



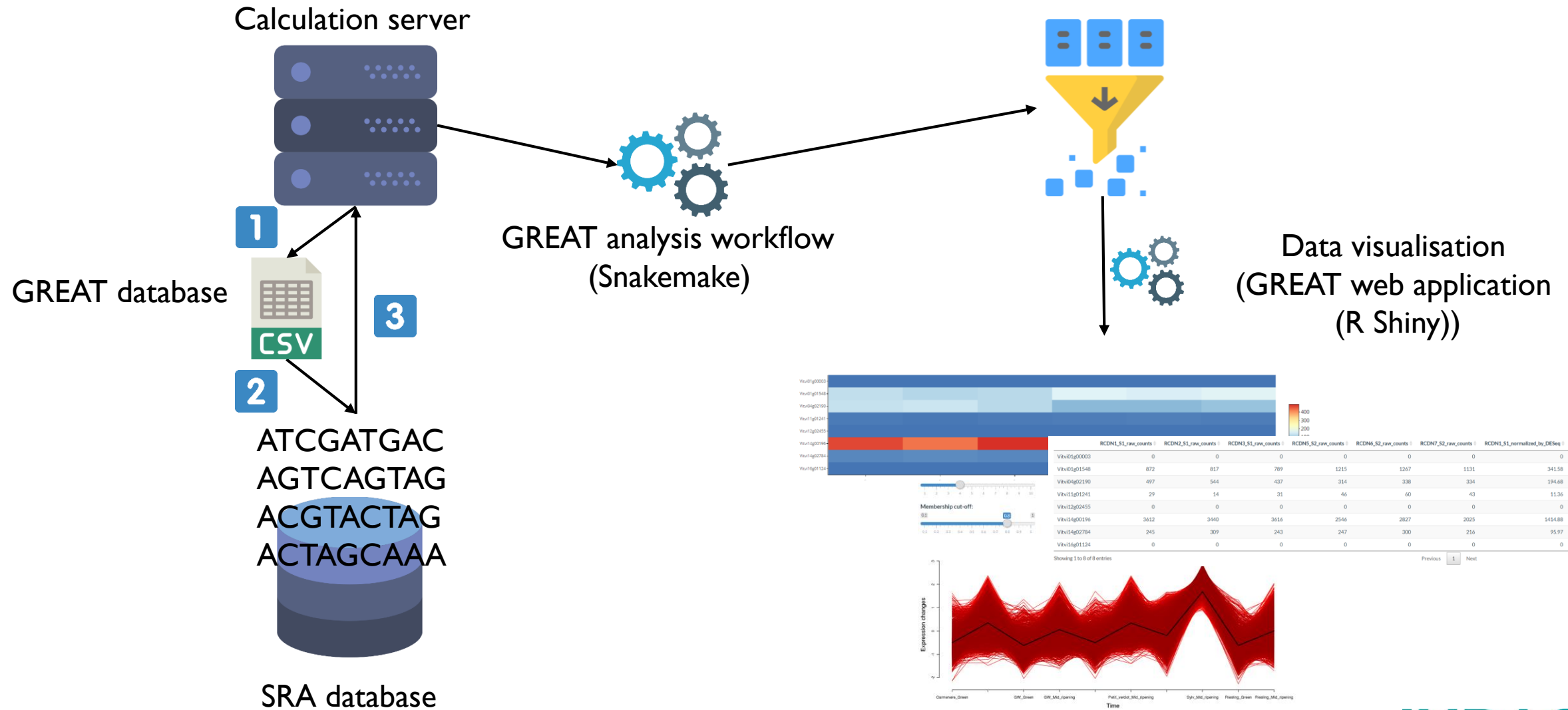
# **GREAT - GRape Expression ATlas**

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

**Amandine Velt, Lauriane Renault, Camille Rustenholz**

# GREAT – Grape Expression ATlas

Operating mode





# GREAT – Grape Expression Atlas

Operating mode

ATCGATGAC  
AGTCAGTAG  
ACGTACTAG  
ACTAGCAAA



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>

NCBI Resources How To Sign in to NCBI

BioProject  Search

Create alert Advanced Browse by Project attributes Help

Project Types  
Primary submission (50)

Data Types  
Genome sequencing (1)  
Other (3)  
Transcriptome (32)

