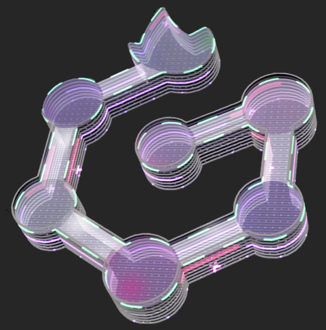
Integrating genetics, omics, and literature-based data for grapevine improvement:



i2sysbio

TOMSBio lab

José Tomás Matus

Transcriptional Orchestration of Plant Secondary Metabolism (TOMS lab)
Institute for Integrative Systems Biology (I2SysBio, UV-CSIC), Spain

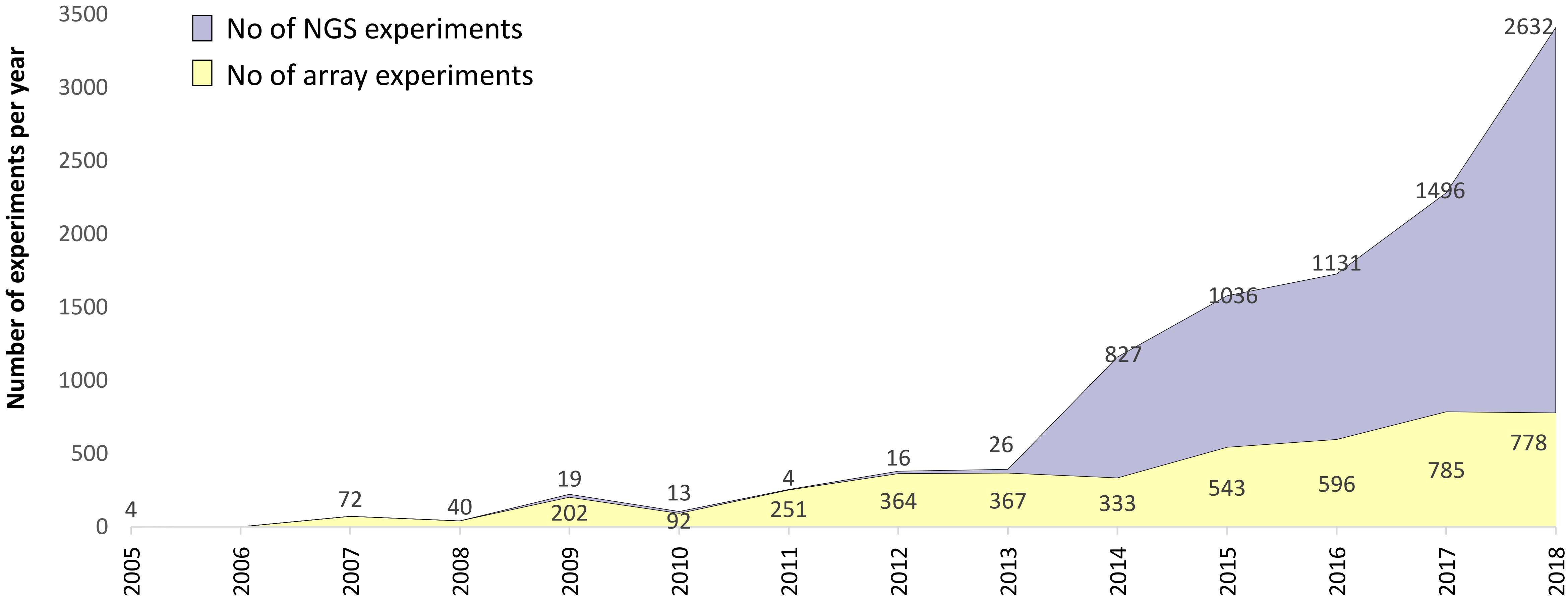
Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

Justin Lashbrooke

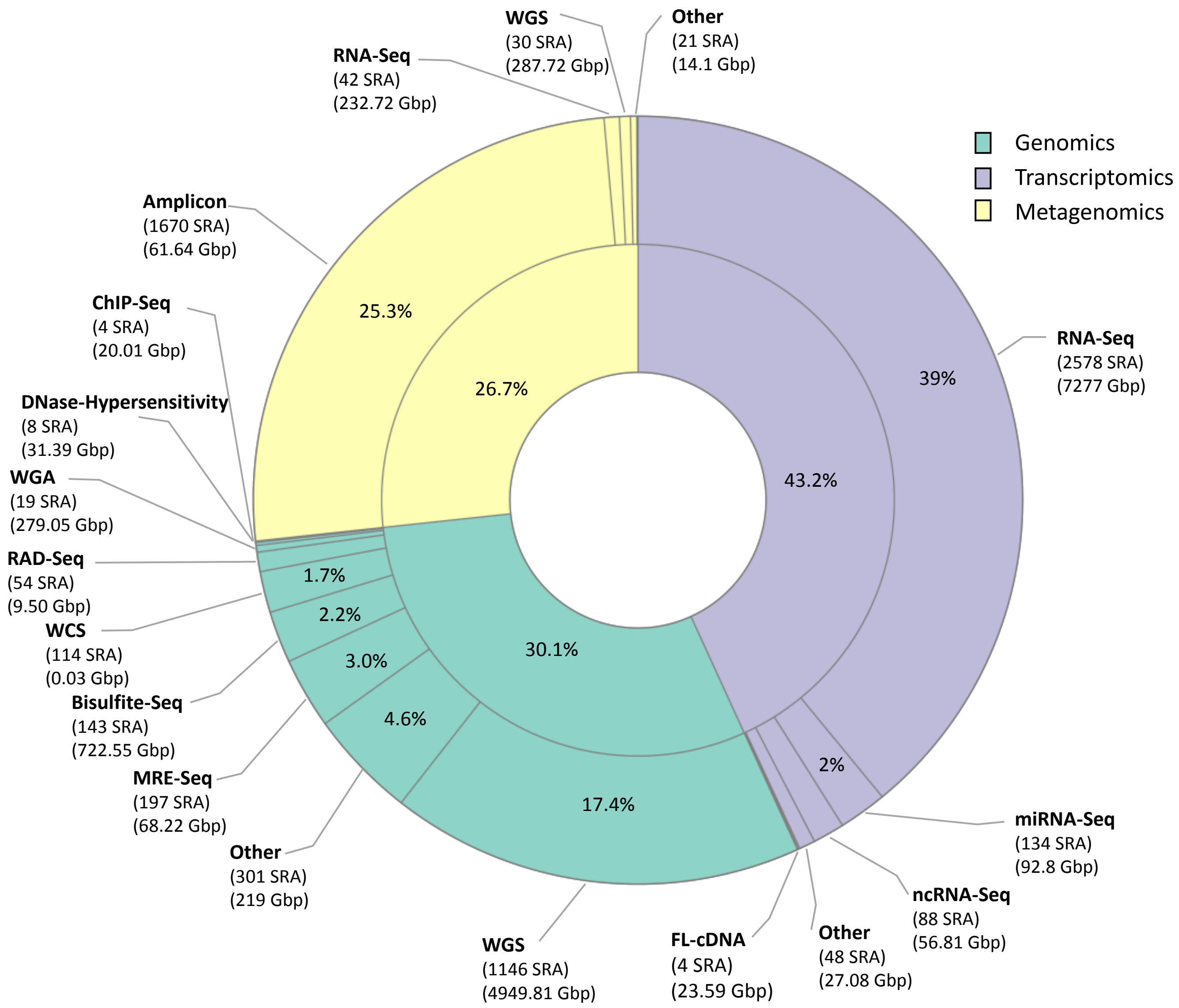
Department of Genetics
Stellenbosch University, South Africa.

