

Durée : 2 journées de cours / TP / TD puis 1/2 journées de travail en autonomie pour la réalisation d'un « mini-projet Galaxy ».

Programme :

Présentation des plateformes et des équipes

Initiation à l'interface Galaxy

Administration et wrappers Galaxy

Pipelines de traitements (bio) et développement de wrappers (info)

Autres sessions Genotoul disponibles : <http://bioinfo.genotoul.fr>

Votre intervenante :

Sarah Maman

Equipe Sigeneae

Ingénieur d'études en bioinformatique

Présentation des plateformes et des équipes

GIS Genotoul

Créé en 2000



PF Genotoul est l'une des 13 plateformes bioinformatiques IBISA.

Depuis 2008, BioInfo Genotoul travaille en collaboration avec la plateforme génomique pour traiter d'énormes volumes de données générés par la nouvelle génération de séquenceurs.

Les fichiers FASTQ produits sont mis à disposition des biologistes via NG6.

Contact us

PROJECTS RUNS DOWNLOAD FAQ

User login

Enter your username and password here in order to log in on the website:

Login

Username:

Password:

Login

Keep up with news

- June 17 2013** NG6 v2.0 is now available. This new version is based upon the Jflow workflow engine instead of ergatis. This version is coming with several new features for runs administrators.
- November 27 2012** NG6 is now only available in english.
- September 14 2012** New user management system. 3 rights levels: administrator (in charge to run workflows), manager (in charge to manage project access) and member (browsing projects/runs/analyses).
- Septembrer 9 2012** Publication of NG6 in [BMC Genomics](#).
- August 7 2012** NG6 is available in french and in english.
- June 4 2012** Fix a bug in the generation of cigarline graphs produced by the AlignmentStats analysis.
- November 23 2011** Since the 3rd of october 2011 the HiSeq quality encoding is in Sanger format, no longer in illumina format.
- August 2nd 2011** New functionalities are available to browse your projects/runs/analyses. A hierarchical display is now available for analysis. For project administrators new functionalities are also available.

Accès à de très hautes performances de calcul : le cluster Genotoul.

Nécessiter de créer un compte pour accéder à ce cluster de calcul.

Utilisation de comptes spécifiques de formation lors de cette semaine.

Si besoin:

- Pour les mini-projets : penser à demander une prolongation de compte.
- Pour les stages : ouvrir un compte avec votre adresse académique (pas de mail perso).

Un fois loggé, vous avez accès:

- * Nombreuses banques de données biologiques
- * Nombreux scripts et outils bioinformatiques.
- * Accès à des sessions de formations
- * Services publics :
 - Hébergement de sites web
 - Hébergement de machines virtuelles (VM).
 - Expertises et support scientifique (biologie / bioinformatique).





The team



Christine Gaspin
DR INRA / Scientific animation
+33 (0)5 61 28 52 82
christine.gaspin@toulouse.inra.fr



Céline Noirot
IE INRA / Development and data analysis
+33 (0)5 61 28 57 24
celine.noirot@toulouse.inra.fr



Didier Laborie
IE INRA / System administrator
+33 (0)5 61 28 54 27
didier.laborie@toulouse.inra.fr



Marie-Stéphane Trotard
IE INRA / System administrator
+33 (0)5 61 28 52 76
marie-stephane.trotard@toulouse.inra.fr

Gaëlle Lefort
IE Biostat - Nov 2015 - Mai 2016
Financement Genotoul pour PF Biostat/PF Bioinfo
Gaëlle.Lefort@toulouse.inra.fr
Alexandre Heurteau
IE Bioinfo/Biostat - Mars 2016 - Août 2016
Financement GenEndurance
alexandre.heurteau@toulouse.inra.fr



Christophe Klopp
IR INRA / Technical animation
+33 (0)5 61 28 50 36
christophe.klopp@toulouse.inra.fr



Claire Hoede
IR INRA / Development and data analysis
+33 (0)5 61 28 53 05
claire.hoede@toulouse.inra.fr