Project Data  
Nucleotide (2)  
SRA (42)  
GEO DataSets (20)

Scope  
Monoisolate (12)  
Multi-isolate (23)  
Multi-species (14)  
Environmental (1)

Organism Groups  
Bacteria (1)  
Fungi (2)  
Plants (48)

[Clear all](#)  
[Show additional filters](#)

Display Settings: Summary, 20 per page, Sorted by Default order Send to: Filters: [Manage Filters](#)

See also 6 genomes matching your organism search

**Search results**  
Items: 1 to 20 of 50

<< First < Prev Page 1 of 3 Next > Last >>

☐ [Leafroll](#)  
1. Project data type: Raw sequence reads  
Scope: Multispecies  
HES-SO University of Applied Sciences and Arts Western Switzerland  
Accession: PRJNA594635 ID: 594635

☐ [Grapevine comparative early transcriptome profiling reveals that Flavescence dorée phytoplasma represses plant responses induced by vector feeding in susceptible varieties](#)  
2. Project data type: Other  
Scope: Monoisolate  
CREA-GB Centro di ricerca Genomica e Bioinformatica  
Accession: PRJEB31325 ID: 549208

☐ [Rpv3-1 Locus Mediated Resistance to Grapevine Downy Mildew is Associated with Specific Host](#)  
3. [Transcriptional Responses and the Accumulation of Stilbenes](#)  
Organism: **Vitis vinifera**  
Taxonomy: *Vitis vinifera* (wine grape)  
Project data type: Transcriptome or Gene expression  
Scope: Monoisolate  
University of Tuebingen, Quantitative Biology Center (QBiC)  
**Accession: PRJNA529179 ID: 529179**

☐ [wine grape leaf transcriptome](#)  
4. Project data type: Raw sequence reads  
Scope: Multispecies  
Yunnan Agricultural university  
Accession: PRJNA528770 ID: 528770

**Find related data**  
Database: Select  
Find items

**Search details**  
("Vitis vinifera"[Organism] OR vitis vinifera[All Fields]) AND rna-seq[All Fields]  
Search See more...

**Recent activity**  
[Turn Off](#) [Clear](#)

Q vitis vinifera rna-seq (50) BioProject

Rpv3-1 Locus Mediated Resistance to Grapevine Downy Mildew is Assoc GEO DataSets

Q GEO DataSets for BioProject (Select 529179) (1) GEO DataSets

Mueller normal isolate infected sample at 6h pi\_rep1 biosample

Q SRA Links for BioProject (Select 529179) (10) SRA

See more...

# SRA database : run selector

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



SRA Run Selector



Log in to NIH

## Filters List

- ☐ Bases
- ☐ BioSample
- ☐ Bytes
- ☐ Cultivar
- ☐ Experiment
- ☐ GEO\_Accession
- ☐ Inoculation
- ☐ phenotype
- ☐ Run
- ☐ Sample Name
- ☐ 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, ncbl.public, s3.us-east-1
Instrument	NextSeq 500

