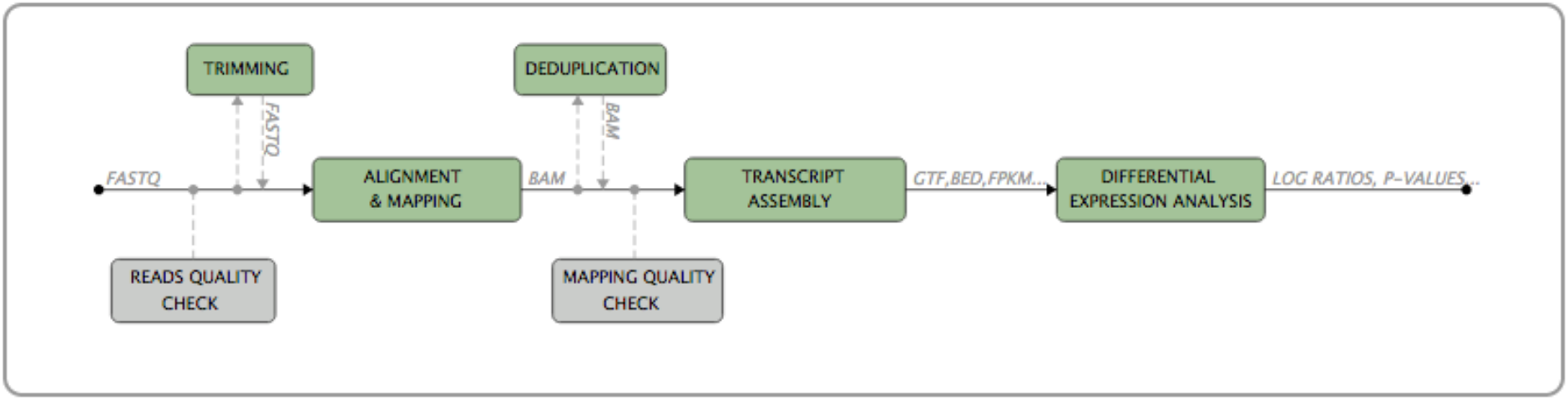
Align Reads & Assemble Transcripts
Map your reads to the genome and assemble them into transcripts. The alignment step will generate BAM files and the assembly step will generate BED and GTF files.
Differential Expression Analysis
Test RNA-Seq samples to determine if transcripts are differentially expressed.
Create GeneList Summary
Create a GeneList file, for use in PATRIC and other differential expression analysis tools.

Additional Tools

BED Tools
Use BEDTools to create summary BED files for analysis of genome coverage.
Remove Duplicate Reads
Optional: PCR amplification can lead to bias. For paired-end reads only: if multiple pairs of reads have the exact same cooridnates mark all except one as a duplicate and remove.
Alignment Only
To use advanced alignment parameters and/or perform alignment against a non-BRC genome.

