

# Mapeamento das leituras no genoma de referência

**Pablo Rodrigo Sanches**

Departamento de Genética – FMRP/USP

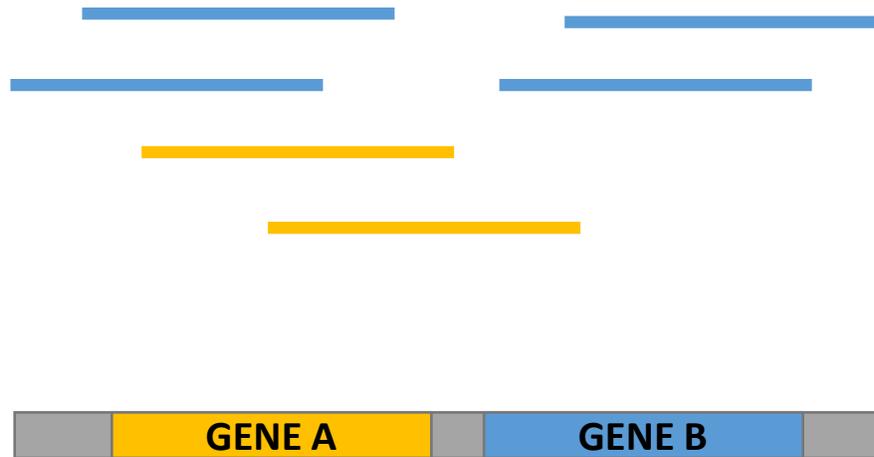
psanches@usp.br

# Roteiro de análise

1. RNA STAR (mapeamento das bibliotecas de leituras contra o genoma de referência)
2. Finalizar a tabela de resultados sobre as características gerais das bibliotecas sequenciadas

# Recordando...

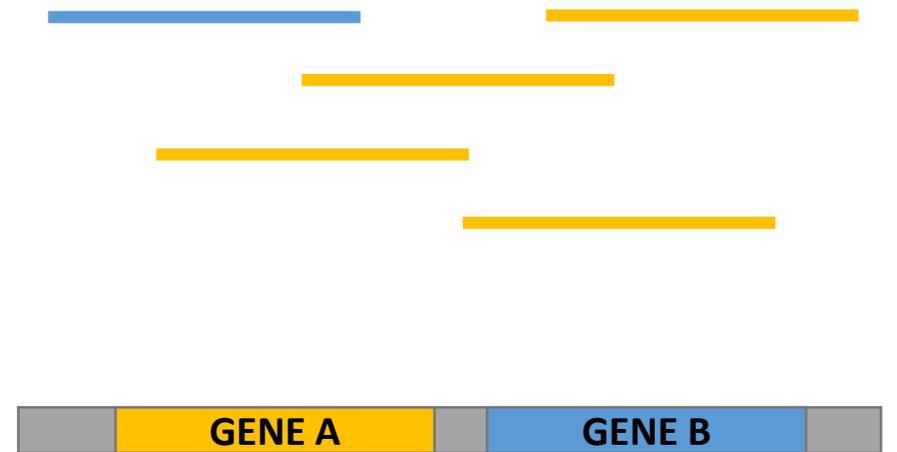
Controle (ex. não tratado)



Reads (leituras)  
do RNA-Seq

Genoma de  
referência

Desafio (ex. tratado com droga)