TABLE GRAPE CONFERENCE, CAPE TOWN. 2023

We are in the Era of Big Data & Databases



Matus et al. Prospects of Systems Biology in Grapevine. In: The Grape Genome. Cantu, Walker, Eds.



Matus et al. Prospects of Systems Biology in Grapevine. In: The Grape Genome. Cantu, Walker, Eds.

Hosted by Cantu Lab @ UC Davis

Current tools available for the grape community

Genome browsers

grapegenomics.com
Hosted by Cantu Lab @ UC Davis

Gramene

Others
(mostly private)

Transcriptome explorers

vespucci

葡萄大数据应用平台

GREAT

EP

VITVIZ

Cultivar collections

Pl@ntGrape

VIVC
Vitis International Variety Catalogue *VIVC*

Metabolite repository

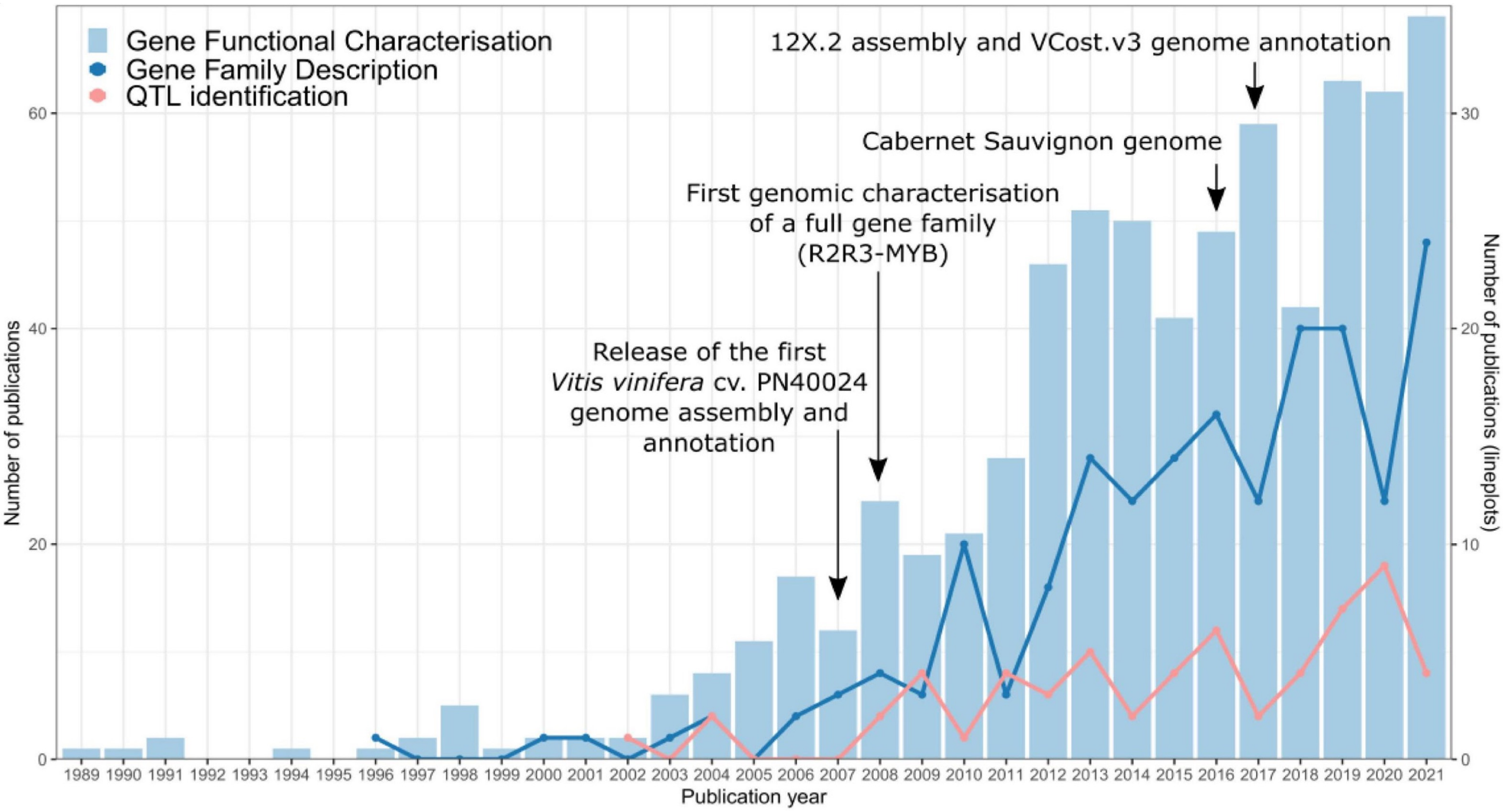
MetaboLights

Literature data?

Gene Networks

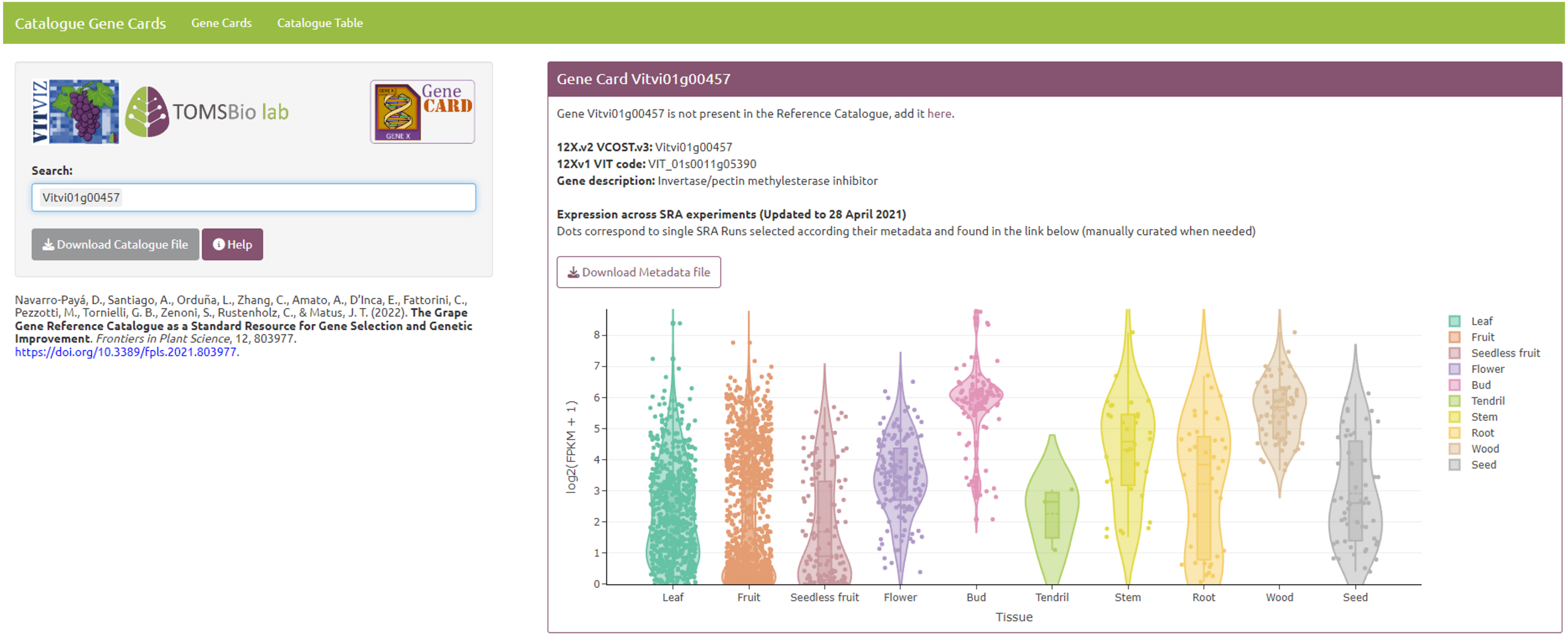
Vitis OneGenE
Single Gene Network Expansion