Features

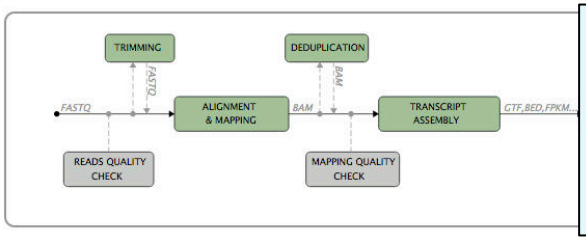
- Guided tour for naïve users



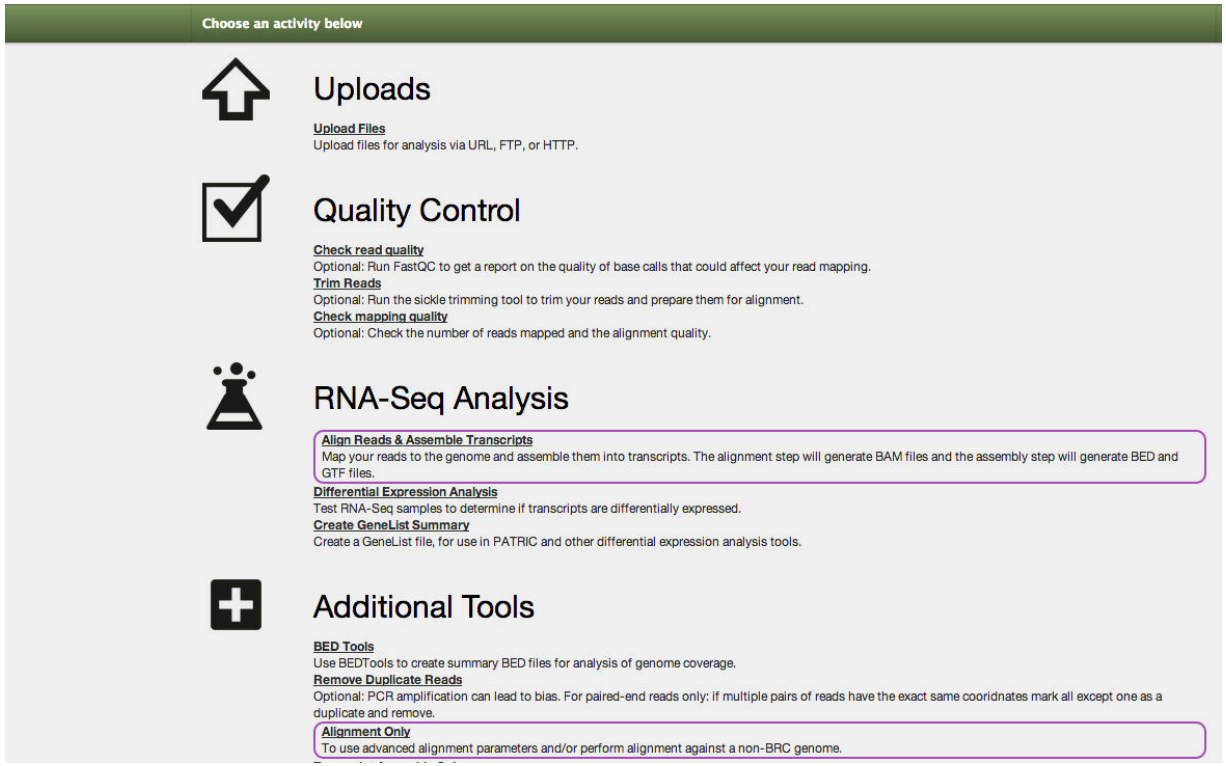
RNA-Seq Pipeline Features

View a list of supported genomes from EuPathDB, PATRIC, and VectorBase.

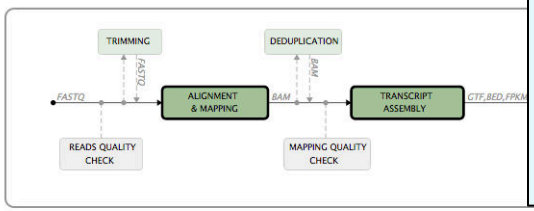
Have a question? Contact the Pathogen Portal Team



- Guided tour for naïve users



RNA-Seq Pipeline Features



- Guided tour for naïve users

Align Reads & Assemble Transcripts

Purpose:
This procedure will map RNA-Seq reads to one of the provided reference genomes and use this mapping to assemble transcripts, map transcripts to existing annotations, and determine the level of expression. Choose the appropriate option for your organism (Eukaryotic/Prokaryotic) and read type (Paired-end/ Single-end) below.
Required Input:
FastQ files
Output:
Read alignments (BAM Files), tab delimited assembly and expression files for known genes, isoforms, and novel transcripts.

Select Analysis Type
- Eukaryotic Single-End Analysis
- Prokaryotic Single-End Analysis
- Eukaryotic Paired-End Analysis
- Prokaryotic Paired-End Analysis

Select an existing Project or create a new Project to be used during this analysis and populate the Project with the necessary files. Output from this analysis will be saved in the selected Project.
Currently Selected Project: None Selected

Select and copy files from Uploads or existing project(s) to populate your current Project.

Target Project:
Select existing project — OR — Create project

← Copy

Source Project:
Select source
imported: CSU Demo

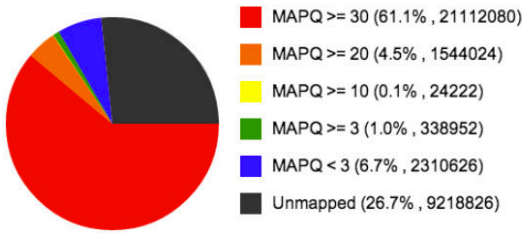
- Upstream_READ1.fastq
- Downstream_READ2.fastq
- Trimmed_Upstream_READ1.fastq
- Trimmed_Downstream_READ2.fastq
- BaseQuality_Upstream_READ1.fastq.html
- BaseQuality_Trimmed_Upstream_READ1.fastq.html
- Align with Bowtie2 Original
- Align with Bowtie2 Trimmed
- SamStat for Align with Bowtie2 on Original
- SamStat for Align with Bowtie2 Trimmed.html