Jérôme Mariette
IE INRA / Development and data analysis
+33 (0)5 61 28 57 25
jerome.mariette@toulouse.inra.fr



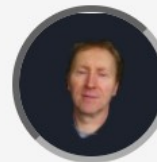
Frédéric Escudié
IE France Génomique / Development and data analysis
+33 (0)5 61 28 55 49
frederic.escudie@toulouse.inra.fr



Régis Ongaro
IE France Génomique / Development and data analysis
+33 (0)5 61 28 57 45
Regis.Ongaro@toulouse.inra.fr

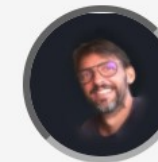
SIGENAE TEAM

The members of the group are part of four INRAE divisions (GA, PHASE, MATHNUM and SA).



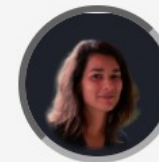
Christophe Klopp
PROJECT COORDINATOR

Genomic, Transcriptomic,
Assembly
Data processing, AWK, Perl,
Python, R.



Philippe Bardou
BIOINFORMATICS ENGINEER

DNaseq, Variant calling,
sRNAseq (miRNA)
Perl, Python, GIT, HTML5,
CSS3, Javascript.



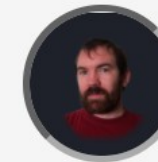
Maria Bernard
BIOINFORMATICS ENGINEER

Variant calling, RADseq,
Metagenomic, RNASeq, ASE
Python, Snakemake, GIT,
Rmarkdown, Galaxy.



Cedric Cabau
BIOINFORMATICS ENGINEER

RNAseq, Assembly,
Annotation
Perl, Awk, Python, GIT,
HTML/CSS/JS.



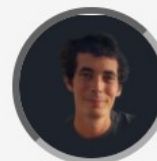
Mathieu Charles
BIOINFORMATICS ENGINEER

Variant calling, CNV,
RNAseq
Perl, Python, Java, C++,
Snakemake, Jupyter, R.



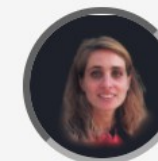
Patrice Dehais
SYSADMIN AND B.E.

Sysadmin HPC Virtualization
Parallelization/Optimization
applied to bioinformatics
software.



Cervin Guyomar
BIOINFORMATICS ENGINEER

Genomics, metagenomics,
functional genomics
R, Python, NextFlow



Sarah Maman
BIOINFORMATICS ENGINEER

RNAseq, Galaxy, elearning
Perl, Python, Cheatah,
Nextflow



Margot Zahm
BIOINFORMATICS ENGINEER

De novo genome assembly,
Genome annotation
Python, Nextflow, R, GIT

DEVELOPMENT PROJECT

The Sigenae team develops or is involved in the development of many tools, databases and workflows.
See below some examples of significant achievements.

ALL TOOLS DATABASES WORKFLOWS



De novo RNA-Seq Assembly Pipeline

Short read RNA-Seq de novo assembly is a well established method to study transcription of organisms lacking a reference genome sequence. Available software packages such as Trinity and Oases have proven to be able to build high quality contigs from short reads. But there is still room for improvement on different points such as:

- compactness: they often produce different contigs which are included in one another or overlapping one another.
- redundancy: the contigs contain different reads or intervals such as duplicated open reading frames.
- substitution, insertion, deletion errors: the consensus sequences held by the assembler contain errors which can be partly corrected using the read alignments.

Inputs and parameters

Parameters: Background average, Density kernel width, Range extension.

Doing vector: Mean proportion.

Species	Height	Weight	No. of legs	Puberty height	Puberty weight	1st lactation height	1st lactation weight
BlackBoschardt 10	1400	100000	4	1000	10000	1400	100000
BlackBoschardt 15	1500	110000	4	1100	11000	1500	110000
BlackBoschardt 20	1600	120000	4	1200	12000	1600	120000

Available and vector detection table with columns: Species, Available, Vector, Height, Weight, No. of legs, Puberty height, Puberty weight, 1st lactation height, 1st lactation weight.