VITVIZ



Navarro-Payá et al. 2022. The Grape Gene Reference Catalogue as a resource for gene selection and Genetic Improvement

Developed tools: Gene Cards



Search: Vitvi01g00457

Navarro-Payá, D., Santiago, A., Orduña, L., Zhang, C., Amato, A., D'Inca, E., Fattorini, C., Pezzotti, M., Tornielli, G. B., Zenoni, S., Rustenholz, C., & Matus, J. T. (2022). **The Grape Gene Reference Catalogue as a Standard Resource for Gene Selection and Genetic Improvement**. *Frontiers in Plant Science*, 12, 803977. https://doi.org/10.3389/fpls.2021.803977.

Gene Card Vitvi01g00457

Gene Vitvi01g00457 is not present in the Reference Catalogue, add it here.

12X.v2 VCOST.v3: Vitvi01g00457
12Xv1 VIT code: VIT_01s0011g05390
Gene description: Invertase/pectin methylesterase inhibitor

Expression across SRA experiments (Updated to 28 April 2021)
Dots correspond to single SRA Runs selected according their metadata and found in the link below (manually curated when needed)

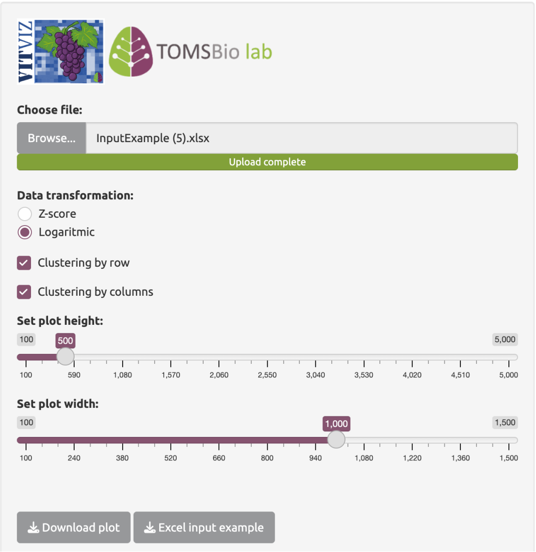
Developed tools: PlantViz platform

http://vitviz.tomsbiolab.com/

Web-based platform with several tools to visualize gene expression data in the form of expression heatmaps of a desired list of genes, and gene- or list-centered coexpression networks. A JBrowser with DAP-seq data published in our lab is also offered. The tool also includes a search engine for looking correspondencies between all PN40024 genome annotations and the Grape Reference Catalogue.

Developed tools: PlantViz platform

EX-ATLAS



Choose file:

Browse... InputExample (5).xlsx

Upload complete

Data transformation:

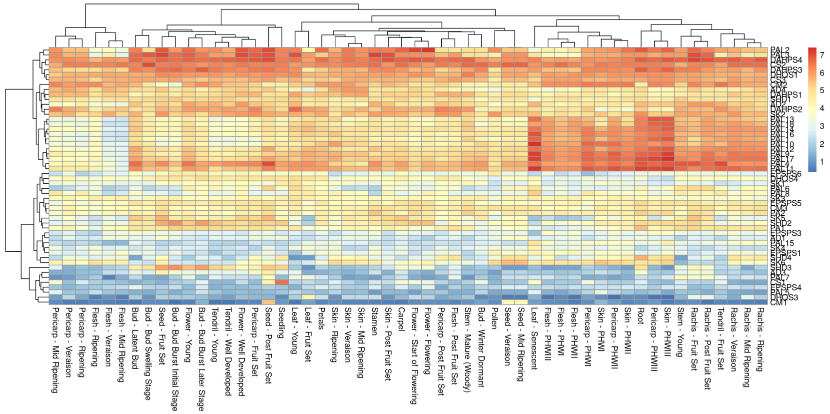
○ Z-score
● Logaritmic

☑ Clustering by row

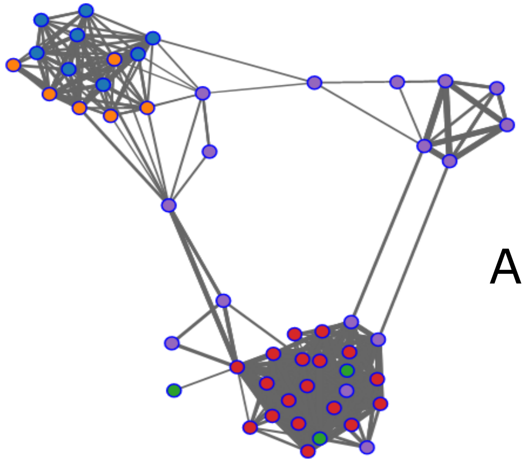
☑ Clustering by columns

Set plot height: 500 (100 – 5,000)

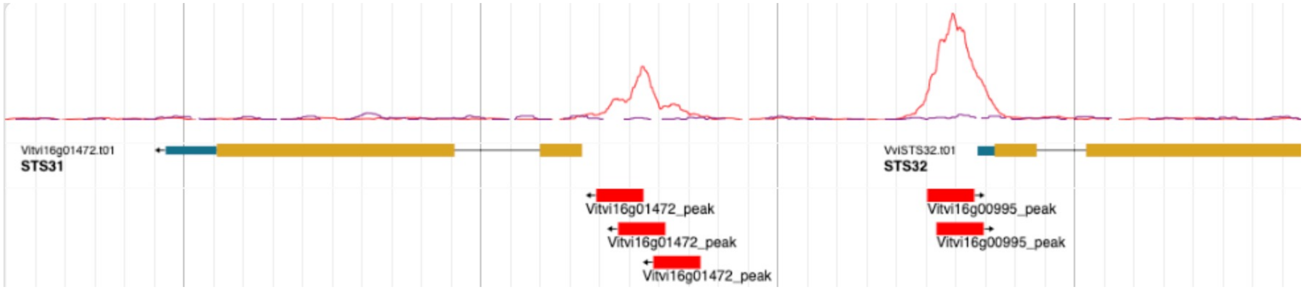
Set plot width: 1,000 (100 – 1,500)