## Select

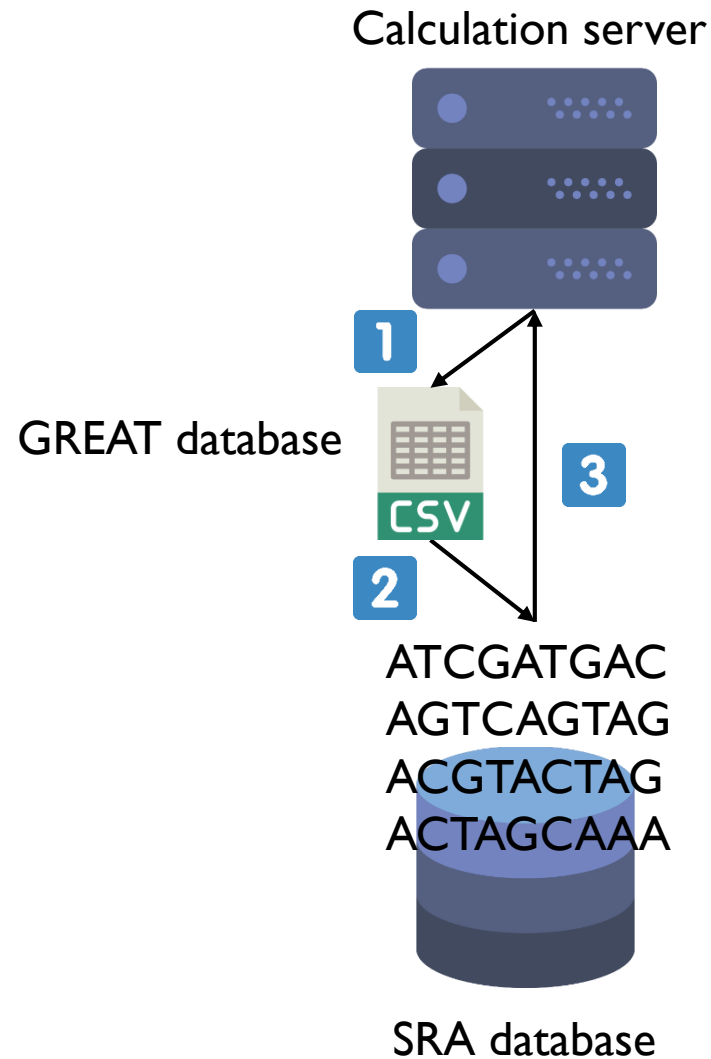
	Runs	Bytes	Bases	Download
Total	10	6.26 Gb	14.54 G	Metadata or Accession List
	0	0	0	Metadata or Accession List or JWT Cart

Found 10 Items

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

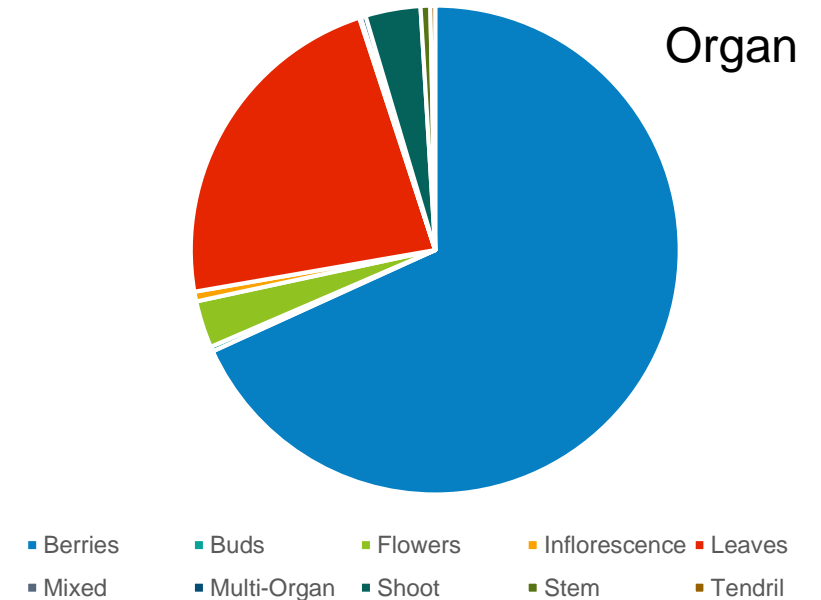
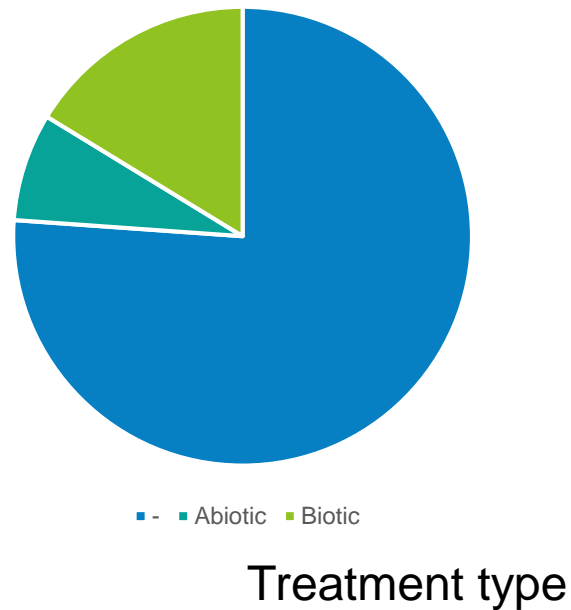
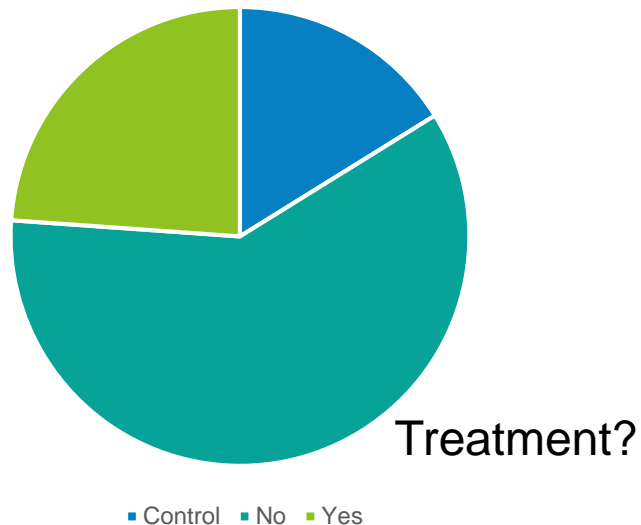
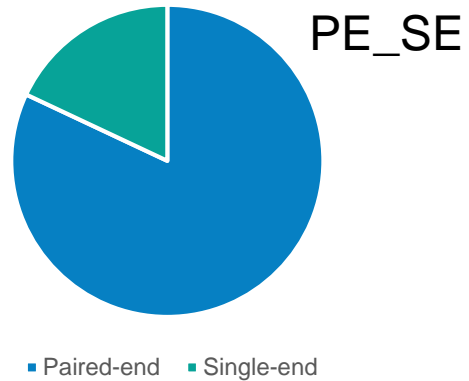
Bioproject	Protocol library	Treatment?
First author	Stranded	Treatment type
Paper	Species	Treatment nature
File name	Variety	Treatment detail
Sequencing platform	Organ	Comments
PE_SE	Tissue	Quality check
Read length bp	Stage_detail	



