GREAT - GRape Expression ATlas

Analysis workflow & web application to analyse
Vitis vinifera public RNA-seq data

Amandine Velt, Lauriane Renault, Camille Rustenholz

INRAE

GREAT – Grape Expression ATlas

Operating mode

Calculation server

GREAT database

CSV

1

2

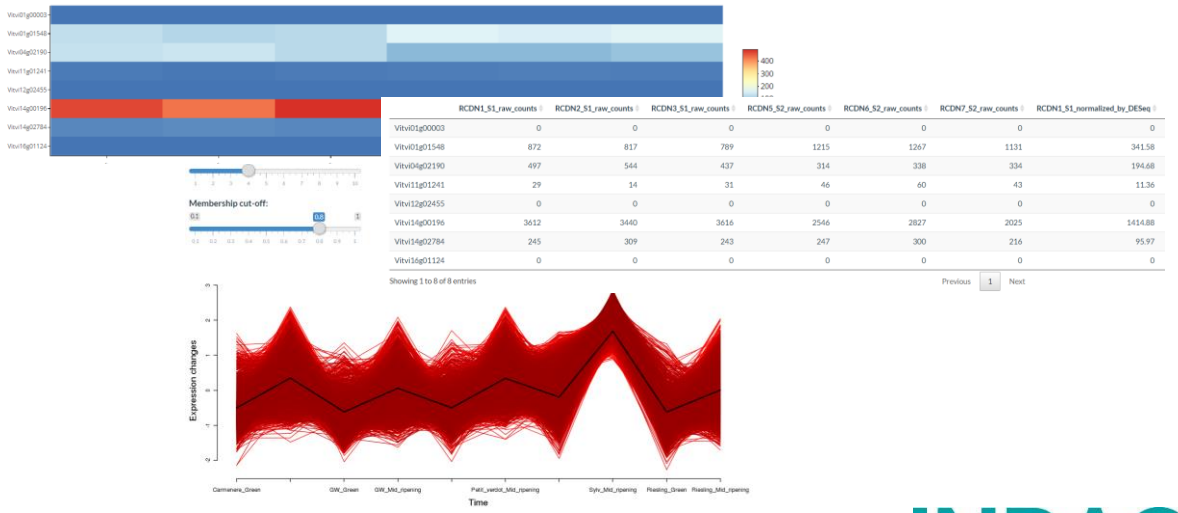
3

ATCGATGAC
AGTCAGTAG
ACGTACTAG
ACTAGCAAA

SRA database

GREAT analysis workflow
(Snakemake)

Data visualisation
(GREAT web application
(R Shiny))



GREAT – Grape Expression ATlas

Operating mode

SRA database

SRA database

Sequence Read Archive (sequencing raw data storage)