RNA-Seq Pipeline Features

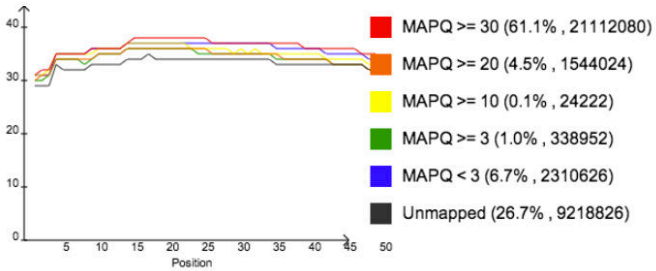
- Quality control tools for read data

Statistics for Align with Bowtie2 on data 1 and data 2: n reads

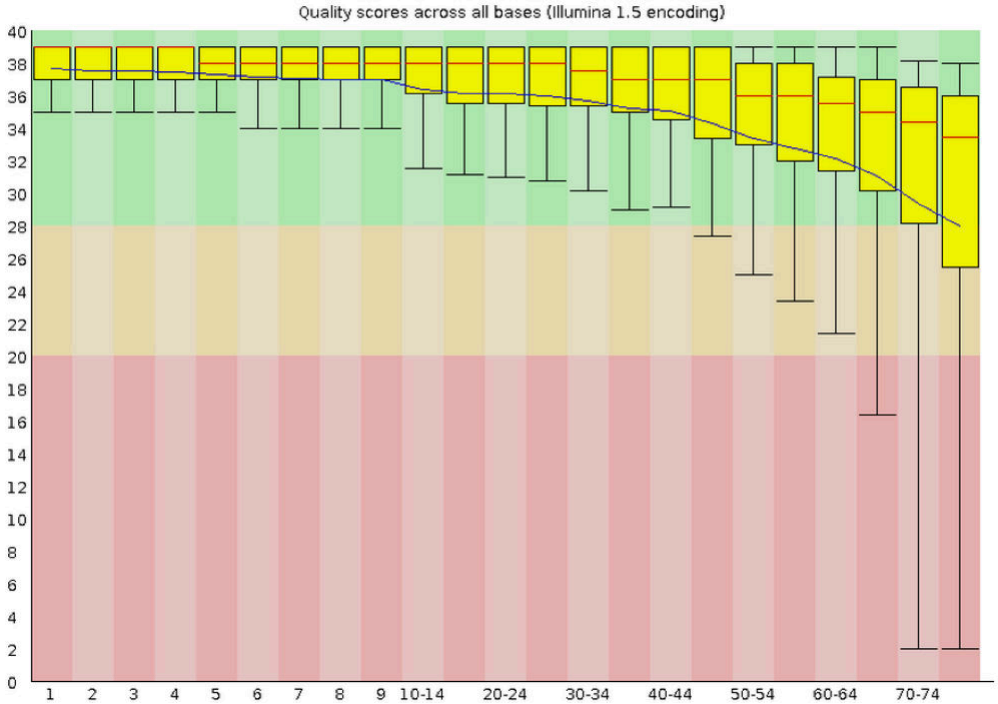
Mapping stats: 73% aligned (25.3M aligned out of 34.5M total)