# GREAT's database

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

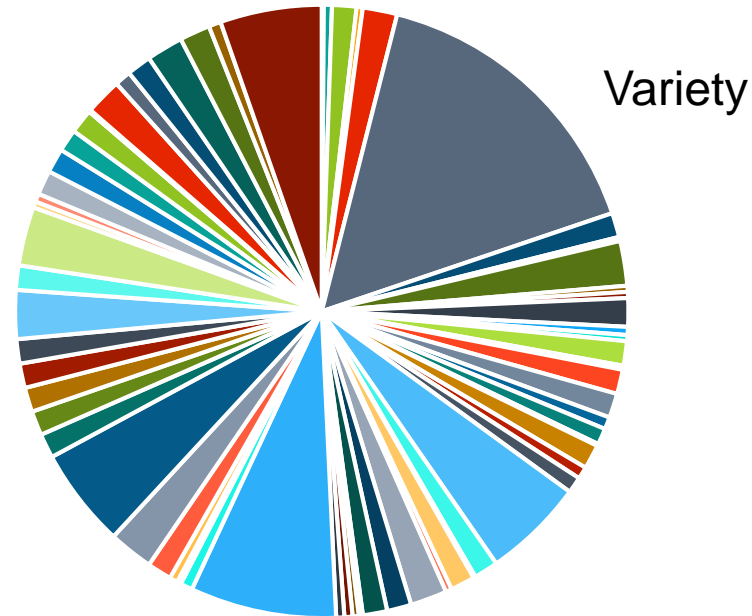
915 samples : identified, annotated and analysed