Public repository for raw sequencing data

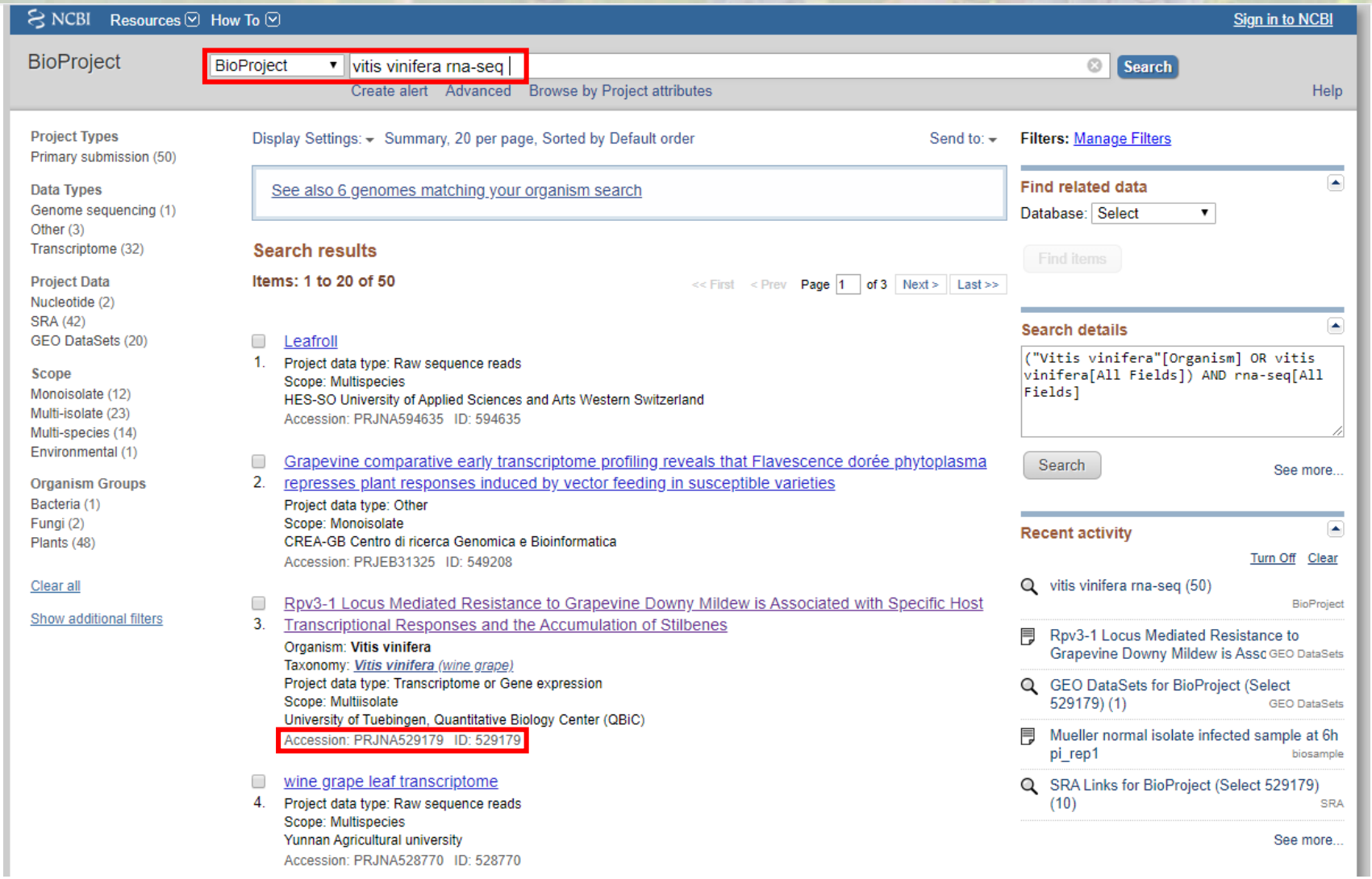
Raw sequencing data used in a publication are usually published on the SRA (NCBI) or ENA (EBI)

SRA provides a tool, SRA Toolkit, to recover data from project ID

The SRA uses the .sra format, more suitable for the amount of data it stores -> easily convertible to fastq

SRA database : search

https://www.ncbi.nlm.nih.gov/sra



SRA database : run selector

https://www.ncbi.nlm.nih.gov/Traces/study/

NCBI **SRA Run Selector** Log in to NIH

Filters List

1 Bases
2 BioSample
3 Bytes
4 Cultivar
5 Experiment
6 GEO_Accession
7 inoculation
8 phenotype
9 Run
10 Sample Name
11 treatment

Accession PRJNA529179 Search

Common Fields

BioProject	PRJNA529179
Consent	PUBLIC
Assay Type	RNA-Seq
AvgSpotLen	75
Center Name	GEO
DATASTORE filetype	FASTQ, SRA
DATASTORE provider	GS, NCBI, S3
DATASTORE region	gs.US, ncbi.public, s3.us-east-1
Instrument	NextSeq 500