Mean Base Quality



Per base sequence quality

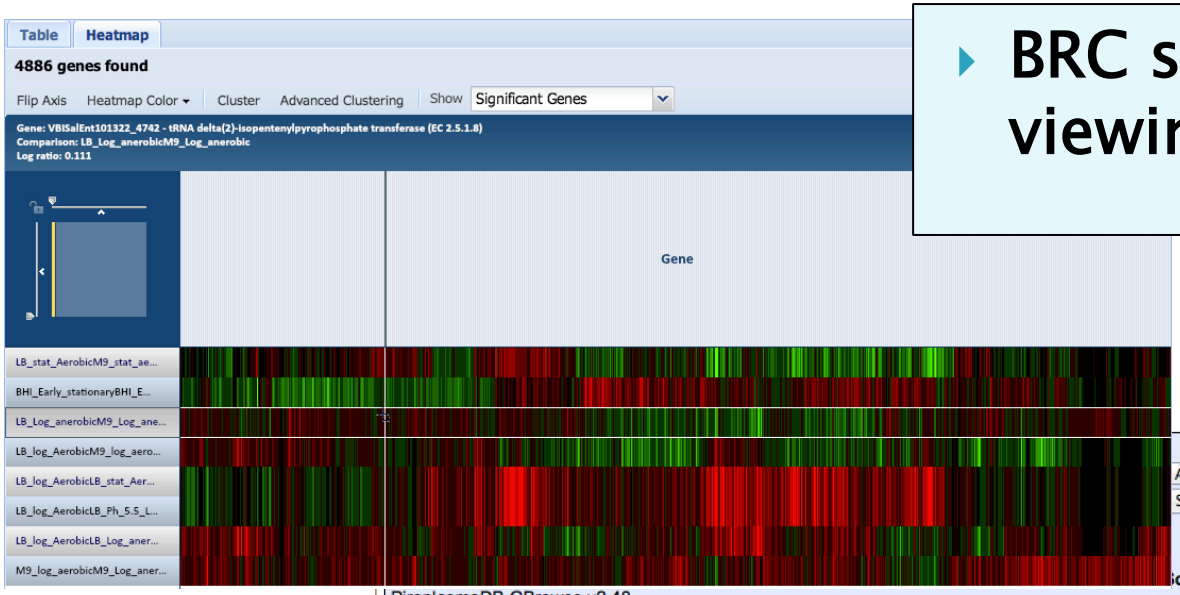


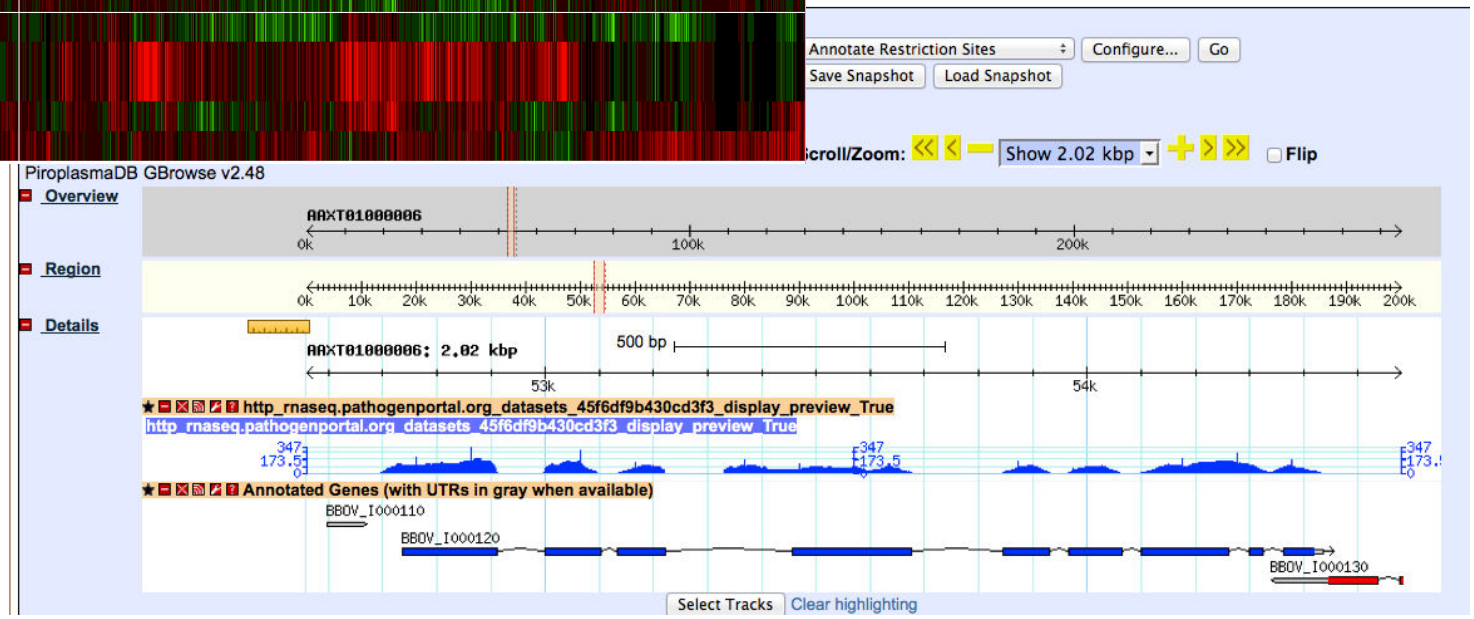
PATHOGENPORTAL
THE BIOINFORMATICS RESOURCE CENTERS PORTAL

CyberInfrastructure DIVISION

RNA-Seq Pipeline Features

- BRC streaming and viewing of results





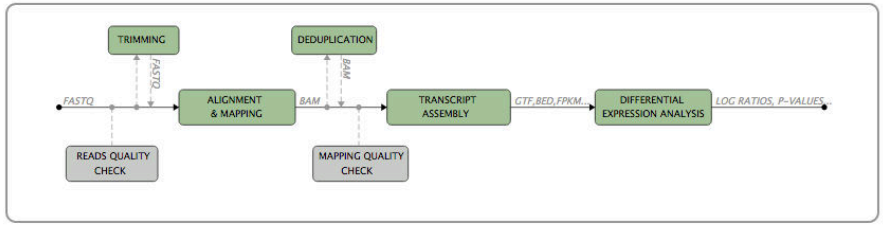
PATHOGENPORTAL
THE BIOINFORMATICS RESOURCE CENTERS PORTAL

CyberInfrastructure
DIVISION

Galaxy Modifications

View a list of supported genomes from EuPathDB, PATRIC, and VectorBase.

Have a question? Contact the Pathogen Portal Team



Choose an activity below

Uploads

Upload Files
Upload files for analysis via URL, FTP, or HTTP.

Quality Control

Check read quality
Optional: Run FastQC to get a report on the quality of base calls that could affect
Trim Reads
Optional: Run the sickle trimming tool to trim your reads and prepare them for alig
Check mapping quality
Optional: Check the number of reads mapped and the alignment quality.

RNA-Seq Analysis

Align Reads & Assemble Transcripts
Map your reads to the genome and assemble them into transcripts. The alignment step will generate BAM files and the assembly step will generate BED and GTF files.
Differential Expression Analysis
Test RNA-Seq samples to determine if transcripts are differentially expressed.
Create GeneList Summary
Create a GeneList file, for use in PATRIC and other differential expression analysis tools.