# Arquivos necessários para o genoma de referência

Arquivo\_de\_supercontig.fasta – *conjunto ordenado e orientado de contigs que ainda possuem lacunas*

```
>Supercontig_2.1 of Trichophyton rubrum CBS 118892
ATAGTATTACTTACTACTACTATAAGTCCCTATTATAGAAGGTATGCTCTACTATTAGTA
TCTAATCTAGATAACTATAAACTATATTTAAATATCTTTAAACTTAGATATATAGCG
GTAGTTAACTACTAGTATACCTCTATTTAGAAGGCGTAGATAGCTTATTTAACTCTCTA
GATAATACTAAAGTATAGAGATTTAATATTAAGTACTACCTATATTAGCCTTTTATAAT
ATTAATAATAATCTTCTATAAAGTAGTAATACCTATAAAAAAGCTATTACTATTCCTATA
GAGTGTAATAATAATATTATATCTAATTCTAATAGTACTTTAGAGGATATTAATATAAAA
GATAACTCCACTACTATTTTTAATAGCCCCTACTATATCTAGACCCCCTAAGTACCGTAAG
TATAAGGACGCGTCTCTAATAGAGCTACTATAGCTAAGTAGCTAAAAGTAGATAGGCTA
ATATCTTATATTATATTAAGGACTAATATAGTACTAAATATAACTAATAAATCTAAAAAA
GATAGTAAAGGTTTTATATTTAGATTTTGAATATATTAATATCTAGAGTATTATCTACT
AAATAACCTACTACTATATACTCTATAAACTCTATTATATTAATATCTAATAATAATATA
ATTAAGTTAGTTTATTTATAGCTACTAATACTTATTTACTACTATAATCTATCTAATATT
TTCTTTTTAGGCCCTATATTACTTAATATCGCTAAAACCTTATTTAAAGCCTTCTCTATA
TTTTTAGACCCCTATAGATACTATCTACTATTTTAGACTATTAATTAGTATATAGATACT
ACCTATTACTTCTATAATAAGCTTAACTAATATATCTTATACTATATAGTATCTACTATA
GCTATTAATAAATAATAACTACTATCTTACTTATTATATTATTTAAATATATTATAG
--More-- (0%)
```



Um contig é um conjunto de segmentos de DNA sobrepostos que juntos representam uma região de consenso do DNA

Arquivo.gtf - *usado para descrever genes e outras informações sobre sua estrutura*

```
Supercontig_2.1 TR_CBS118892_V2_FINAL_CALLGENES_5 start_codon 14925 14927 . - 0 gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig_2.1 TR_CBS118892_V2_FINAL_CALLGENES_5 stop_codon 12841 12843 . - 0 gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig_2.1 TR_CBS118892_V2_FINAL_CALLGENES_5 exon 14609 14927 . - . gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig_2.1 TR_CBS118892_V2_FINAL_CALLGENES_5 CDS 14609 14927 . - 0 gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig_2.1 TR_CBS118892_V2_FINAL_CALLGENES_5 exon 13933 14529 . - . gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig_2.1 TR_CBS118892_V2_FINAL_CALLGENES_5 CDS 13933 14529 . - 2 gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig_2.1 TR_CBS118892_V2_FINAL_CALLGENES_5 exon 13702 13869 . - . gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig_2.1 TR_CBS118892_V2_FINAL_CALLGENES_5 CDS 13702 13869 . - 2 gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig_2.1 TR_CBS118892_V2_FINAL_CALLGENES_5 exon 13470 13638 . - . gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig_2.1 TR_CBS118892_V2_FINAL_CALLGENES_5 CDS 13470 13638 . - 2 gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig_2.1 TR_CBS118892_V2_FINAL_CALLGENES_5 exon 12841 13414 . - . gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig_2.1 TR_CBS118892_V2_FINAL_CALLGENES_5 CDS 12844 13414 . - 1 gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig_2.1 TR_CBS118892_V2_FINAL_CALLGENES_5 start_codon 16137 16139 . + 0 gene_id "TERG_00003"; transcript_id "TERG_00003T0";
Supercontig_2.1 TR_CBS118892_V2_FINAL_CALLGENES_5 stop_codon 17793 17795 . + 0 gene_id "TERG_00003"; transcript_id "TERG_00003T0";
Supercontig_2.1 TR_CBS118892_V2_FINAL_CALLGENES_5 exon 16137 16610 . + . gene_id "TERG_00003"; transcript_id "TERG_00003T0";
Supercontig_2.1 TR_CBS118892_V2_FINAL_CALLGENES_5 CDS 16137 16610 . + 0 gene_id "TERG_00003"; transcript_id "TERG_00003T0";
Supercontig_2.1 TR_CBS118892_V2_FINAL_CALLGENES_5 exon 16685 17514 . + . gene_id "TERG_00003"; transcript_id "TERG_00003T0";
--More-- (0%)
```