Select

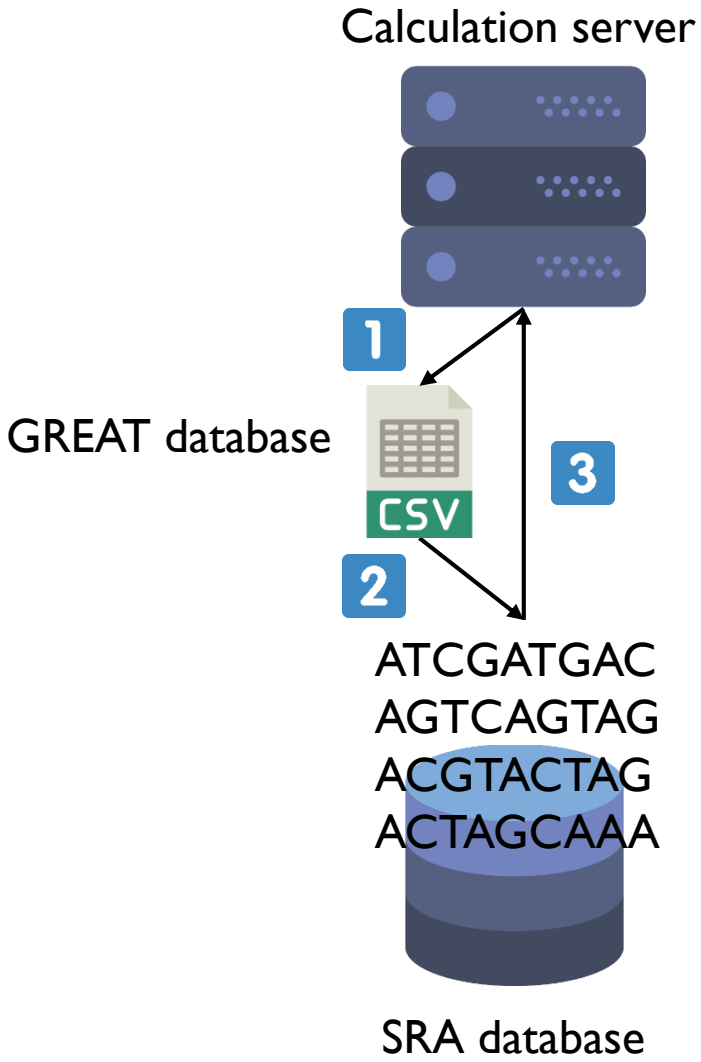
	Runs	Bytes	Bases	Download
Total	10	6.26 Gb	14.54 G	Metadata or Accession List
	0	0	0	Metadata or Accession List or JWT Cart

Found 10 Items

	Run	BioSample	Bases	Bytes	Cultivar	Experiment	GEO_Accession	inoculation	phenotype	Sample Name	treatment
1	SRR8786886	SAMN11258051	1.36 G	600.16 Mb	Regent	SRX5576720	GSM3687409	avirulent P. viticola isolate avrRpv3+	resistant	GSM3687409	x
2	SRR8786888	SAMN11258048	1.71 G	753.38 Mb	Regent	SRX5576722	GSM3687411	virulent P. viticola isolate avrRpv3-	resistant	GSM3687411	x_aggressive
3	SRR8786882	SAMN11258055	1.35 G	593.28 Mb	Mueller	SRX5576716	GSM3687405	avirulent P. viticola isolate avrRpv3+	susceptible	GSM3687405	x
4	SRR8786889	SAMN11258047	1.25 G	553.48 Mb	Regent	SRX5576723	GSM3687412	virulent P. viticola isolate avrRpv3-	resistant	GSM3687412	x_aggressive
5	SRR8786884	SAMN11258053	1.72 G	756.26 Mb	Mueller	SRX5576718	GSM3687407	mock control (water)	susceptible	GSM3687407	m
6	SRR8786887	SAMN11258050	1.17 G	515.33 Mb	Regent	SRX5576721	GSM3687410	avirulent P. viticola isolate avrRpv3+	resistant	GSM3687410	x
7	SRR8786891	SAMN11258044	1.35 G	593.97 Mb	Regent	SRX5576725	GSM3687414	mock control (water)	resistant	GSM3687414	m
8	SRR8786890	SAMN11258045	1.25 G	553.73 Mb	Regent	SRX5576724	GSM3687413	mock control (water)	resistant	GSM3687413	m
9	SRR8786883	SAMN11258054	1.53 G	669.77 Mb	Mueller	SRX5576717	GSM3687406	avirulent P. viticola isolate avrRpv3+	susceptible	GSM3687406	x
10	SRR8786885	SAMN11258052	1.86 G	819.17 Mb	Mueller	SRX5576719	GSM3687408	mock control (water)	susceptible	GSM3687408	m

GREAT – Grape Expression ATlas

Operating mode



GREAT's database

Grouping information about all public RNA-seq samples for Vitis vinifera

Excel file created and maintained by Lauriane Renault

GREAT's database update every 6 months and creation of a more complete version of GREAT

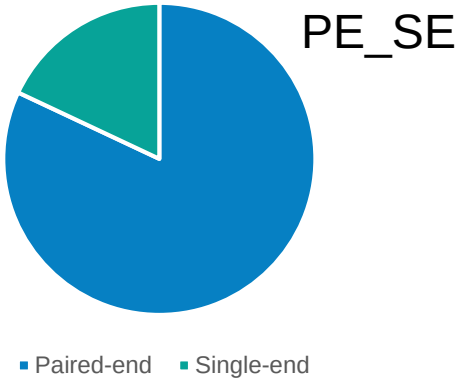
information collected :