# GREAT's database

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

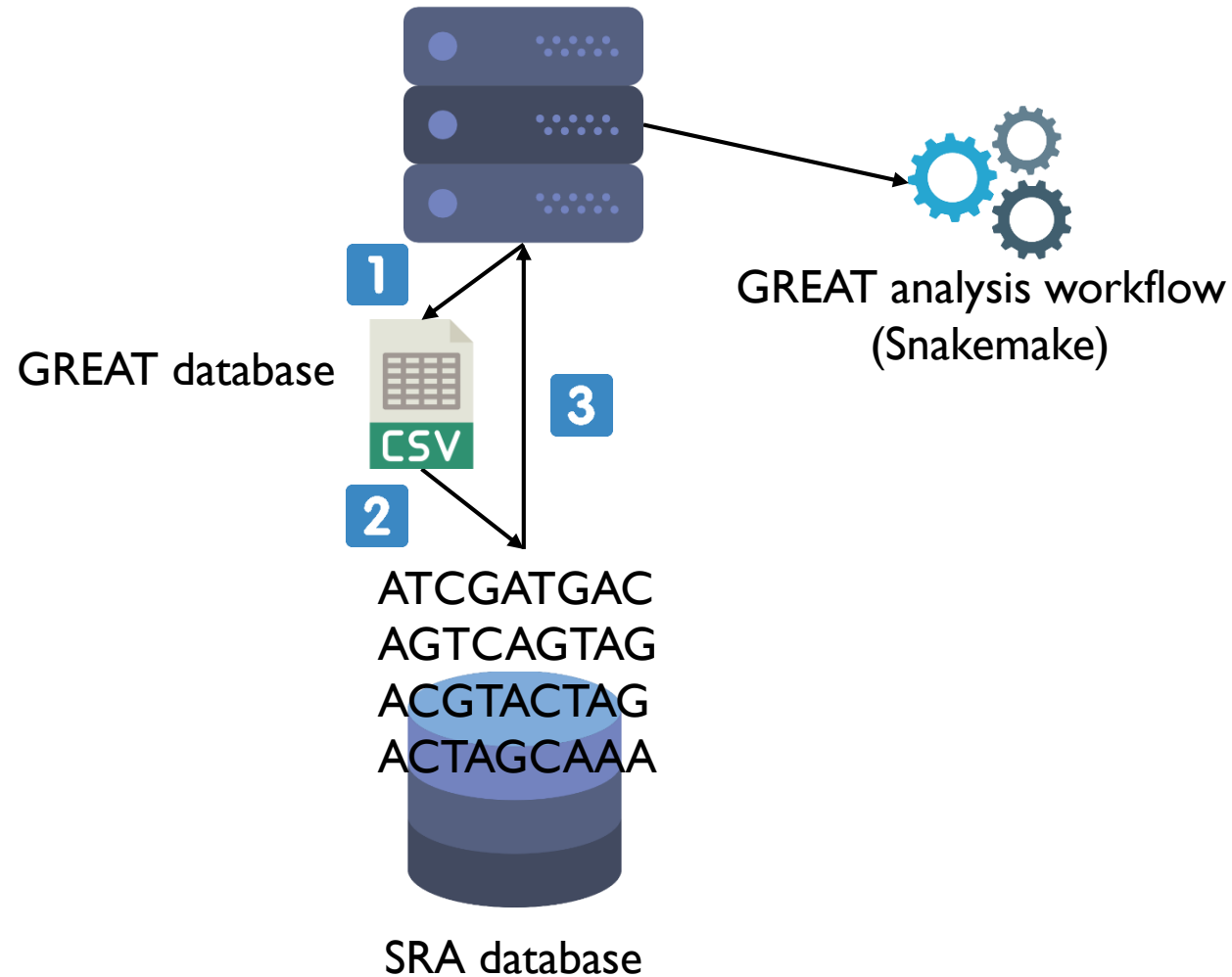
915 samples : identified, annotated and analysed



- |                         |                   |                                  |                     |                          |                              |                              |
|-------------------------|-------------------|----------------------------------|---------------------|--------------------------|------------------------------|------------------------------|
| ■ 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](https://gitlab.com/avelt/great_rnaseq_workflow)