# Enviar arquivos do genoma para o Galaxy

The screenshot displays the Galaxy web interface with a modal dialog box titled "Download from web or upload from disk". The dialog has tabs for "Regular", "Composite", "Collection", and "Rule-based". A large dashed box in the center contains the text "Drop files here". Below this box are two dropdown menus: "Type (set all):" set to "Auto-detect" and "Genome (set all):" set to "---- Additional S...". At the bottom of the dialog are four buttons: "Choose local files", "Choose FTP files", "Paste/Fetch data", and "Start", "Pause", "Reset", "Close". A yellow arrow points to the "Upload Data" button in the left sidebar, and a black arrow points to the "Choose local files" button in the dialog. The background shows the Galaxy interface with a "History" panel on the right listing various datasets.

Galaxy is an open source platform for genomic data analysis. You can install your own Galaxy instance or use the public Galaxy instances.

Resources. You can find more information about Galaxy at [galaxyproject.org](http://galaxyproject.org).

Galaxy can best be used by a research group or faculty.

Galaxy was established by the University of Texas at Austin. The JXTC (The Joint XTC) is a computational biology center that provides training sessions and resources for researchers.

Galaxy was found on the [www.oregonhealthscience.edu](http://www.oregonhealthscience.edu) website.

Want to learn the best practices for the analysis of SARS-CoV-2 data using Galaxy? Visit the Galaxy SARS-CoV-2 portal at [covid19.galaxyproject.org](http://covid19.galaxyproject.org)

**PennState** **JOHNS HOPKINS UNIVERSITY** **OREGON HEALTH & SCIENCE UNIVERSITY** **TACC** **CYVERSE**

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

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

**History**

search datasets

**UDA**

137 shown, 100 deleted, 6 hidden

79.81 GB

data 130, data 129, and others: log

137: 0h\_III RNA STAR on data 130, data 129, and others: mapped.bam

136: 0h\_III RNA STAR on data 130, data 129, and others: splice junctions.bed

135: 0h\_III RNA STAR on data 130, data 129, and others: log

133: 0h\_II RNA STAR on data 130, data 129, and others: mapped.bam

132: 0h\_II RNA STAR on data 130, data 129, and others: splice junctions.bed

131: 0h\_II RNA STAR on data 130, data 129, and others: log

130: trichophyton\_rubrum\_cbs\_118892\_2\_transcripts.gtf

129: trichophyton\_rubrum\_cbs\_118892\_2\_supergenes.fasta

128: FastQC on data 10

0: RawData

# Arquivo\_de\_supercontig.fasta no Galaxy

The screenshot displays the Galaxy web interface. At the top, the browser address bar shows 'usegalaxy.org'. The main navigation bar includes 'Analyze Data', 'Workflow', 'Visualize', 'Shared Data', 'Help', and 'User'. A notification banner at the top center states: 'This dataset is large and only the first megabyte is shown below. Show all | Save'. The left sidebar contains a 'Tools' section with a search bar and an 'Upload Data' button. Below this is a 'Get Data' section with 'Collection Operations' and 'GENERAL TEXT TOOLS'. The 'FASTA/FASTQ' section is highlighted, showing options for 'FASTQ Quality Control', 'SAM/BAM', 'BED', 'VCF/BCF', 'Nanopore', and 'Convert Formats'. The main content area displays the first megabyte of a FASTA file named '>Supercontig21 of Trichophyton rubrum CBS 118892'. The right sidebar shows a 'History' section with a search bar and a list of jobs. The jobs listed include '132: 0h\_II RNA STAR on data 130, data 129, and others: splice junctions.bed', '131: 0h\_II RNA STAR on data 130, data 129, and others: log', '130: trichophyton\_rubrum\_cbs\_118892\_2\_transcripts.gtf', '129: trichophyton\_rubrum\_cbs\_118892\_2\_supercontigs.fasta', '128: FastQC on data 100: RawData', '127: FastQC on data 100: Webpage', '126: FastQC on data 99: RawData', '125: FastQC on data 99: Webpage', '124: FastQC on data 94: RawData', and '123: FastQC on data 94: Webpage'.