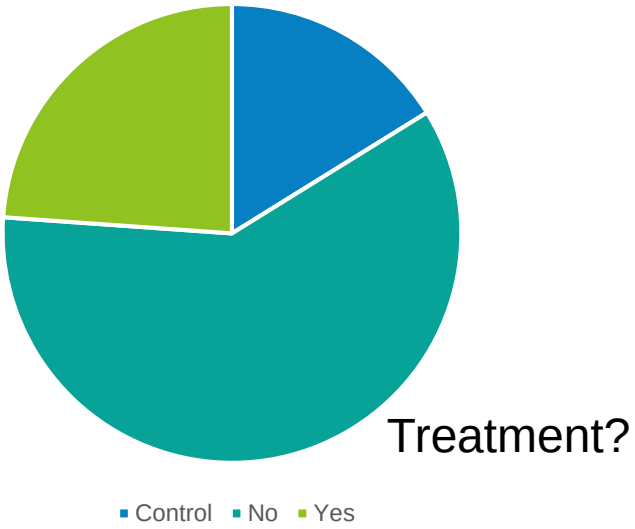
Bioproject	Protocol library	Treatment?
First author	Stranded	Treatment type
Paper	Species	Treatment nature
File name	Variety	Treatment detail
Sequencing platform	Organ	Comments
PE_SE	Tissue	Quality check
Read length bp	Stage_detail	

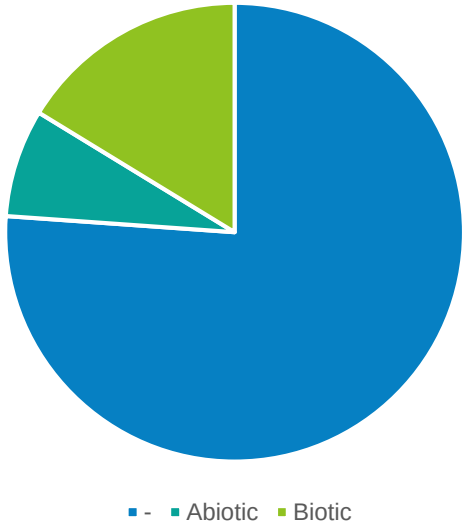
GREAT's database

Grouping information about all public RNA-seq samples for Vitis vinifera

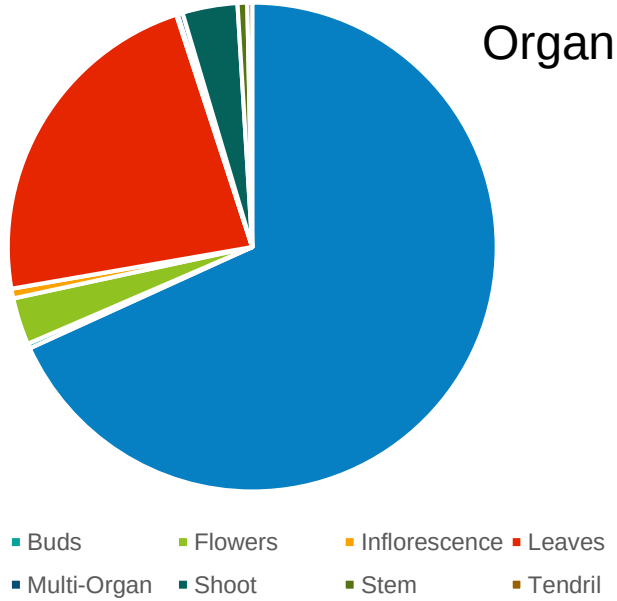
915 samples : identified, annotated and analysed