Additional Tools

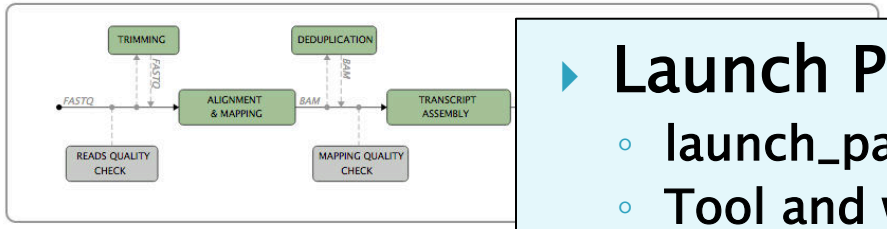
BED Tools
Use BEDTools to create summary BED files for analysis of genome coverage.
Remove Duplicate Reads
Optional: PCR amplification can lead to bias. For paired-end reads only: if multiple pairs of reads have the exact same cooridnates mark all except one as a duplicate and remove.
Alignment Only
To use advanced alignment parameters and/or perform alignment against a non-BRC genome.

- Launch Pad
  - D3 library concept diagram

Galaxy Modifications

View a list of supported genomes from EuPathDB, PATRIC, and VectorBase.

Have a question? Contact the Pathogen Portal Team



- **Launch Pad**
  - launch_pad.py controller
  - Tool and workflow lists

Choose an activity below

Uploads

Upload Files
Upload files for analysis via URL, FTP, or HTTP.

Quality Control

Check read quality
Optional: Run FastQC to get a report on the quality of base calls that could affect your read mapping.
Trim Reads
Optional: Run the sickle trimming tool to trim your reads and prepare them for alignment.
Check mapping quality
Optional: Check the number of reads mapped and the alignment quality.