# Arquivo.gtf no Galaxy

The screenshot displays the Galaxy web interface at usegalaxy.org. The main panel shows a table of genomic features for Supercontig21. The table columns include contig name, feature name, type, start and end coordinates, strand, phase, and gene/transcript IDs. The left sidebar lists tool categories, and the right sidebar shows a history of datasets.

Contig	Feature	Type	Start	End	Strand	Phase	Gene/Transcript IDs	
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	start_codon	14925	14927	.	-	0	gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	stop_codon	12841	12843	.	-	0	gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	exon	14609	14927	.	-	.	gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	CDS	14609	14927	.	-	0	gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	exon	13933	14529	.	-	.	gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	CDS	13933	14529	.	-	2	gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	exon	13702	13869	.	-	.	gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	CDS	13702	13869	.	-	2	gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	exon	13470	13638	.	-	.	gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	CDS	13470	13638	.	-	2	gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	exon	12841	13414	.	-	.	gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	CDS	12844	13414	.	-	1	gene_id "TERG_00002"; transcript_id "TERG_00002T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	start_codon	16137	16139	.	+	0	gene_id "TERG_00003"; transcript_id "TERG_00003T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	stop_codon	17793	17795	.	+	0	gene_id "TERG_00003"; transcript_id "TERG_00003T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	exon	16137	16610	.	+	.	gene_id "TERG_00003"; transcript_id "TERG_00003T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	CDS	16137	16610	.	+	0	gene_id "TERG_00003"; transcript_id "TERG_00003T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	exon	16685	17514	.	+	.	gene_id "TERG_00003"; transcript_id "TERG_00003T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	CDS	16685	17514	.	+	0	gene_id "TERG_00003"; transcript_id "TERG_00003T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	exon	17603	17795	.	+	.	gene_id "TERG_00003"; transcript_id "TERG_00003T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	CDS	17603	17792	.	+	1	gene_id "TERG_00003"; transcript_id "TERG_00003T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	start_codon	18581	18583	.	+	0	gene_id "TERG_00004"; transcript_id "TERG_00004T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	stop_codon	19811	19813	.	+	0	gene_id "TERG_00004"; transcript_id "TERG_00004T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	exon	18581	19813	.	+	.	gene_id "TERG_00004"; transcript_id "TERG_00004T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	CDS	18581	19810	.	+	0	gene_id "TERG_00004"; transcript_id "TERG_00004T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	start_codon	29218	29220	.	+	0	gene_id "TERG_00008"; transcript_id "TERG_00008T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	stop_codon	30984	30986	.	+	0	gene_id "TERG_00008"; transcript_id "TERG_00008T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	exon	29218	29382	.	+	.	gene_id "TERG_00008"; transcript_id "TERG_00008T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	CDS	29218	29382	.	+	0	gene_id "TERG_00008"; transcript_id "TERG_00008T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	exon	29454	30052	.	+	.	gene_id "TERG_00008"; transcript_id "TERG_00008T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	CDS	29454	30052	.	+	0	gene_id "TERG_00008"; transcript_id "TERG_00008T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	exon	30102	30120	.	+	.	gene_id "TERG_00008"; transcript_id "TERG_00008T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	CDS	30102	30120	.	+	1	gene_id "TERG_00008"; transcript_id "TERG_00008T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	exon	30174	30484	.	+	.	gene_id "TERG_00008"; transcript_id "TERG_00008T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	CDS	30174	30484	.	+	0	gene_id "TERG_00008"; transcript_id "TERG_00008T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	exon	30541	30811	.	+	.	gene_id "TERG_00008"; transcript_id "TERG_00008T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	CDS	30541	30811	.	+	1	gene_id "TERG_00008"; transcript_id "TERG_00008T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	exon	30867	30986	.	+	.	gene_id "TERG_00008"; transcript_id "TERG_00008T0";
Supercontig21	TR_CBS118892_V2_FINAL_CALLGENES_5	CDS	30867	30983	.	+	0	gene_id "TERG_00008"; transcript_id "TERG_00008T0";