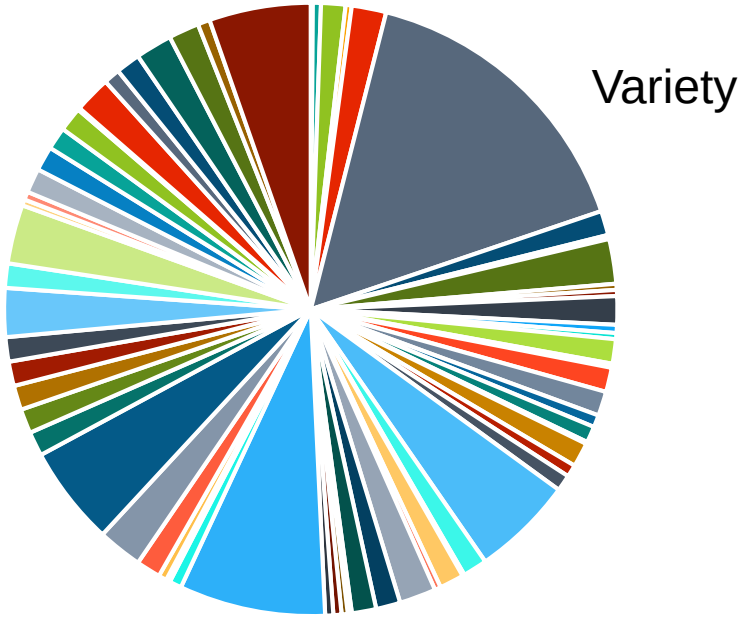
Treatment type



GREAT's database

Grouping information about all public RNA-seq samples for Vitis vinifera

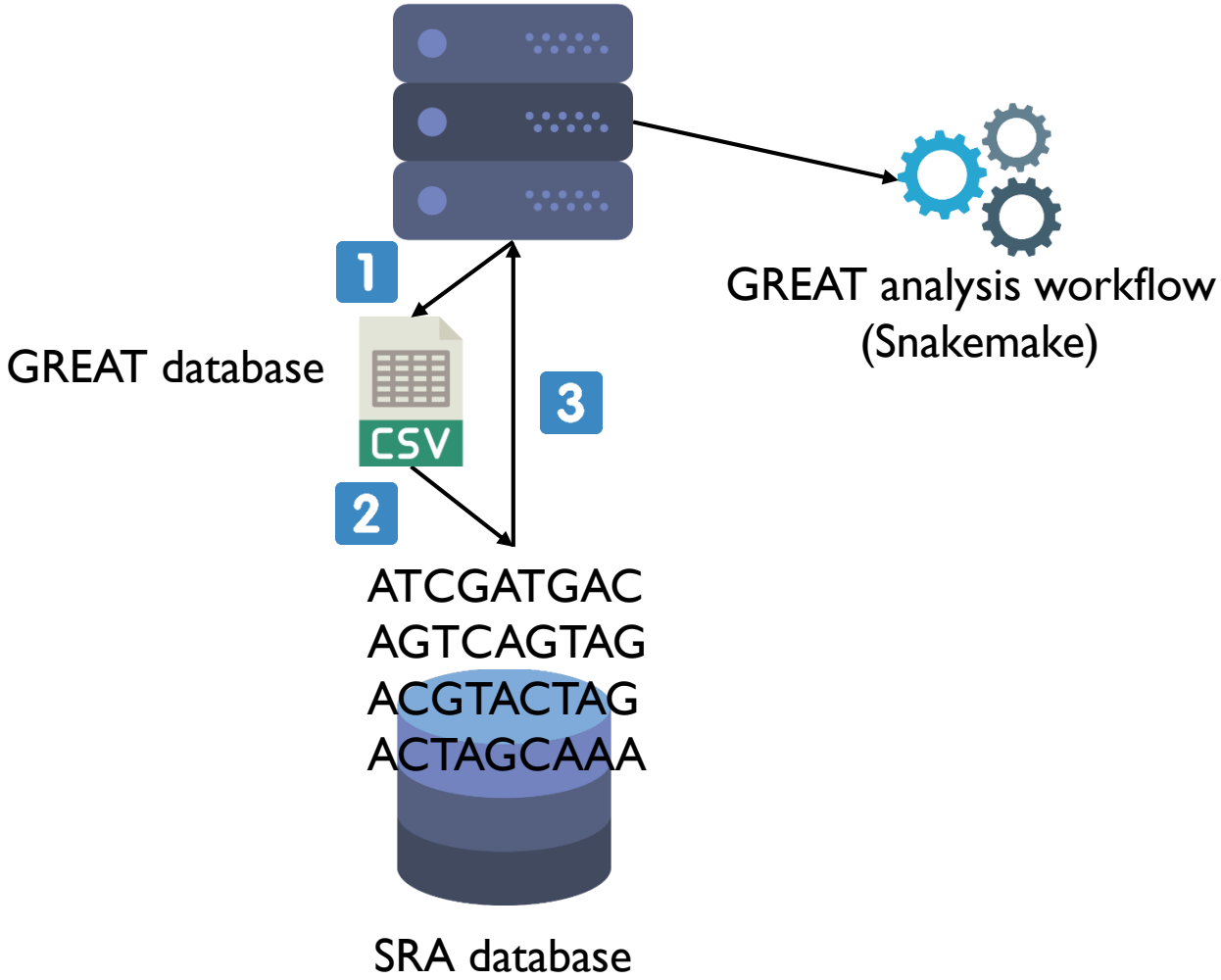
915 samples : identified, annotated and analysed