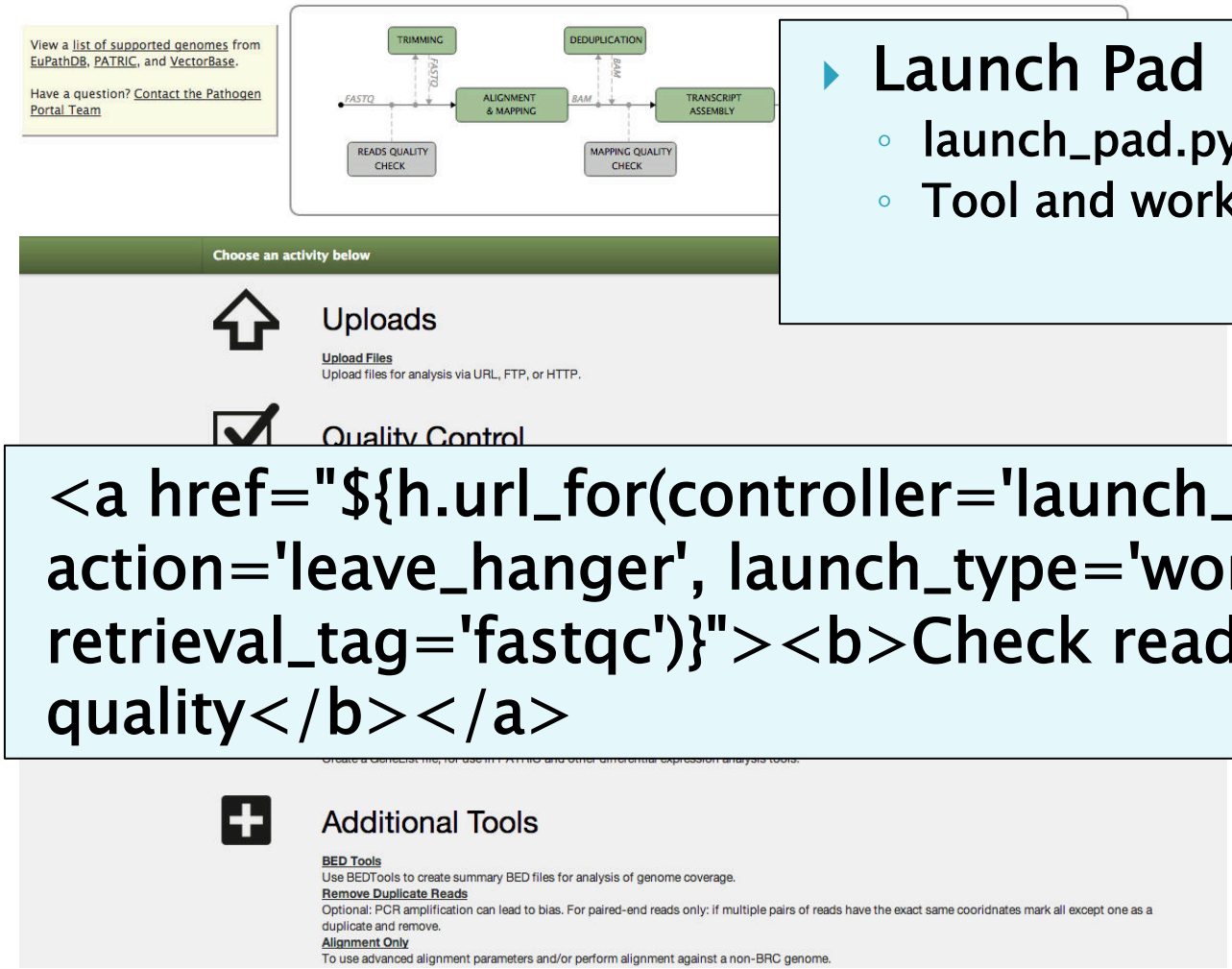
RNA-Seq Analysis

Align Reads & Assemble Transcripts
Map your reads to the genome and assemble them into transcripts. The alignment step will generate BAM files and the assembly step will generate BED and GTF files.
Differential Expression Analysis
Test RNA-Seq samples to determine if transcripts are differentially expressed.
Create GeneList Summary
Create a GeneList file, for use in PATRIC and other differential expression analysis tools.

Additional Tools

BED Tools
Use BEDTools to create summary BED files for analysis of genome coverage.
Remove Duplicate Reads
Optional: PCR amplification can lead to bias. For paired-end reads only: if multiple pairs of reads have the exact same cooridnates mark all except one as a duplicate and remove.
Alignment Only
To use advanced alignment parameters and/or perform alignment against a non-BRC genome.