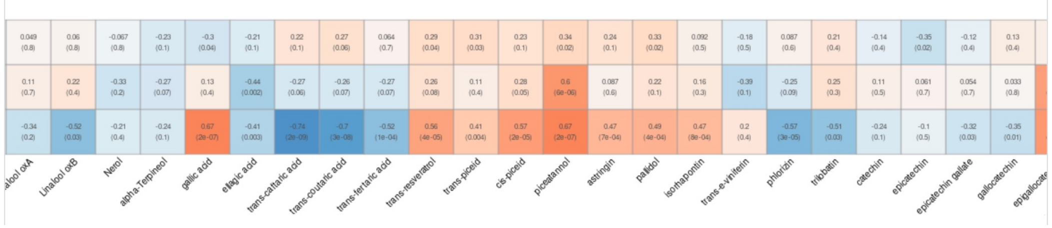
AggGCNs



DAPBrowse



TransMETA



alool oxA	Linalool oxB	Nerol	alpha-Terpineol	gallic acid	ellagic acid	trans-caftaric acid	trans-coutaric acid	trans-fertaric acid	trans-resveratrol	trans-piceid	cis-piceid
0.049 (0.8)	0.06 (0.8)	-0.067 (0.8)	-0.23 (0.1)	-0.3 (0.04)	-0.21 (0.1)	0.22 (0.1)	0.27 (0.06)	0.064 (0.7)	0.29 (0.04)	0.31 (0.03)	0.23 (0.1)
0.11 (0.7)	0.22 (0.4)	-0.33 (0.2)	-0.27 (0.07)	0.13 (0.4)	-0.44 (0.002)	-0.27 (0.06)	-0.26 (0.07)	-0.27 (0.07)	0.26 (0.08)	0.11 (0.4)	0.28 (0.05)
-0.34 (0.2)	-0.52 (0.03)	-0.21 (0.4)	-0.24 (0.1)	0.67 (2e-07)	-0.41 (0.003)	-0.74 (2e-09)	-0.7 (3e-08)	-0.52 (1e-04)	0.56 (4e-05)	0.41 (0.004)	0.57 (2e-05)

piceatannol	astringin	pallidol	isorhapontin	trans-e-viniferin	phlorizin	trilobatin	catechin	epicatechin	epicatechin gallate	gallocatechin
0.34 (0.02)	0.24 (0.1)	0.33 (0.02)	0.092 (0.5)	-0.18 (0.5)	0.087 (0.6)	0.21 (0.4)	-0.14 (0.4)	-0.35 (0.02)	-0.12 (0.4)	0.13 (0.4)
0.6 (6e-06)	0.087 (0.6)	0.22 (0.1)	0.16 (0.3)	-0.39 (0.1)	-0.25 (0.09)	0.25 (0.3)	0.11 (0.5)	0.061 (0.7)	0.054 (0.7)	0.033 (0.8)
0.67 (2e-07)	0.47 (7e-04)	0.49 (4e-04)	0.47 (8e-04)	0.2 (0.4)	-0.57 (3e-05)	-0.51 (0.03)	-0.24 (0.1)	-0.1 (0.5)	-0.32 (0.03)	-0.35 (0.01)

visit http://tomsbiolab.com

Funding sources

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
AGENCIA ESTATAL DE INVESTIGACIÓN
FEDER
Fondo Europeo de Desarrollo Regional
GENERALITAT VALENCIANA
HFSP
International Human Frontier Science Program Organization
cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY
INSTITVT D'ESTVDIS CATALANS MCMVII
Institut d'Estudis Catalans

Current tools available for the grape community

Genome browsers

grapegenomics.com
Hosted by Cantu Lab @ UC Davis

Gramene

Others (mostly private)

Transcriptome explorers

vespucci

葡萄大数据应用平台

GREAT

eFP

VITVIZ

Cultivar collections

Pl@ntGrape

VIVC

Vitis International Variety Catalogue *VIVC*

Metabolite repository

MetaboLights

Literature data

VITVIZ

Gene Networks

Vitis OneGenE
Single Gene Network Expansion

VITVIZ

Available tools

Not yet F.A.I.R.

Not yet available

Phenotypes

QTL

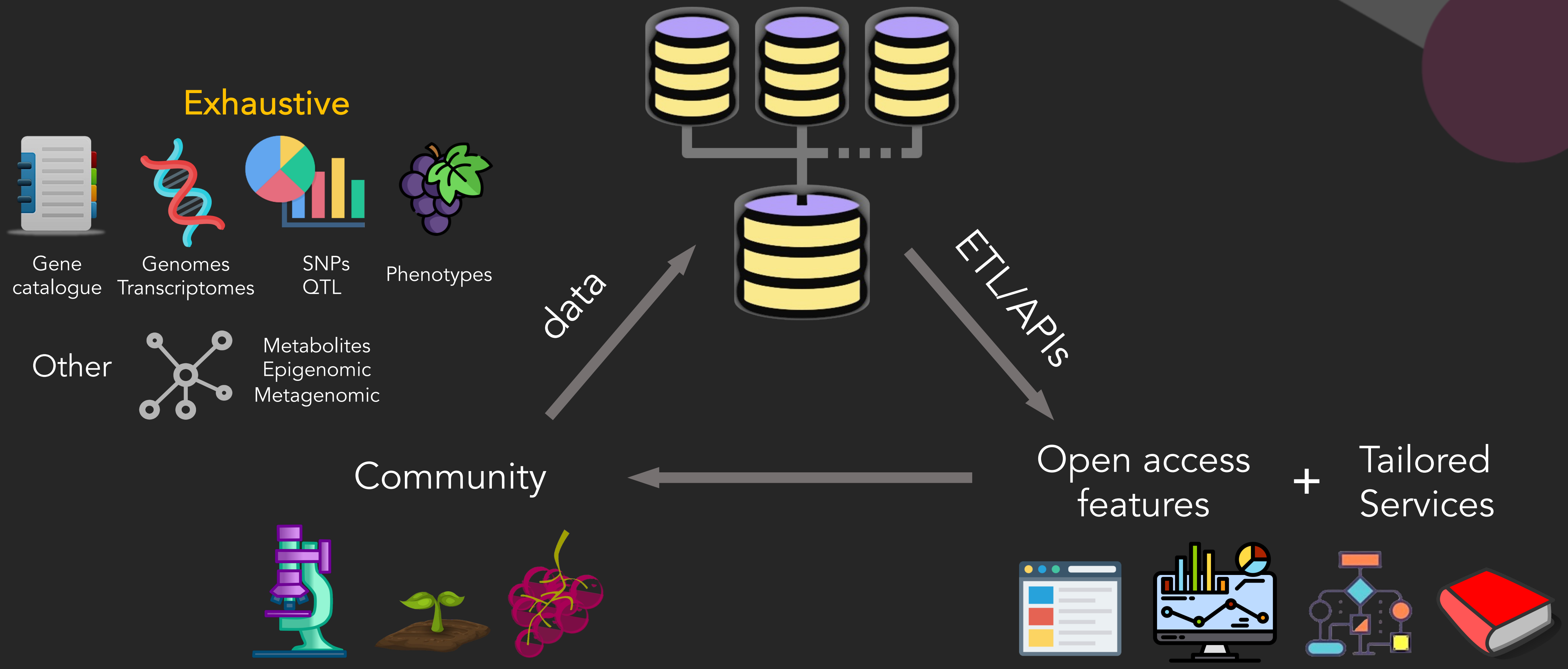
Climate data

Grapedia

The Grapevine Genomics Encyclopedia

an innovative portal that integrates biological knowledge, genetic and genomic resources and customized services for the grape scientific community and industry