- Alicante Bouschet
- Bangalore blue
- Barbera
- Brachetto
- Cabernet Franc
- Cabernet Sauvignon
- Carignan
- Carménère
- Chardonnay
- Chasselas Blanc
- Chasselas cioutat
- Corvina
- Dilkhush
- Fujiminori
- Garganega
- Gewurztraminer
- Glera
- Jingxiangyu
- Koshu
- L1 Pinot Meunier microvine
- Lambrusco
- L Carignane
- Loureira
- Manicure Finger
- Merlot
- Moscato Bianco
- Moscato Rosa
- Muscat Blanc Petits Grains
- Muscat Hamburg
- Muscat Superior
- Negro amaro
- Passerina
- Petit Verdot
- picovine 06C008V0003
- picovine 06C021V0046
- Pinot Meunier L2
- Pinot Noir
- Prieto Picudo
- Primitivo
- Red Globe
- Refosco
- Riesling
- Sangiovese
- Sativa Husseine
- Sativa Karadzhandal
- Sativa Khalchili
- Sativa Late Vavilov
- Sativa Sochal
- Sauvignon blanc
- Sauvignon Blanc
- Semillon
- Shine Muscat
- Shiraz
- Sylvestris DVIT3351.27
- Sylvestris O34-16
- Tannat
- Tempranillo
- Teroldego
- Tocai friulano (Sauvignonasse)
- Touriga Nacional
- Vermentino
- Victoria
- Vitis sylvestris
- Vitis sylvestris C1-2
- Zinfandel

GREAT – Grape Expression ATlas

Operating mode



GREAT - step 1 : RNA-seq analysis workflow

Developed with Snakemake

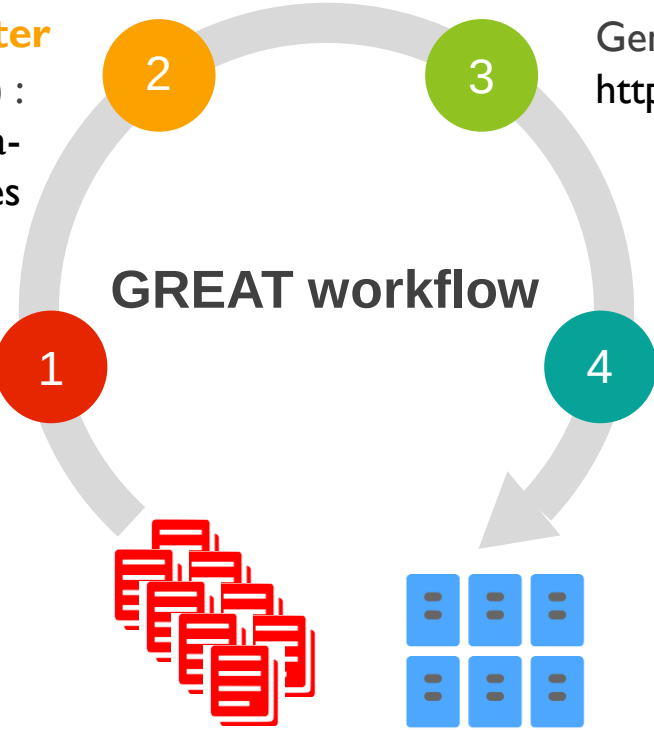
Snakemake = Python extension to create workflows
GREAT workflow available on gitlab : https://gitlab.com/avelt/great_rnaseq_workflow

Conversion to fastq + quality control (fastQC)

This step allows to validate or not the samples

Read alignment with STAR + quality filter

Alignment on PN40024 (PN12XV2) : https://urgi.versailles.inra.fr/Species/Vitis/Data-Sequences/Genome-sequences