Galaxy Modifications

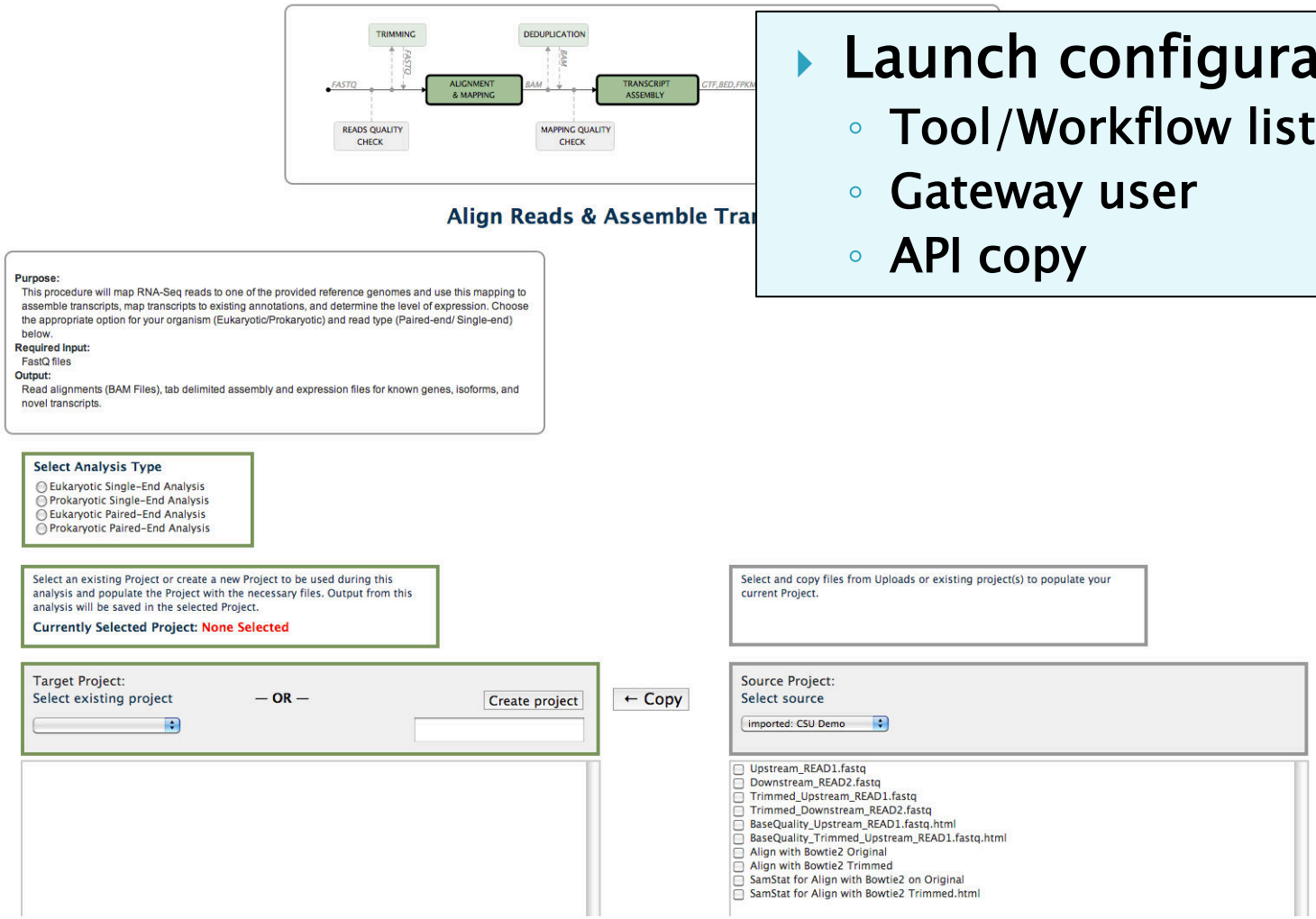


- **Launch Pad**
  - launch_pad.py controller
  - Tool and workflow lists

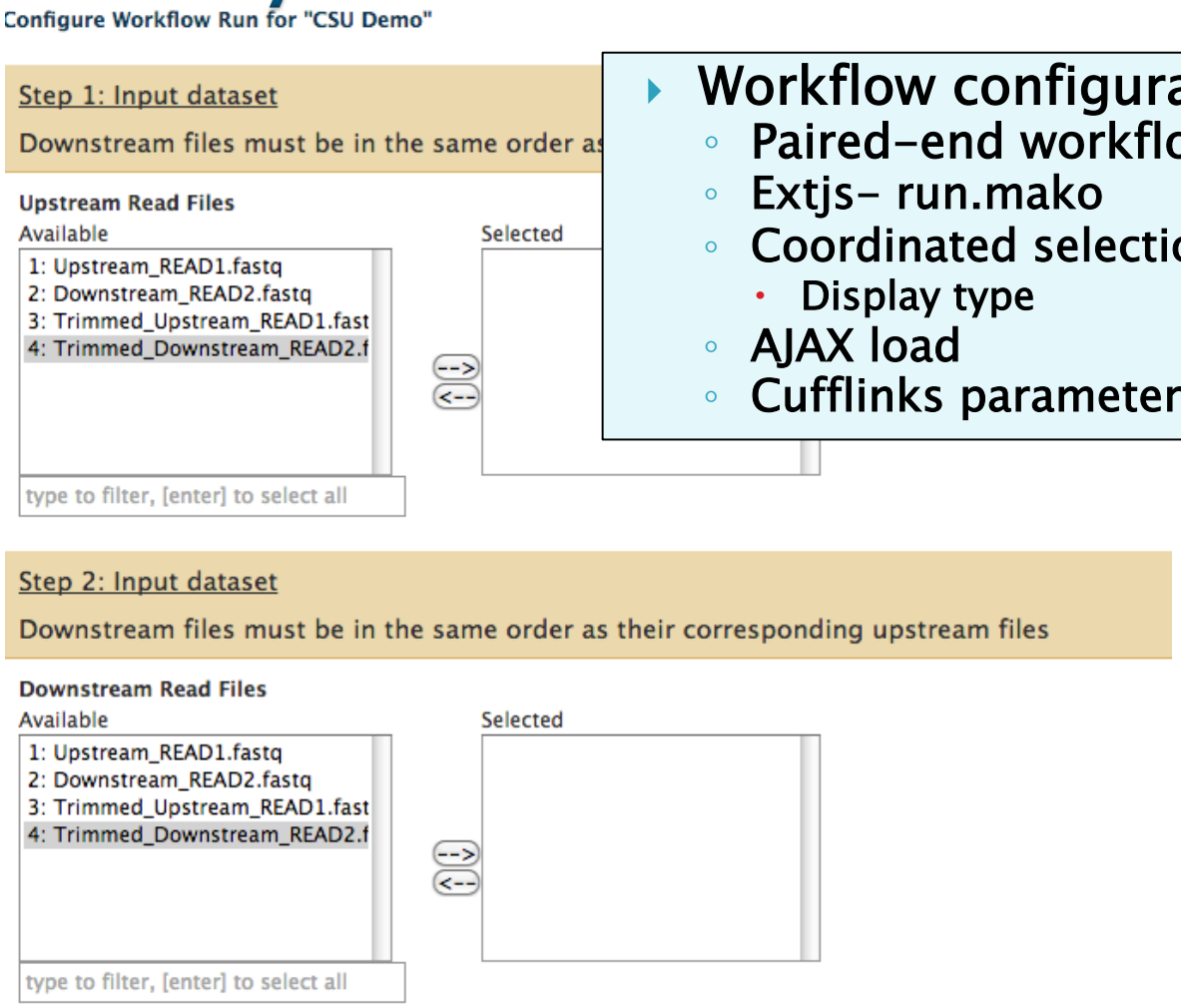
```
<a href="${h.url_for(controller='launch_pad', action='leave_hanger', launch_type='workflow', retrieval_tag='fastqc')}"><b>Check read quality</b></a>
```

