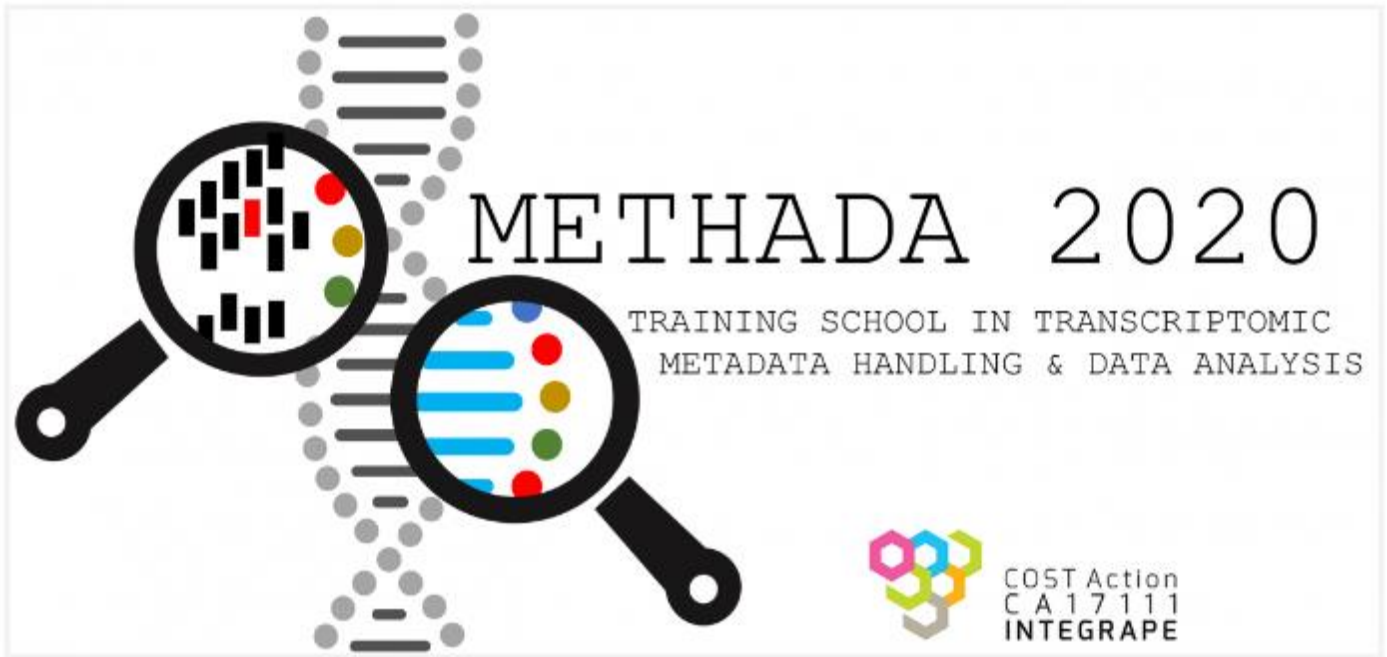


METHADA 2020



Training school on Transcriptomic Metadata Handling and Data Analysis



APPLICATION DEADLINE  
29th November 2019

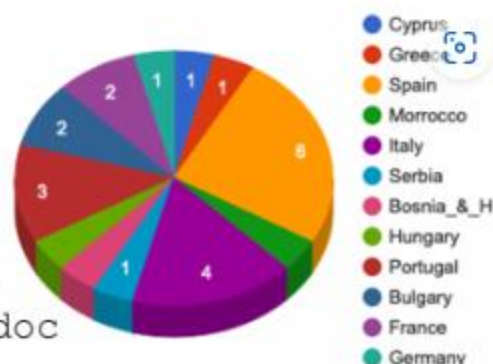
Now Closed

See Presentations and Report Below



-24 trainees  
 -20 grants awarded  
 -gender balanced

10 PhD  
 9 Postdoc  
 5 PI



6 Inclusiveness Target Countries included

“The major benefit was to discuss how to analyse grape expression data”

trainee



“There are still many differences in metadata of public datasets. Therefore, the early teaching of young scientists from the grape community fills the gap to hopefully make datasets comparable in the near future.”

trainee

## PRESENTATIONS



### TRAINER SLIDES

For more details please contact the organizers.

### REPORT

Presented at the The Second Annual Meeting of COST Action CA17111 (Ljubljana, SLOVENIA).

[Vitis phenotyping guidelines](#)

[Galaxy Resources](#)

[GREAT database](#)

[TOMEXPRESS platform](#)

## OVERVIEW



**TITLE:** METHADA 2020 – Transcriptomic Metadata Handling and Data Analysis

**DATE:** 5-7 February 2020

**VENUE:** Institute for Integrative Systems Biology (I2SysBio). Valencia, Spain

### **BRIEF PROGRAM:**

The rise of the latest technologies combining physics, optics, chemistry and its application to molecular biology has led to high-throughput experiments, yielding an explosion of publicly available data. This data ranges from Next Generation Sequencing (NGS) to transcriptomics, phenomics, metabolomics to large scale single cell data. In the case of transcriptomics, which generates to date the biggest amount of data compared to other omics, protocols for data submission are not fully standardized for grapevine data and not controlled by the research community. Public available gene expression datasets have a hidden true potential in the light of data reanalysis and integration. In line with the FAIR (Findable Accessible Interoperable Reusable) principles our next challenge as a community relies on correct sample and experiment annotations, using controlled vocabularies to ensure both human readability and computational tractability. This training school addresses transcriptomics data handling and analysis, and it is organized in two modules. On the first unit, trainees will work to learn how to correctly annotate experiments and handle metadata in order to exploit standards and bio-ontologies for data annotation. Secondly, attendees will be trained in a reduced set of foundational skills to analyze and explore transcriptomic datasets, including resources freely available for the grapevine community. All trainees will learn on how to use Jupyter Notebook, an open-source language-agnostic web application (it supports over 40 programming languages including Python and R) that allows a wide range of workflows in data science and scientific computing without burdening the users with installation and maintenance tasks