visit http://grapedia.org



Federative Database System

Modular (FDBS)

data

ETL/APIs

Exhaustive

- Gene catalogue
- Genomes Transcriptomes
- SNPs QTL
- Phenotypes

Other

- Metabolites
- Epigenomic
- Metagenomic

Community

- researchers
- breeders
- wineries & producers

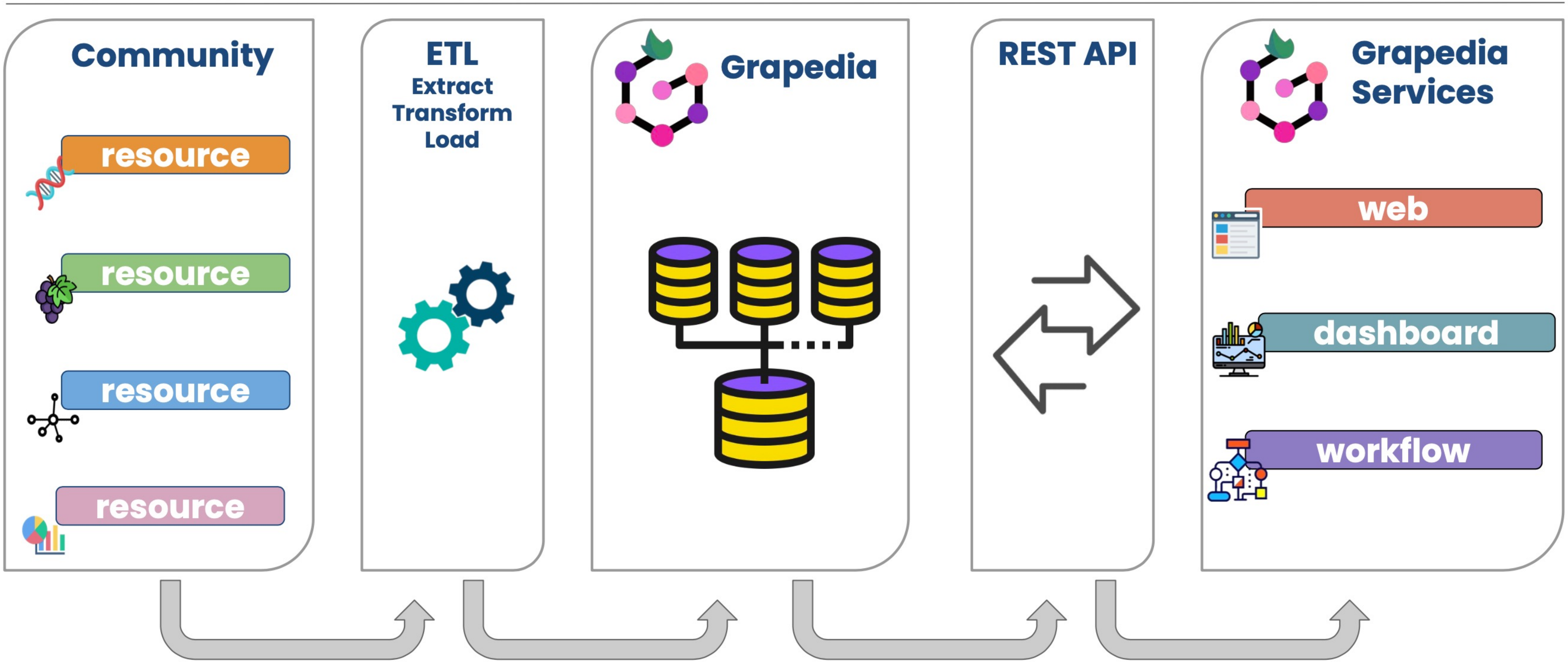
Diverse

Open access features + Tailored Services

- browsers
- dashboard
- Analyses
- Consulting

Interactive / Graphic

BACKEND FRONTEND



Community

resource

resource

resource

resource

ETL

Extract
Transform
Load

Grapedia

REST API

Grapedia Services

web

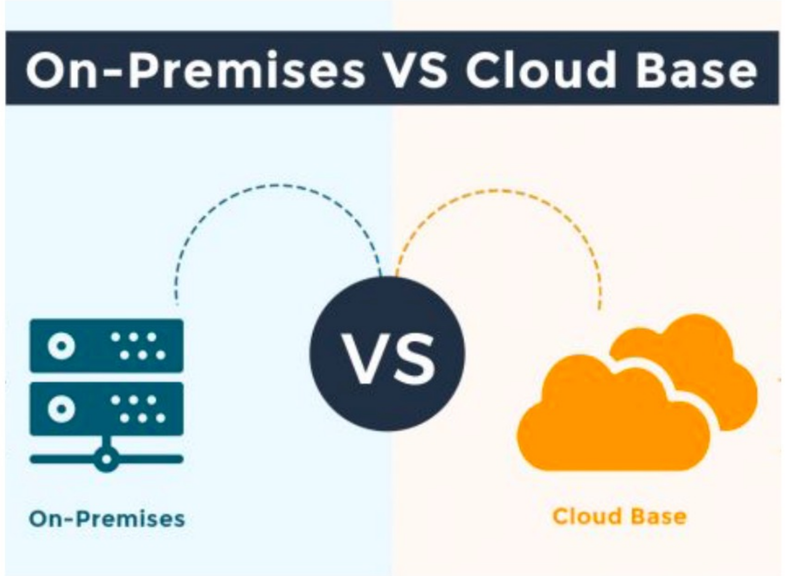
dashboard

workflow

Infrastructure

Scalability
Flexibility
Reproducibility

SEQUENTIA
DECODE • DISCOVER • DISRUPT



Grapedia
The Grapevine Genomics Encyclopedia →

On-Premises

Pros

- » Full Data Control
- » Full Hard- and Software Control
- » Full Access Reliability
- » No Operating Software Cost
- » Performance
- » Customizations

Cons

- » Maintenance and Acquisition Costs
- » Internal Knowledge
- » Full Responsibility
- » Long-term Reliability

Cloud-Base

Pros

- » Easy Set-up
- » Low Acquisition and Maintenance Costs
- » Flexibility and Scalability
- » Accessibility and Integration
- » Updates and Security
- » Back-up and Data Restore
- » Disaster Recovery