## Read alignment with STAR + quality filter

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

## Conversion to fastq + quality control (fastQC)

This step allows to validate or not the samples

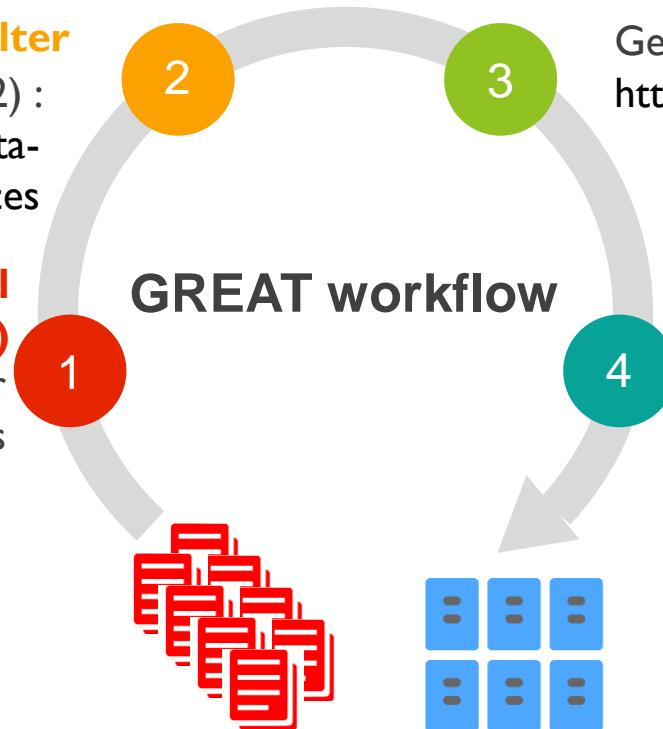
## GREAT workflow

## 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

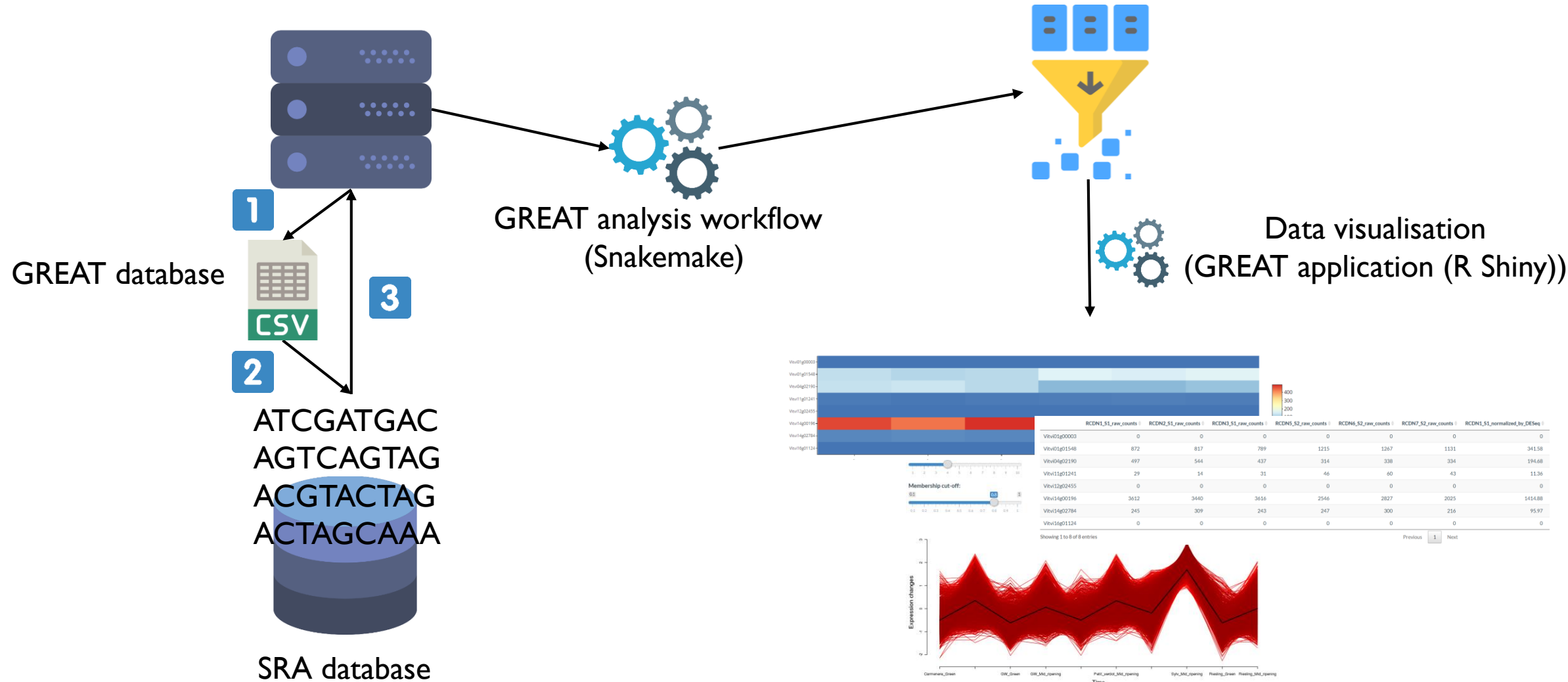
- Normalization of count data & samples merging
- Calculation of correlation coefficients
- Variance-stabilizing transformation (for PCA)