Galaxy Modifications

- Launch configuration
  - Tool/Workflow lists
  - Gateway user
  - API copy

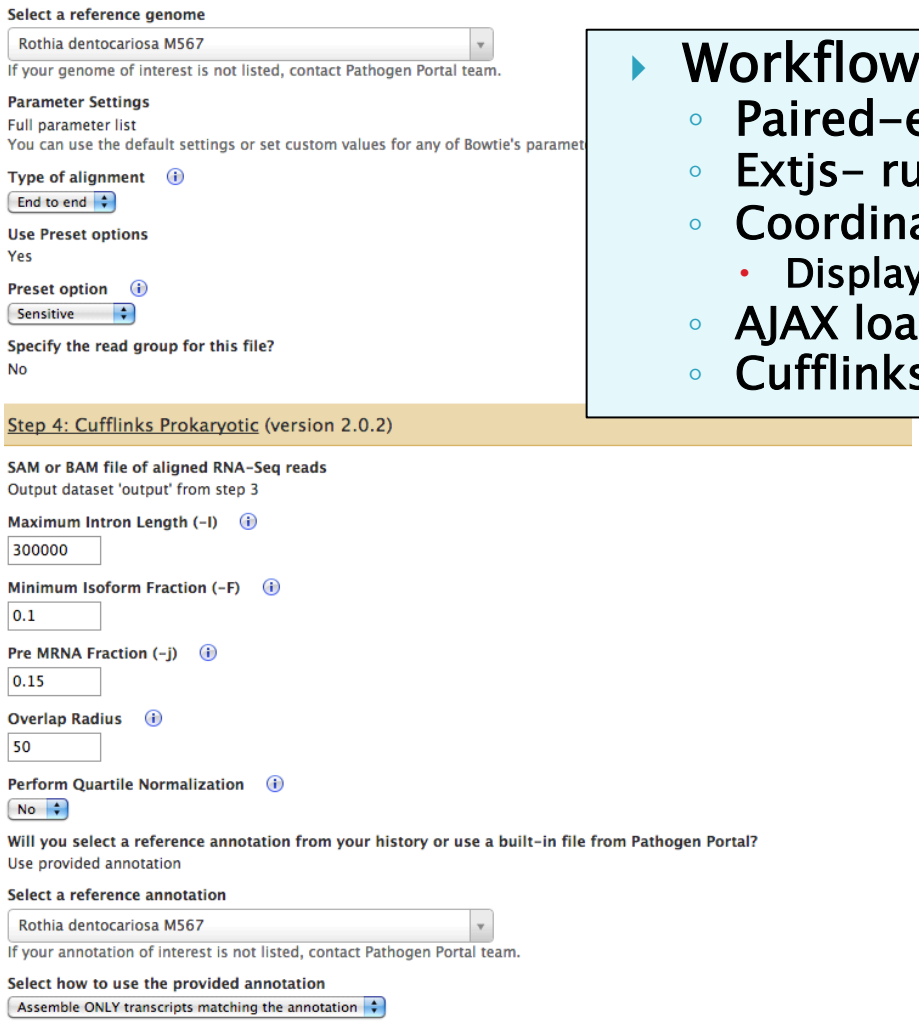


Galaxy Modifications



- Workflow configuration
  - Paired-end workflows
  - Extjs- run.mako
  - Coordinated selection
    - Display type
  - AJAX load
  - Cufflinks parameters e.g.

Galaxy Modifications



- Workflow configuration
  - Paired-end workflows
  - Extjs- run.mako
  - Coordinated selection
    - Display type
  - AJAX load
  - Cufflinks parameters e.g.

Other Modifications

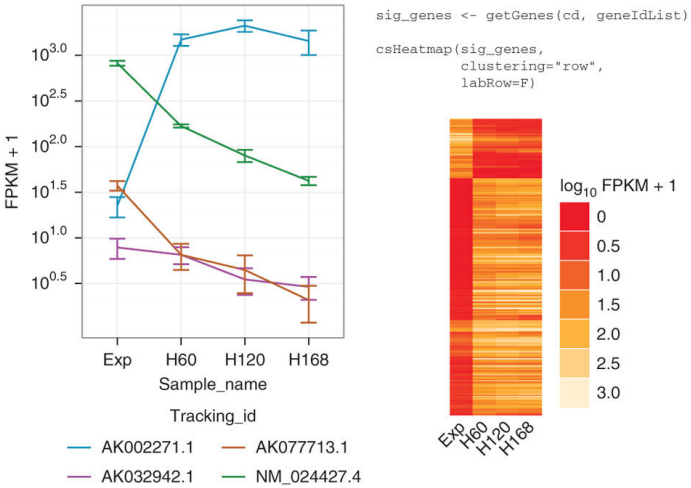
- Cuffdiff/cuffmerge integration
- SAM Stat
- Select2 wait time
- Power users

Challenges

- Controller system
- Workflow vs tool runners
  - Fixed parameters
  - Inputs
- Advanced parameters

Future Plans

- Visualization of results
  - CummeRbund
- Organism specific tools
- Link outs to BRCs
  - Annotation differences
  - Differential Expression Visualization
  - Alignment results streaming



Acknowledgements

pathogenportal.org

- Eric Nordberg
- Dustin Machi
- Chunhong Mao
- This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN272200900040C, awarded to BWS Sobral.

PATHOGENPORTAL
THE BIOINFORMATICS RESOURCE CENTERS PORTAL

CyberInfrastructure
DIVISION