Genes quantification with featureCounts

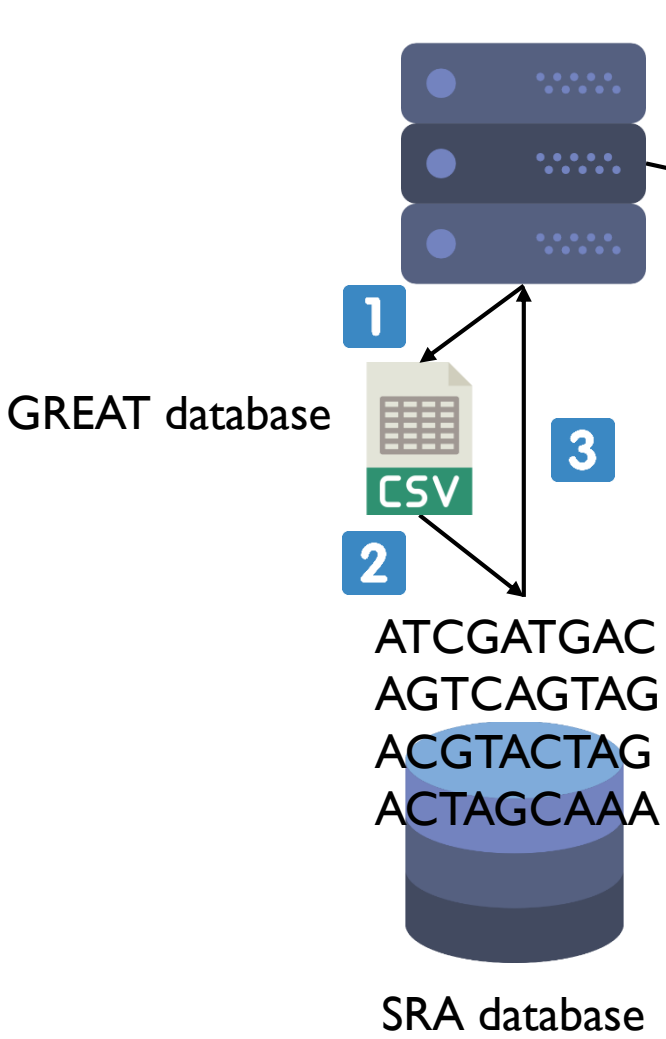
Genes annotation Vcost V3 of PN12XV2 : https://urgi.versailles.inra.fr/Species/Vitis/Annotations

Creation of final files for the R Shiny application

- Normalization of count data & samples merging
- Calculation of correlation coefficients
- Variance-stabilizing transformation (for PCA)

GREAT – Grape Expression ATlas

Operating mode

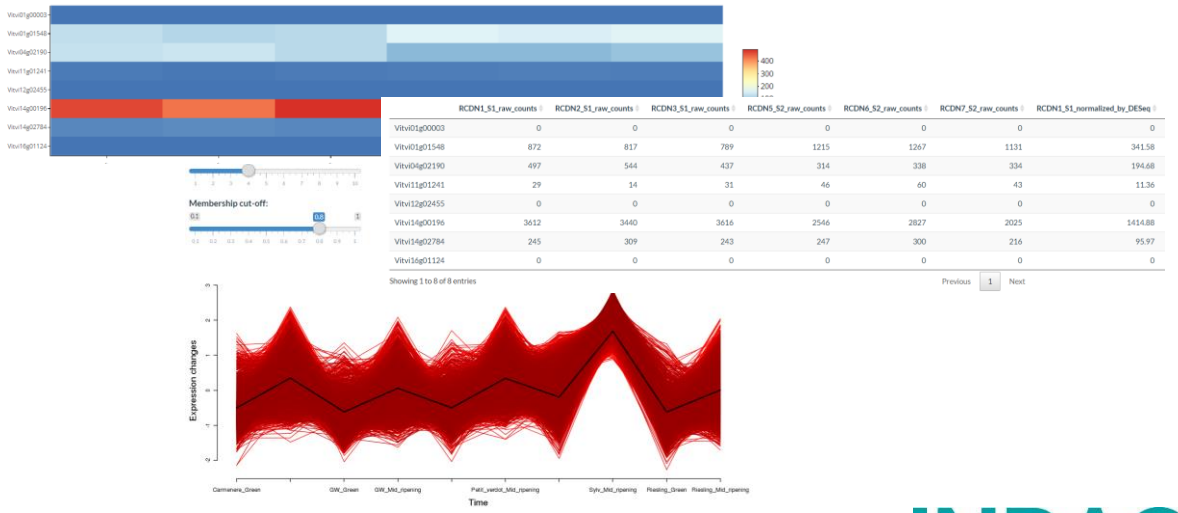


GREAT database

SRA database

GREAT analysis workflow
(Snakemake)

Data visualisation
(GREAT application (R Shiny))