Cons

- » On-going Software Costs
- » Performance Limitations
- » Internet Access

Menu

Gene(s) name

Vitvi02g01019

Genome

Vitis vinifera
PN40024 v4
Cabernet Sauv v1
Thompson seedless
…

Vitis riparia
Vitis berlandieri
…

Single Gene Dashboard

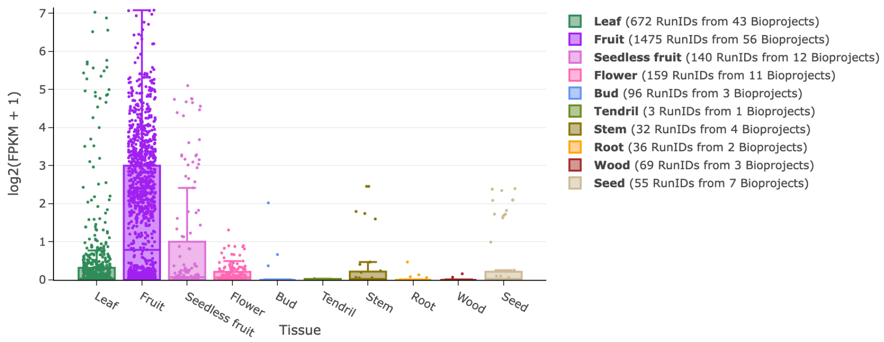
This dashboard is meant to explore a single gene in different contexts. From genome browser to its expression level across different tissues, as well as gene co-expression networks

Catalogue & Gene Cards Explorator

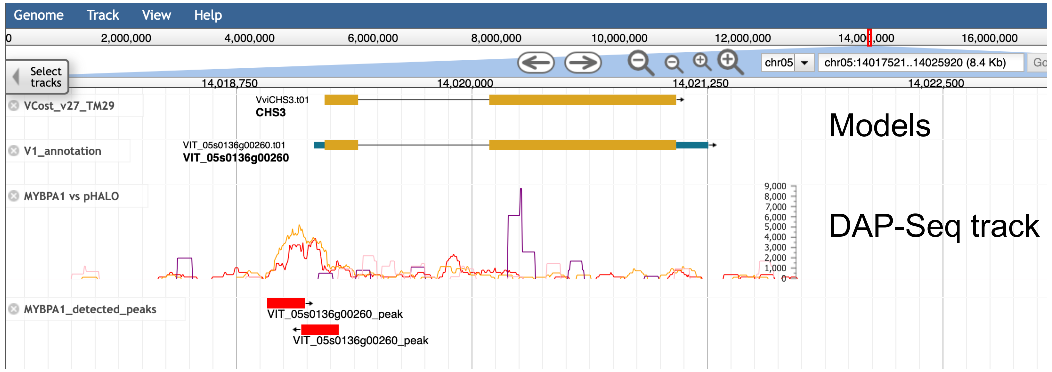
type: TF
Validation level: 6
Functional association: Regulator of anthocyanin synthesis. Berry color locus.
Paper references for functional association: Kobayashi et al., 2004
DOI of functional association: NULL
Paper references for family description: Wong et al. 2016 (DNA Res)
DOI of family description: doi.org/10.1093/dnares/dsw028
Phylogeny constructed: Maximum Likelihood (Bayesian)

Expression across SRA experiments (Updated to 28 April 2021)
Dots correspond to single SRA Runs selected according their metadata and found in the link below (manually curated when needed)

Download Metadata file



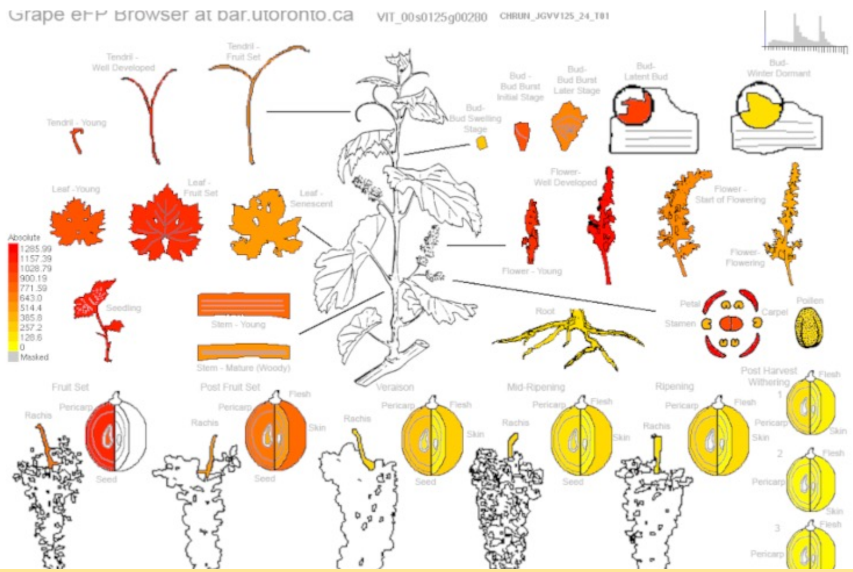
Integrated Genome Browser



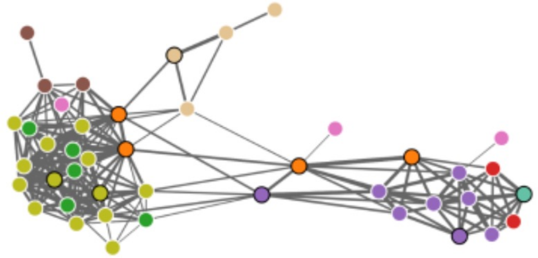
Models

DAP-Seq track

Gene Expression Atlases



Co-expression network



Heatmaps

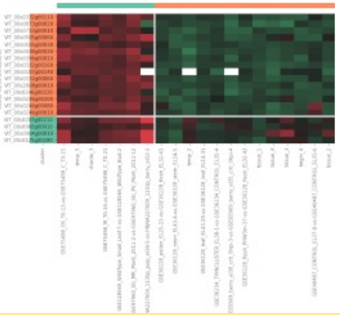
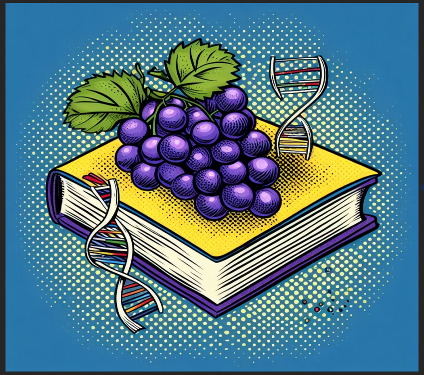


TABLE GRAPE CONFERENCE, CAPE TOWN. 2023

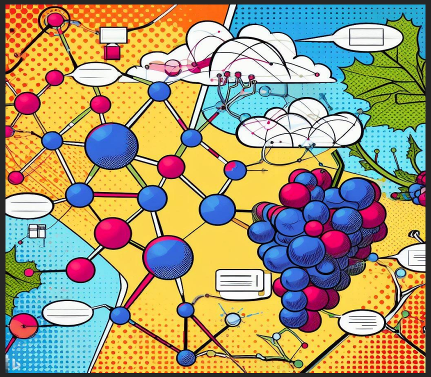
Dashboards - Implemented models within first release (Early 2024)