# RNA STAR – Via Galaxy

Galaxy

usegalaxy.org

Galaxy Analyze Data Workflow Visualize Shared Data Help User

Using 31%

Tools

rna star

Upload Data

Show Sections

fastp - fast all-in-one preprocessing for FASTQ files

ABRicate List all of abricate's available databases.

BamLeftAlign indels in BAM datasets

Cuffmerge merge together several Cufflinks assemblies

RNA STAR Gapped-read mapper for RNA-seq data

RNA STARsolo mapping, demultiplexing and gene quantification for single cell RNA-seq

Trinity de novo assembly of RNA-Seq data

STAR-Fusion detect fusion genes in RNA-Seq data

Align reads and estimate abundance on a de novo assembly of RNA-Seq data

FEELnc FExible Extraction of LncRNA

Salmon quant Perform dual-phase, reads or mapping-based estimation of transcript abundance from RNA-seq reads

Seurat - toolkit for exploration of single-cell RNA-seq data

RNA Structure Prediction predict RNA structures with or without experimental constraints from the Reactivity Calculation module

QualiMap RNA-Seq QC

Remove Unwanted Variation from RNA-seq data

siRNA Finds siRNA duplexes in mRNA

RNA STAR Gapped-read mapper for RNA-seq data (Galaxy Version 2.7.8a)

Favorite Versions Options

Single-end or paired-end reads

Paired-end (as individual datasets)

Paired-end

RNA-Seq FASTQ/FASTA file, forward reads

69: Trimmomatic on 0h\_II\_R1.fastq.gz (R1 paired)

Fastq "trimado" (R1 paired)

RNA-Seq FASTQ/FASTA file, reverse reads

70: Trimmomatic on 0h\_II\_R2.fastq.gz (R2 paired)

Fastq "trimado" (R2 paired)

Custom or built-in reference genome

Use reference genome from history and create temporary index

Built-ins were indexed using default options

Select a reference genome

129: trichophyton\_rubrum\_cbs\_118892\_2\_supercontigs.fasta

Arquivo\_de\_supercontig.fasta

(--genomeFastaFiles)

Length of the SA pre-indexing string

14

Typically between 10 and 15. Longer strings will use much more memory, but allow faster searches. For small genomes, the parameter --genomeSAindexNbases must be scaled down to min(14, log2(GenomeLength)/2 - 1) (--genomeSAindexNbases)

Build index with or without known splice junctions annotation

build index with gene-model

To build an index with known splice junctions annotated, you will have to provide a GTF or GFF3 dataset that describes the gene models (the location of genes, transcripts and exons) known for the reference genome.

Gene model (gff3,gtf) file for splice junctions

130: trichophyton\_rubrum\_cbs\_118892\_2\_transcripts.gtf

Arquivo.gtf

Exon junction information for mapping splices (--sjdbGTFfile)

Length of the genomic sequence around annotated junctions

100

Used in constructing the splice junctions database. Ideal value is ReadLength-1 (--sjdbOverhang)