# GREAT – Grape Expression ATlas

Operating mode



# GREAT – step 2 : data exploration

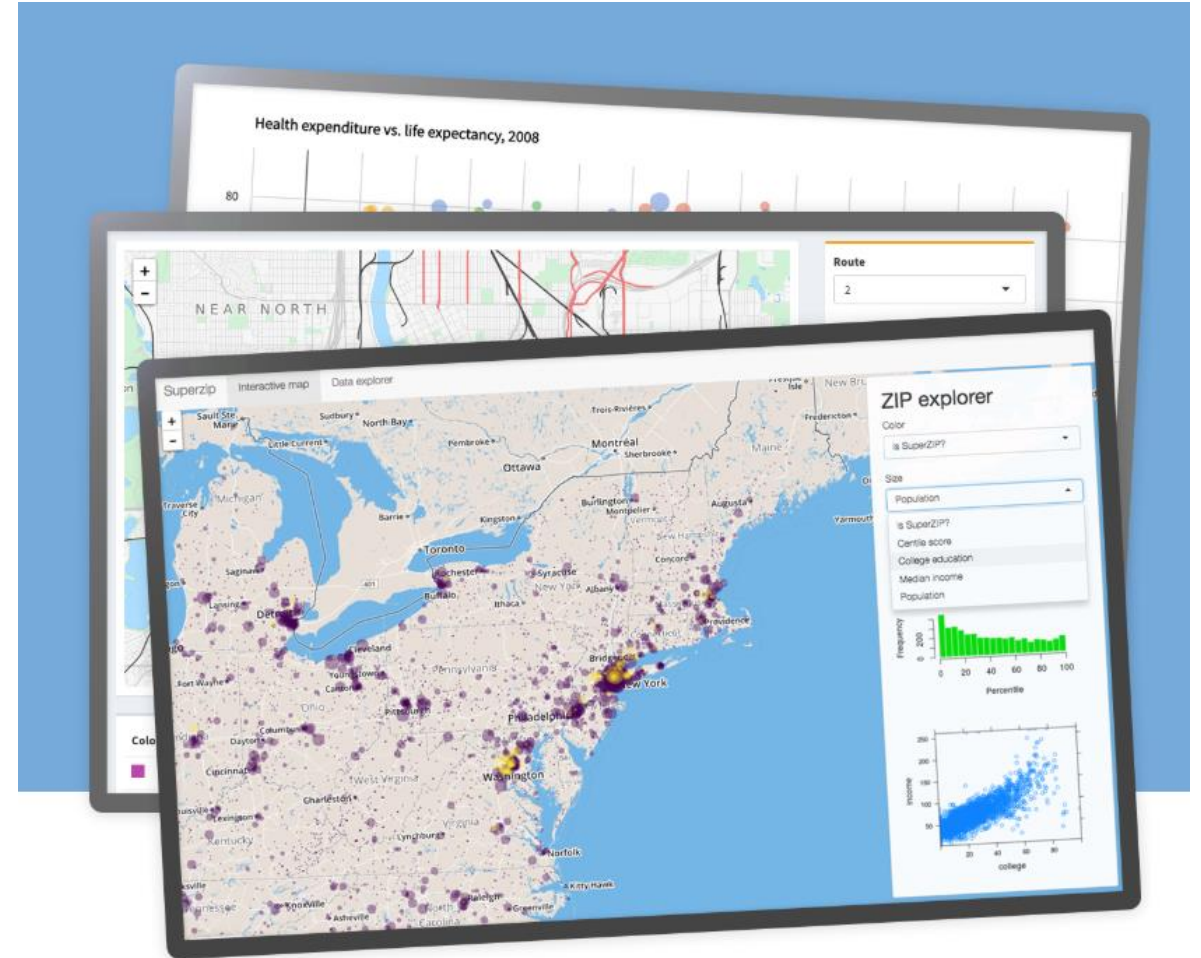
Web application developed with R Shiny



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](mailto: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 ?**