Genome JBrowser2

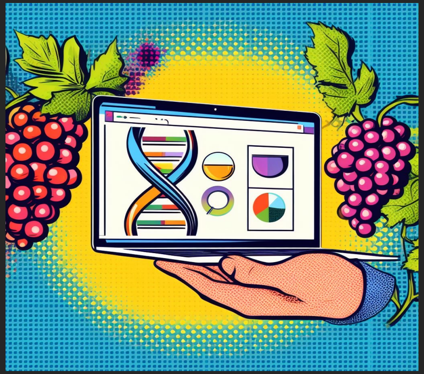
QTL Browser
Gene Catalogue/
Cards

Expression
atlases



Co-expression &
regulatory
networks

Workflows (Nextflow)



Expression profiling,
gene annotation, others

Guidelines

**& Curation /
Submission platforms**

Funding

- CIG Funding
- Donations & Sponsoring
- Own-funding resources
- Public grant applications

Establish DB

Commercial exploitation

DB long-term maintenance and growth

Services
- Externalized
- Tailored

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

E&J. Gallo Winery

novatech
SÉLECTION - ANALYSES - RECHERCHE

SEQUENTIA
DECODE • DISCOVER • DISRUPT

GRAPEDIA Annual Meeting, Valencia, Spain (2023)

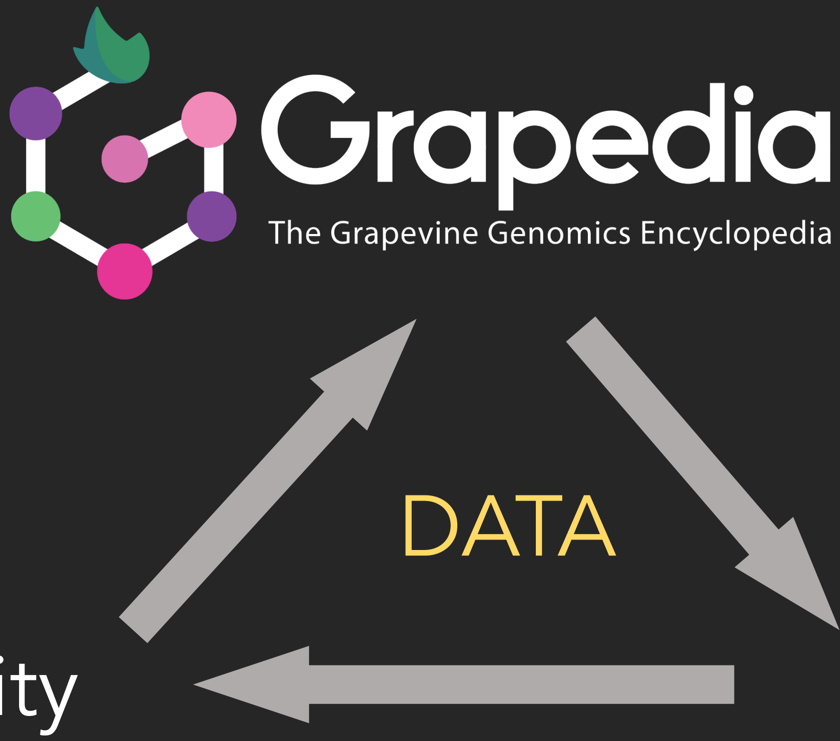
Grapedia

The Grapevine Genomics Encyclopedia

DATA

Community

Solutions



Grapedia

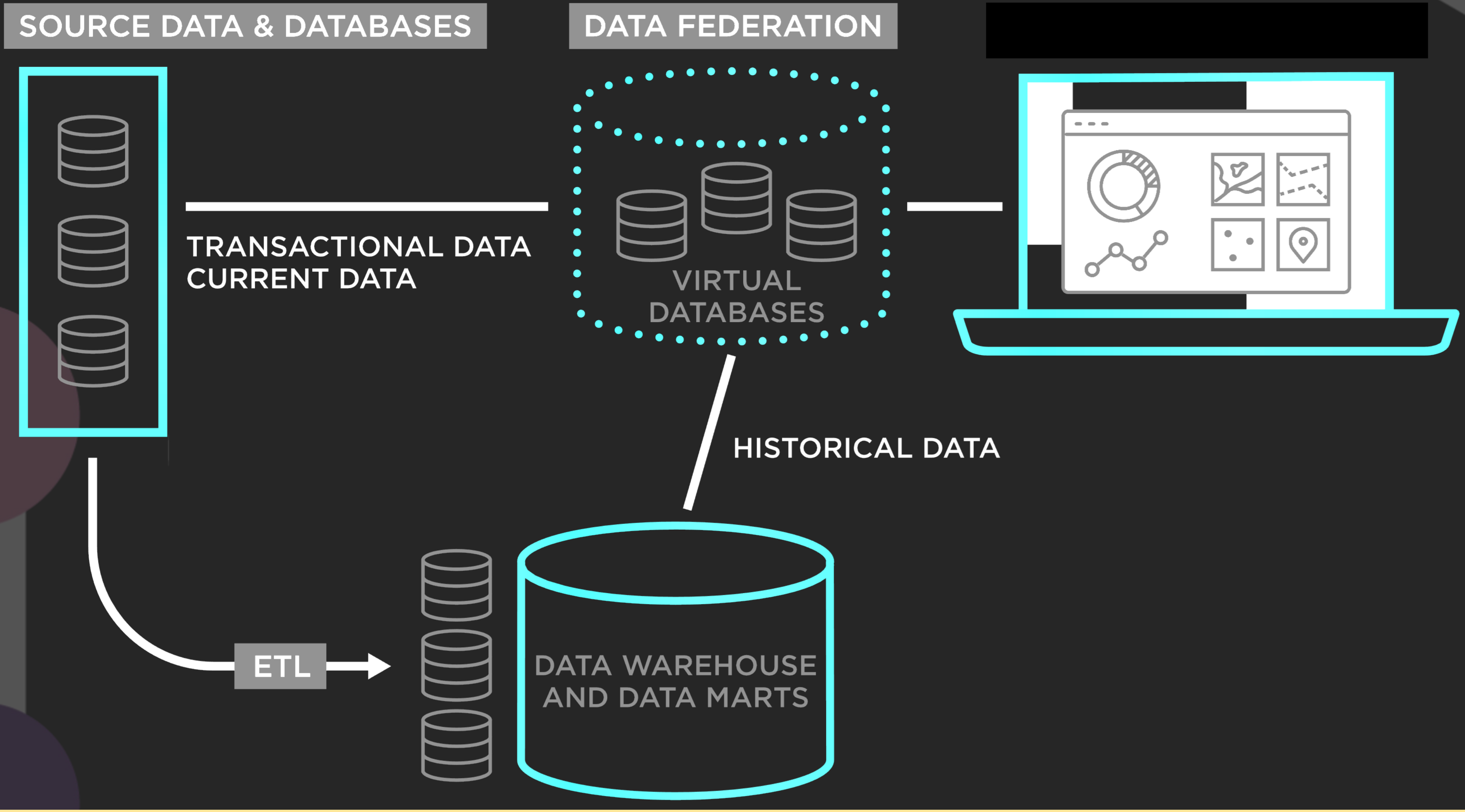
The Grapevine Genomics Encyclopedia

DATA

Community

Solutions

Federative Database System (FDBS)



SOURCE DATA & DATABASES

DATA FEDERATION

[redacted]