Marco Moretto

Computational Biology Unit,  
Fondazione Edmund Mach -  
Istituto Agrario San Michele  
All'Adige



Paolo Sonogo

Computational Biology Unit,  
Fondazione Edmund Mach -  
Istituto Agrario San Michele  
All'Adige



Elie Maza

INP-ENSAT Toulouse



José Tomás Matus

I2SysBio, Valencia



Jérôme Grimplet

CITA-Aragon.es

**Day 1 (5th Feb, 9:00-18:00)**

9:00	Welcome from I2SysBio Director/Vice-director and COST Action MC Chair (Mario Pezzotti). Presentation of the Training School (Jerome Grimplet, Tomás Matus).
9:30	Introduction and setting up of the Jupyter Notebook working environment (Marco Moretto).
10:30	Coffee
11:00	Metadata handling (Marco Moretto). -Background. General standards: ISA tab tools, MIAMI protocol. -Recurrent pitfalls in metadata submission. The NCBI-SRA case.
13:00	Lunch
14:30	Metadata handling (Continue, Marco Moretto and Tomas Matus). -European metadata cases: upload from excel using ABI (Application Binary Interface). -Relation between experimental designs and metadata annotation. Complex versus complicated designs.
15:45	Coffee
16:00	Best practices and useful tips for the analysis of gene expression data (Paolo Sonogo). -How to take decisions regarding: type of sequencing, number of replicates, depth, where to map reads (genome vs transcriptome, genome accessions) and annotations.

**Day 2 (6th Feb, 9:00-18:00)**

9:00	Transcriptomics data analysis (Paolo Sonogo) -Practical session with a typical workflow on grapevine real data downloaded from GEO/SRA. DE Analysis, EDA and visualization of transcriptomics data using R/Bioconductor (Paolo Sonogo).
10:30	Coffee
11:00	Transcriptomics data analysis (Continue)
12:00	Tools available at Galaxy for transcriptomic analysis (Jerome Grimplet). <a href="#">SLIDES</a>
13:00	Lunch
14:30	Resources: The TomExpress RNA-Seq platform, what can we learn from the tomato community regarding public RNA-Seq data handling, visualization and mining? (Mohamed Zouine) <a href="#">SLIDES</a>
16:30	Coffee
16:45	Grape-specific platforms (first part): -GREAT (video-remote lecture: Camille Rustenholz, Amandine Velt, INRAE COLMAR) <a href="#">SLIDES</a> -VitisNet/eFP-Browser Corvina (Jerome Grimplet).
21:00	Gala Dinner

**Day 3 (7th Feb, 9:00-16:00)**

9:00	Grape-specific platforms Marco Moretto) -Tutorial and Outputs from the VESPUCCI platform.
10:30	Coffee
11:00	Access to Phyton/R packages in Jupyter to see how VESPUCCI is built.
12:00	Metadata annotation in VESPUCCI as a case of study.
14:00	Lunch
15:30	Discussion and roundup of Training school (Mario Pezzotti).

COST gives each eligible trainee 700 euros as a financial contribution towards overall travel, accommodation and meal expenses of the Grantee.

Priority will be given to early career researchers and eligible candidates must be from COST member countries or MC Observers from Near-neighbouring countries (eligibility details available on pages 30/31 of the COST Vademecum [https://www.cost.eu/wp-content/uploads/2019/07/Vademecum\\_June2019.pdf](https://www.cost.eu/wp-content/uploads/2019/07/Vademecum_June2019.pdf)). Evaluation of applicants will be based on the following criteria:

Postdocs, PhD students and early career PIs working in transcriptomics, genomics, and computational biology-related aspects of grapevine, with high motivation in spreading good metadata handling practices learned during the training school.

The training school is directed to computational biologists or biologists/biotechnologists with the following bioinformatics skills:

- Basic knowledge of bioinformatics and computational biology theory
- Basic programming in R and/or Python.

Selection of participants will be made by COST Integrate local organisers on the basis of all submitted data:

- Reason for participating
- Curriculum vitae
- Recommendation letter (in case of PhD students and Postdocs)



## Dr José Tomás Matus

Institute for Integrative Systems Biology  
Carrer del Catedràtic Agustín Escardino Benlloch,  
46980 Paterna, València  
[tomas.matus@uv.es](mailto:tomas.matus@uv.es)

## Dr Jérôme Grimplet

Centro de Investigación y Tecnología Agroalimentaria de Aragón  
Avda. Montañana 930  
50059 – Zaragoza  
Phone: +34 976 71 3635  
[jgrimplet@cita-aragon.es](mailto:jgrimplet@cita-aragon.es)