High

Input: Species, Height, Weight, No. of legs, Puberty height, Puberty weight, 1st lactation height, 1st lactation weight.

Output: High (table with columns: Species, Height, Weight, No. of legs, Puberty height, Puberty weight, 1st lactation height, 1st lactation weight).

D-Genies

It allows to compare two genomes. It supports large genomes and you can interact with the data to improve the visualization.

We use reference version 2 to align the two genomes. Then the FM file is sorted and stored into an intermediate sort order with 4.5 kb bins.

D-Genies also allows to display dot plots from other aligners like Bowtie2 or BWA.

PIPELINES

Tools: Trinity, Oases, Velvet, SOAPdenovo, SPAdes, SPARKdenovo, SPARKdenovo2, SPARKdenovo3, SPARKdenovo4, SPARKdenovo5, SPARKdenovo6, SPARKdenovo7, SPARKdenovo8, SPARKdenovo9, SPARKdenovo10.

Galaxy

Tools

search tools

MANAGE YOUR DATA FILES

- Get Data
- Download Data
- Jobs statistics

FILES MANIPULATION

- Text Manipulation
- Filter and Sort
- Join, Subtract and Group
- GFF
- BED Tools
- Convert Formats
- Fetch Sequences
- Statistics
- Graph/Display Data

SEQUENCES MANIPULATION

- FASTA manipulation
- FastQC: fastq/sam/bam
- Illumina fastq
- Generic FASTQ manipulation
- FASTX-Toolkit for FASTQ data

SAM/BAM MANIPULATION : PICARD (BETA)

- Conversion
- QC/Metrics for sam/bam
- BAM/SAM Cleaning
- SAM/BAM manipulation: SAMtools
- Sequences Queries
- VCF Tools

SGS MAPPING

- BLAST
- BWA - Bowtie
- Indel Analysis
- Variant calling
- SNP annotation
- RNAseq Alignment
- RNAseq Raw Expression
- RNAseq Cufflinks



How to cite Galaxy workbench ?

Depending on the help provided you can cite us in acknowledgements, references or both.

Examples :

Research teams can thank the Toulouse Midi-Pyrenees bioinformatics platform and Sigenae group, using in their publications the following sentence : "We are grateful to the genotoul bioinformatics platform Toulouse Midi-Pyrenees and Sigenae group for providing help and/or computing and/or storage ressources thanks to Galaxy instance <http://sigenae-workbench.toulouse.inra.fr/>".

In cases of collaboration, you can directly quote the person who participated to the project : Name, Sigenae group, GenPhySE, INRA Auzeville CS 52627 31326 Castanet Tolosan cedex.

References
X. SIGENAE [<http://www.sigenae.org/>]

History

search datasets

TESTIS-PIG
7 shown, 110 deleted

78.98 GB

109: /work/robic/Reads_externes_2020/TESTIS/Pig/ERR3418012_2.fastq.gz				
6: ERR3418012_1.fastq.gz				
5: ERR3417952_2.fastq.gz				
4: ERR3417952_1.fastq.gz				
3: ERR3417904_2.fastq.gz				
2: ERR3417904_1.fastq.gz				
1: Upload				



Ready to use Workflows

- 1/ FROGS - Find, Rapidly, Otus with Galaxy Solution
- 2/ Tax4Fun workflow
- 3/ Function Table for Tax4Fun matrix workflow



Galaxy News

- 1/ How to use datasets collections ?
- 2/ How to upload several files in Galaxy ?
- 3/ New tools : Tax4Fun (manual), VIP (manual), Salmon, multiQC.
- 4/ Tools availables for virologie analyses
 - 5/ Virology tools
- 6/ How to use Function Table tool ?
- 7/ How to export and import your histories ?



E-learning

<https://inra.classila.com/Login>

2 e-learning availables : Galaxy initiation and Quality sequences analysis

Training ressources : [BioInfo Genotoul / Sigenae](#)

Merci pour votre écoute

-

Remerciements



Fonds Européen
de Développement Régional