TRANSACTIONAL DATA
CURRENT DATA

VIRTUAL
DATABASES

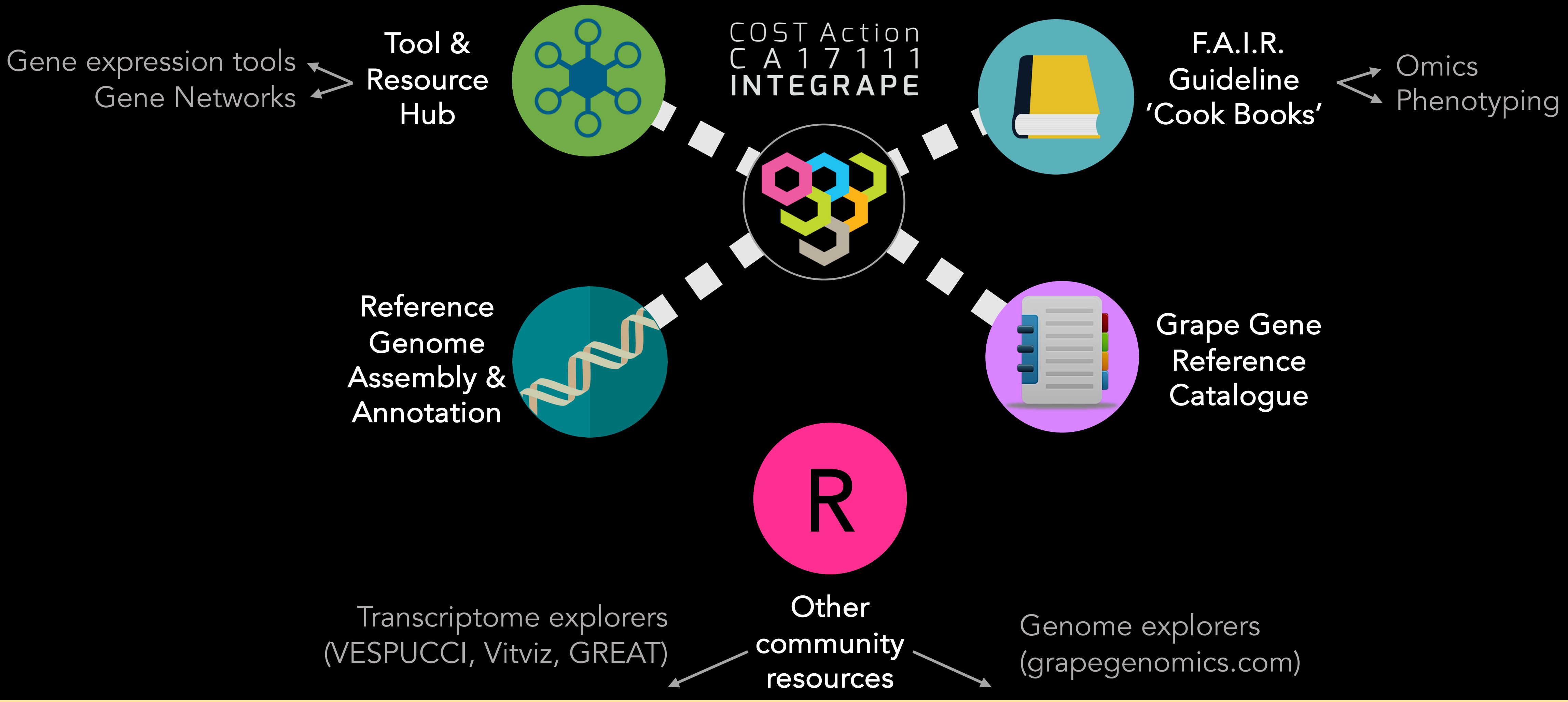
HISTORICAL DATA

ETL

DATA WAREHOUSE
AND DATA MARTS

Grapedia
The Grapevine Genomics Encyclopedia

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY



COST Action
CA17111
INTEGRAPE

Gene expression tools
Gene Networks

Tool & Resource Hub

F.A.I.R. Guideline 'Cook Books'

Omics
Phenotyping

Reference Genome Assembly & Annotation

Grape Gene Reference Catalogue

Other community resources

Transcriptome explorers (VESPUCCI, Vitviz, GREAT)

Genome explorers (grapegenomics.com)

Grapedia
The Grapevine Genomics Encyclopedia

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

Community

COST Action
CA17111
INTEGRAPE

Tool & Resource Hub

Gene expression tools
Gene Networks

F.A.I.R. Guideline 'Cook Books'

Omics
Phenotyping

Reference Genome Assembly & Annotation

Grape Gene Reference Catalogue

Other community resources

Transcriptome explorers (VESPUCCI, Vitviz, GREAT)

Genome explorers (grapegenomics.com)

Commercial aspects

Services

DB long-term management

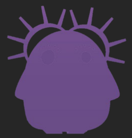
SEQUENTIA
DECODE • DISCOVER • DISRUPT

Externalized

Tailored

1. Link grapevine resources (API develop.)
2. Integrate Open Source tools
3. Develop Cloud Software
4. Customize algorithms
5. Ensure Security & Maintenance

AIR, Artificial intelligence RNAseq



GAIA, Metagenomics data analysis

GROWMICS, RNA analysis for agricultural R&D

GREEN GINO, Genomics data exploitation

Grapedia services

Clear interpretation
Unique perspectives
Applicable results

SEQUENTIA
DECODE · DISCOVER · DISRUPT

web

Advanced UI

Responsive

Accesibility

dashboard

- Data integration & visualization
  - Free custom
  - Open source (eg Jbrowse)
- Tools
  - Free custom (e.g Drago)
  - Open source (e.g BLAST)
- Upload, download, sharing sections
- User profile and data management

workflow

Free custom pipelines to run bioinformatics analysis

nextflow

docker

Community

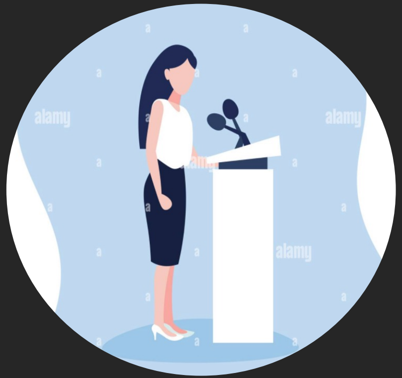
To build GRAPEDIA by building a strong unified community

CIG tools offer

Working
Group
Meetings

Training
Schools
Hackathons

Short-term
Scientific
Missions



Dissemination
activities

Business
plan

Grapedia
The Grapevine Genomics Encyclopedia

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

COST Action
Integrape
CA17111


Mario Pezzotti
University of Verona
(Italy)


Anne-Francoise Adam-Blondon
INRAE Versailles
(France)

FUNDED BY

FUNDED BY THE
EUROPEAN UNION

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

COST
Innovators Grant
IG17111

Grapedia Technology

Innovation at multiple levels

Grapedia Services

IT
infrastructure
Cloud and hybrid solution

Scalability
Flexibility
Reproducibility

Bioinformatics
algorithms
pipelines
database

Modularity
Precise
Versatile

Products
Custom & cloud SOFTWARE
Data delivery platform
Automated interpretation

Clear interpretation
Unique perspectives
Applicable results