Use 2-pass mapping for more sensitive novel splice junction discovery

History

search datasets

UDA

137 shown, 100 deleted, 6 hidden

79.81 GB

d others: mapped.bam

136: 0h\_III RNA STAR on data 130, data 129, and others: splice junctions.bed

135: 0h\_III RNA STAR on data 130, data 129, and others: log

133: 0h\_II RNA STAR on data 130, data 129, and others: mapped.bam

132: 0h\_II RNA STAR on data 130, data 129, and others: splice junctions.bed

131: 0h\_II RNA STAR on data 130, data 129, and others: log

130: trichophyton\_rubrum\_cbs\_118892\_2\_transcripts.gtf

129: trichophyton\_rubrum\_cbs\_118892\_2\_supercontigs.fasta

128: FastQC on data 100: RawData

127: FastQC on data 100: Webpage

126: FastQC on data 99: RawData

125: FastQC on data 99: Webpage

# RNA STAR – Via Galaxy (cont.)

The screenshot displays the Galaxy web interface for configuring the RNA STAR tool. The browser address bar shows 'usegalaxy.org'. The top navigation bar includes 'Analyze Data', 'Workflow', 'Visualize', 'Shared Data', 'Help', and 'User'. The left sidebar lists various tools under the 'Tools' section, including 'rna star', 'fastp', 'ABRicate List', 'BamLeftAlign', 'Cuffmerge', 'RNA STAR', 'RNA STARsolo', 'Trinity', 'STAR-Fusion', 'Align reads and estimate abundance', 'FEELnc', 'Salmon quant', 'Seurat', 'RNA Structure Prediction', 'QualiMap RNA-Seq QC', 'Remove Unwanted Variation', and 'sirna'. The main configuration area is titled 'Use 2-pass mapping for more sensitive novel splice junction discovery' and includes the following sections:

- Use 2-pass mapping for more sensitive novel splice junction discovery:** A dropdown menu set to 'No'. Below it, a paragraph explains that for multiple samples, 2-pass mapping is the most sensitive approach.
- Per gene/transcript output:** A dropdown menu set to 'No per gene or transcript output'. Below it, a paragraph explains that STAR can provide analysis results with respect to the reference genome, genes, and transcripts.
- Report chimeric alignments?:** A dropdown menu set to 'Don't report chimeric alignments'. Below it, a paragraph explains how chimeric alignments should be reported.
- BAM output format specification:** A section with a 'Read alignment tags to include in the BAM output' checkbox and a list of tags to include or unselect. The 'ch' tag (used to indicate chimeric alignments) is checked.
- HI tag values should be:** Radio buttons for 'one-based' (selected) and 'zero-based'. Below it, the text '(--outSAMattrIHstart)' is visible.
- MAPQ value for unique mappers:** A slider and input field set to '60'. Below it, a paragraph explains that STAR bases the mapping quality scores on the number of alternative mappings.
- Output filter criteria:** A section at the bottom of the configuration area.

The right sidebar shows the 'History' section with a search bar and a list of datasets. The datasets listed include:

- 136: 0h\_III RNA STAR on data 130, data 129, and others: splice junctions.bed
- 135: 0h\_III RNA STAR on data 130, data 129, and others: log
- 133: 0h\_II RNA STAR on data 130, data 129, and others: mapped.bam
- 132: 0h\_II RNA STAR on data 130, data 129, and others: splice junctions.bed
- 131: 0h\_II RNA STAR on data 130, data 129, and others: log
- 130: trichophyton\_rubrum\_cbs\_118892\_2\_transcripts.gtf
- 129: trichophyton\_rubrum\_cbs\_118892\_2\_supercontigs.fasta
- 128: FastQC on data 100: RawData
- 127: FastQC on data 100: Webpage
- 126: FastQC on data 99: RawData
- 125: FastQC on data 99: Webpage

# RNA STAR – Via Galaxy (cont.)

The screenshot displays the Galaxy web interface for the RNA STAR tool. The browser address bar shows `usegalaxy.org`. The top navigation bar includes links for `Analyze Data`, `Workflow`, `Visualize`, `Shared Data`, `Help`, and `User`. The left sidebar lists various tools, with `rna star` selected. The main configuration area is titled `Output filter criteria` and includes the following sections:

- Exclude the following records from the BAM output:**
  - Select/Unselect all
  - Unmapped reads
  - Alignments that have junctions with inconsistent strands
  - Alignments across unannotated non-canonical junctions
  - All alignments across non-canonical junctions (recommended for compatibility with Cufflinks)
- Would you like to set additional output filters?:** No
- Algorithmic settings:** Configure seed, alignment and limits options (Use Defaults)
- Performance tweaks / Troubleshooting:** (icon)
- Email notification:** No (Send an email notification when the job completes.)
- Execute:** (button)
- What it does:** STAR is an ultrafast universal RNA-seq aligner.
- Compatibility Notes:** STAR has a huge amount of options to filter alignments and to configure the exact format of its output. Some tools you may plan to use in your downstream analysis of the results are known to be sensitive to these settings or combinations of them.
- STAR-Fusion:** STAR-Fusion can use the chimeric junctions output of STAR as input, but you need to enable **chimeric alignment detection** by STAR for that dataset to be generated. Hence, be sure to select: **Report chimeric alignments?: As separate tabular "Junctions" output (Junctions).**
- In addition, for best results it is recommended that you:**
  - use **2-pass mapping** for more sensitive novel splice junction discovery
  - under **BAM output format specification**, **Read alignment tags to include in the BAM output**: select `XS` as an additional tag to generate (this is the equivalent of using `--outSAMstrandField intronMotif` on the command line)
  - under **Algorithmic settings**, **Configure seed, alignment and limits options**: use *parameters suggested for STAR-Fusion*.
- Cufflinks:** (icon)

The right sidebar shows the **History** panel with a search bar and a list of jobs. The jobs listed include:

- 136: 0h\_III RNA STAR on data 130, data 129, and others: splice junctions.bed
- 135: 0h\_III RNA STAR on data 130, data 129, and others: log
- 133: 0h\_II RNA STAR on data 130, data 129, and others: mapped.bam
- 132: 0h\_II RNA STAR on data 130, data 129, and others: splice junctions.bed
- 131: 0h\_II RNA STAR on data 130, data 129, and others: log
- 130: trichophyton\_rubrum\_cbs\_118892\_2\_transcripts.gtf
- 129: trichophyton\_rubrum\_cbs\_118892\_2\_supercontigs.fasta
- 128: FastQC on data 100: RawData
- 127: FastQC on data 100: Webpage
- 126: FastQC on data 99: RawData
- 125: FastQC on data 99: Webpage

