



# Les évolutions de Galaxy

Interface et sessions de formations



## Des outils plus complets

<http://galaxy-workbench.toulouse.inra.fr/>

**Galaxy Workbench** Analyze Data Workflow Shared Data Visualization Admin Help User Welcome smaman Using 37%

**Tools** Options

- FASTA manipulation
- FASTQ manipulation
- SAM/BAM manipulation : Picard (beta)
- SAM/BAM manipulation : SAMtools
- Fetch Sequences
- Sequences Queries
- 4 - SGS MAPPING
- BWA - Bowtie
- 5 - SNP / INDEL
- GATK Tools (beta)
- SAMtools
- Indel Analysis
- 6 - TRANSCRIPTOMIC
- RNAseq
- RNAseq ALIGNMENT
  - \* Tophat for Illumina Find splice junctions using RNA-seq data
  - RNAseq RAW EXPRESSION
    - \* htseq count
    - \* Merge Htseq count output file into a global counting file
    - \* Sigdifflinks to obtain raw count of reads
    - \* Merge sigdifflinks count file
- 8 - TRAININGS
- Galaxy Initiation
- Reads alignment and SNP calling
- RNA-Seq
- sRNAseq
- SNP annotation

**\* EN COURS DE TEST \* Tophat for Illumina (version 1.0.0)**

Your RNA-Seq FASTQ file (read 1):  
2: M\_Pf\_2\_ACAGTG\_L00..fastqsanger

Your RNA-Seq FASTQ file (read 2):  
2: M\_Pf\_2\_ACAGTG\_L00..fastqsanger

Your RNA-seq FASTQ file are zipped:  
☐ Yes  
Please check this option if your files are zipped.

Choose your reference genome:  
Select a reference genome  
Please choose either use a bank available on your BioInfo Genotoul Plateform or use your own FASTA reference file (this FASTA file will automatically be indexed by Galaxy)

Select a reference genome:  
Danio rerio Zv9 62 chr 22

Number of threads used to align reads:  
16

Maximum intron length:  
5000

Expected (mean) inner distance between mate pairs:  
200

More options ?:  
No more option  
Please choose Show if you want to see more options.

Execute

What is Tophat? reads to a genome in order to identify exon-exon splice junctions. It is built on the ultrafast short read mapping program Bowtie. TopHat was designed to work with reads produced by the Illumina Genome Analyzer, although users have been successful in using TopHat with reads from other technologies. In 1.1.0, we began supporting Applied Biosystems' Colospace format. The software is optimized for reads 75bp or longer.

Mixing paired- and single- end reads together is not supported.

How does TopHat find junctions?

TopHat can find splice junctions without a reference annotation. By first mapping RNA-Seq reads to the genome, TopHat identifies potential exons, since many RNA-Seq reads will contiguously align to the genome. Using this initial mapping information, TopHat builds a database of possible splice junctions and then maps the reads against these junctions to them.

Short read sequencing machines can currently produce reads 100bp or longer but many exons are shorter than this so they would be missed in the initial mapping. TopHat solves this problem mainly by splitting all input reads into smaller segments which are then mapped independently. The segment alignments are put back together in a final step of the program to produce the end-to-end read alignments.

TopHat generates its database of possible splice junctions from two sources of evidence. The first and strongest source of evidence for a splice junction is when two segments from the same read (for reads of at least 45bp) are mapped at a certain distance on the same genomic sequence or when an internal segment fails to map - again suggesting that such reads are spanning multiple exons. With this approach, "GT-AG", "GC-AG" and "AT-AC" introns will be found ab initio. The second source is pairings of "coverage islands", which are distinct regions of

**History** Options

Test / Phylofish 3.9 Gb

15:  
{M\_Pf\_2\_ACAGTG\_L007\_R2.fastqsanger}-BWA.bam  
-head.txt  
100 lines  
format: txt, database: 2  
Info: Epilog : job finished at mar. juil. 16 11:35:24 CEST 2013

85Q VW:1.3 50:coordinate  
85Q SN:ENSDART00000112953 IN:2064  
85Q SN:ENSDART00000122537 IN:579  
85Q SN:ENSDART00000129800 IN:810  
85Q SN:ENSDART00000099219 IN:2553  
85Q SN:ENSDART00000079354 IN:1245

14:  
{M\_Pf\_2\_ACAGTG\_L007\_R2.fastqsanger}-BWA.bam  
8.7 Mb  
format: bam, database: 2  
Info: Etape 1  
Indexation : /usr/local/bioinfo/bin/bwa index -a is /work/galaxy/database/files/011/dataset\_11741.dat >> ./bwaindex.log 2>&1  
Etape 2  
Alignement du premier fastq : /usr/local/bioinfo/bin/bwa -in /work/galaxy/database/files/011/dataset\_11741.d

Binary bam alignment file

ENSDART00000112953 2064 0 0  
ENSDART00000122537 579 0 0  
ENSDART00000129800 810 0 0  
ENSDART00000099219 2553 0 0  
ENSDART00000079354 1245 0 0  
ENSDART0000024641 1605 0 0

**(\*) Outils Siganae**

junctions using RNA-seq data

sRNAseq

7 - CHIP-SEQ

Operate on Genomic Intervals

Nebula

8 - TRAININGS

Galaxy Initiation

Reads alignment and SNP calling

RNA-Seq

sRNAseq


SNP annotation

**Accès à plus d'options de la ligne de commande**

**Affichage de la ligne de commande et des étapes de traitement**



[Home](#) [About us](#) [Resources](#) [Services](#) [Help](#) [Login](#)



## Event Detail

You are here: » [Services](#) » [Training](#)

### Galaxy training days

This session group the 4 trainings into 3 days :

- Galaxy : First step
- Galaxy : Reads alignment and SNP calling
- Galaxy : RNAseq alignment and transcripts assemblies
- Galaxy : sRNAseq

You will learn how to use galaxy environment, NGS standard formats and handle different type of data.

- Date : 06.-08.11.2013
- Number : 43
- Time : will be announced
- Place :  
Computer training room  
INRA Auzeville, Auzeville
- Academic : € 450,00
- Private : € 1.500,00
- Organizers**  
Sarah Maman  
Philippe Bardou  
Delphine Labourdette  
Jerome Mariette  
Christophe Klopp  
**Organizing partners**  
Monique Falieres
- Vacancies : 6

**Registration : [Register now](#)**

[Back to listing](#)

<http://bioinfo.genotoul.fr>