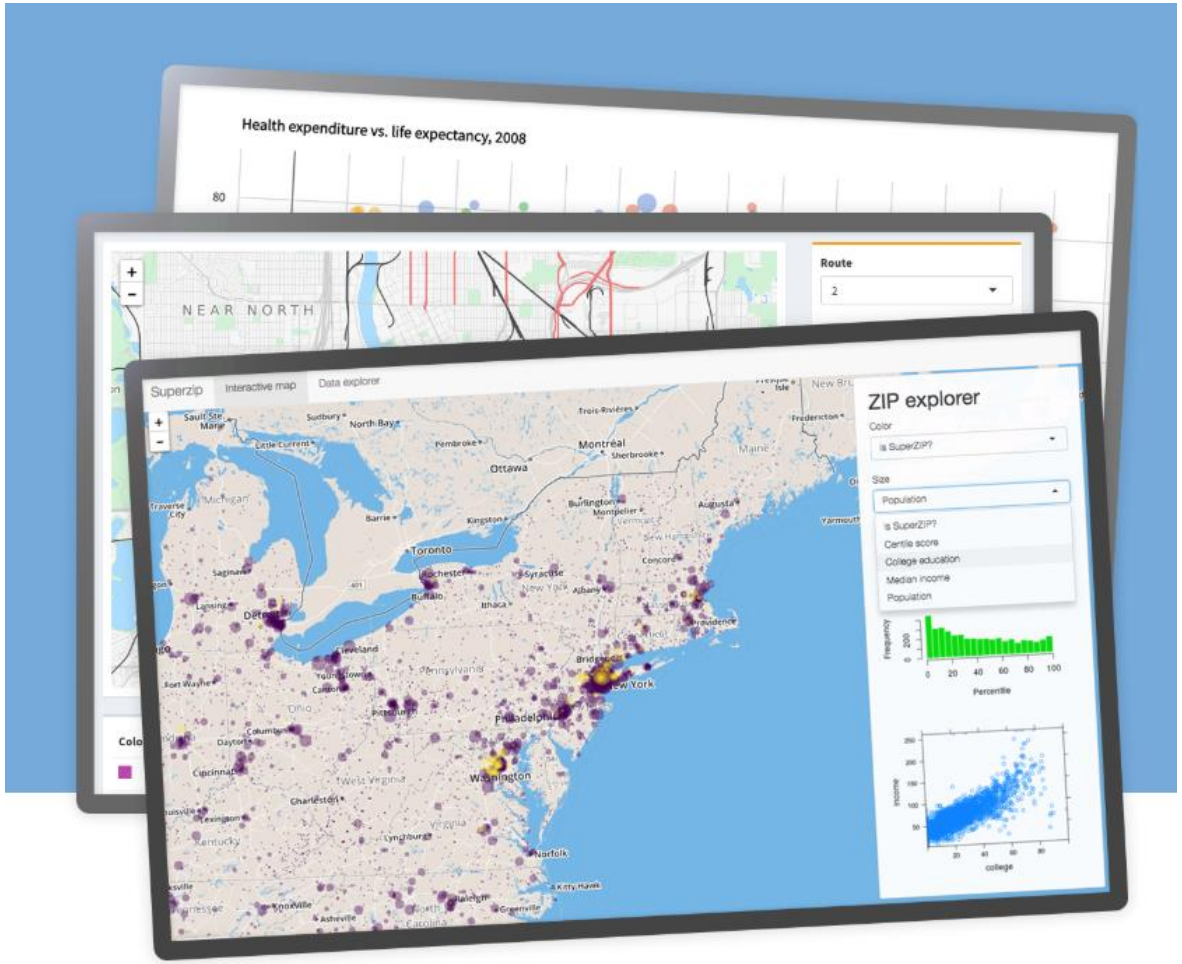
GREAT – step 2 : data exploration

Web application developed with R Shiny

Two R scripts :

ui.R : to create the interface

server.R : to perform calculs and generate tables and graphs



Demo of GREAT

R Shiny application

GREAT is available online. Please, contact amandine.velt@inra.fr to request an account.

Link : http://great.colmar.inra.fr/

Perspectives

Update of GREAT

New samples have been added to the GREAT database : 1,162 samples!

Publication in progress !

Add new features according to users feedback.

Acknowledgements

Lauriane Renault
Camille Rustenholz

Gautier Arista

The beta-testers

MSV team – INRAE Colmar
Philippe Hugueney
Thuy-Thanh Truong

GAV team – INRAE Colmar
Eric Duchene

Thank you for your attention
Questions ?