# Exemplo de Resultado RNA STAR (Log)

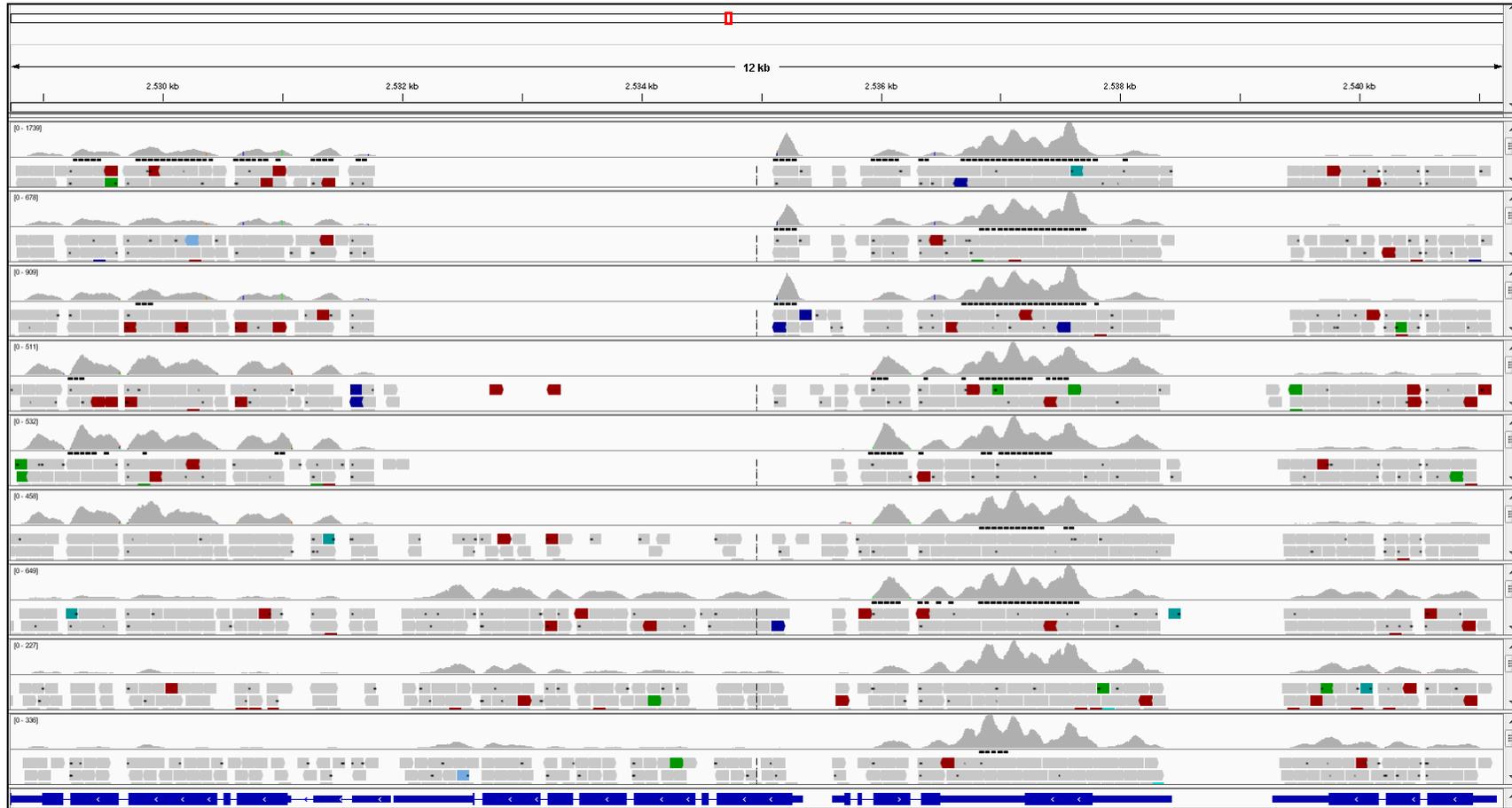
Started job on		Apr 09 17:17:25
Started mapping on		Apr 09 17:17:56
Finished on		Apr 09 17:50:15
Mapping speed, Million of reads per hour		49.76
Number of input reads		26800883
Average input read length		91
UNIQUE READS:		
Uniquely mapped reads number		23563054
Uniquely mapped reads %		87.92%
Average mapped length		91.49
Number of splices: Total		3281588
Number of splices: Annotated (sjdb)		2879421
Number of splices: GT/AG		3259914
Number of splices: GC/AG		11595
Number of splices: AT/AC		795
Number of splices: Non-canonical		9284
Mismatch rate per base, %		0.18%
Deletion rate per base		0.00%
Deletion average length		1.30
Insertion rate per base		0.00%
Insertion average length		1.07
MULTI-MAPPING READS:		
Number of reads mapped to multiple loci		0
% of reads mapped to multiple loci		0.00%
Number of reads mapped to too many loci		115902
% of reads mapped to too many loci		0.43%
UNMAPPED READS:		
Number of reads unmapped: too many mismatches		0
% of reads unmapped: too many mismatches		0.00%
Number of reads unmapped: too short		3049969
% of reads unmapped: too short		11.38%
Number of reads unmapped: other		71958
% of reads unmapped: other		0.27%
CHIMERIC READS:		
Number of chimeric reads		0
% of chimeric reads		0.00%

Nro. Reads de entrada  
(já trimados)

Nro. Reads mapeadas



# IGV – Software para visualização gráfica de um arquivo .bam



# Finalizar a tabela de resultados

**Table 1.** General features of RNA-seq reads mapped to the *T. rubrum* reference genome

Sample	Raw reads	High-quality reads	Mapped reads STAR	Total mapped reads (%)
0h II				
0h III				
3h II				
3h III				
12h II				
12h III				