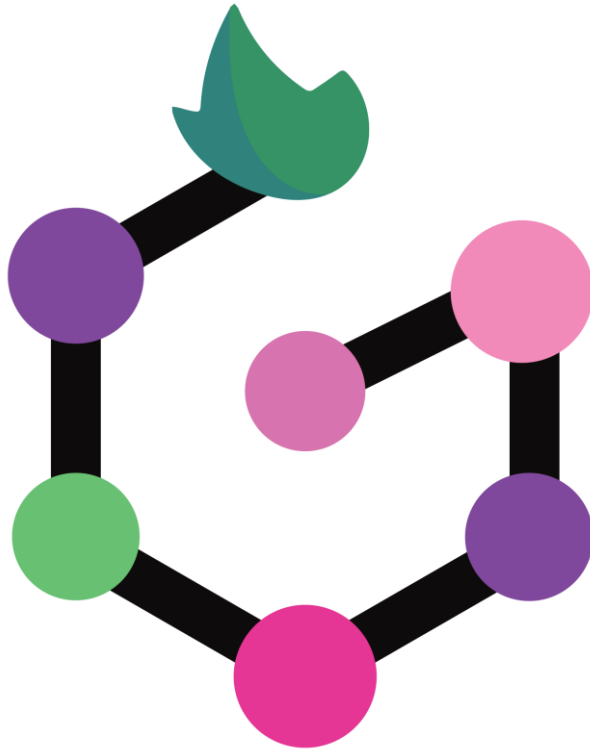


Grapedia

The Grapevine Genomics Encyclopedia



The Annual Meeting Booklet

11-13th September 2023, Valencia (Spain)

The birth of a centralized, federative portal



Foreword

Welcome to the GRAPEDIA Annual Meeting, hopefully the first of many future meetings. We have gathered here the latest advances in computational biology studies in grapevine, including data integration, pangenomes, annotation pipelines, QTL database development and so much more. We hope that you enjoy these three days to also explore the various facets of GRAPEDIA's creation.

The idea behind GRAPEDIA was conceived over four years ago but it was really born within the end of the COST Action Integrape CA17111. Thus, a special thankyou to the drivers behind that Action (particularly, to Mario Pezzotti and Anne-Françoise Adam-Blandon) and all the people who contributed to it by settling the bases for standardized data in grapevine. We are now ready to integrate resources into one single platform.

It has been a year of many activities, engaging researchers from all over the world to implement a first prototype to be delivered by the end of 2023. During the 2022-2023 Grant Period of the COST Innovators Grant IG17111, we have launched scientific missions, training schools, hackathons and workshops in order to build GRAPEDIA. But we are not ready. We are continuing to a beta testing model and then to a final product. Extra sources of data need to be implemented, such as phenotyping and QTL data, and why not also dream of climate data? Together we hope GRAPEDIA will enable researchers and the industry to design the vineyard of the future with more quality products, resistant to pests and climate transition.

We would like to acknowledge our sponsors Gallo Winery, Novatech (Mercier Group) and Sequentia Biotech, for believing



in our idea and funding us whenever needed. Special thanks to our database collaborators, the grapegenomics.com portal and Gramene. Also, much mention needs to be given to particular features being implemented in GRAPEDIA, such as ChromoMap and Nextflow-based workflows.

We also invite you to be a part of our initiative. Become an ambassador of GRAPEDIA and help us disseminate the portal and its benefits to the entire grapevine community.



José Tomás Matus
GRAPEDIA CIG Chair
Convener

Camille Rustenholz
GRAPEDIA CIG Vice-Chair



Preface

In the pursuit of a centralized resource for grapevine-related studies.

By David Navarro-Payá, José Tomás Matus and Marco Moretto.

Established research communities dedicated to model organisms have developed and maintained a diverse array of online tools and databases, funded by both public and private sources, which serve as repositories for genome and gene function data. However, a significant gap exists when it comes to resources for grapevine, an economically vital non-model species. This gap has spurred the grapevine community to amass a collection of valuable assets, including the first grape gene reference catalogue, reference genome sequence assemblies, functional gene annotations, guidelines for generating various omics data, phenotypic data collections, *Vitis* ontologies, *Vitis* variety catalogues and grapevine-specific gene expression databases. Alongside these data resources, a substantial number of grapevine-specific analysis tools have been developed, encompassing data visualization, gene networks and metabolic pathways analysis, among others.

GRAPEDIA (GRAPE Genomics EncycloPEDIA) is a community-driven project catering to both the scientific community and grape and wine industries. It centralizes these previously mentioned resources within a single accessible database, offering customized services for researchers and industry professionals, thus providing economic value to this platform. GRAPEDIA stands as a multi-tier application with multifaceted architecture.



The backbone of this system, accessible via a programmatic GraphQL interface, serves as a semi-federated database designed to integrate diverse grapevine resources into a coherent and unified data model. At its core, this data model revolves around the concept of the gene. It encapsulates both qualitative and quantitative information pertaining to genes, like gene annotation, genomic coordinates, genome-specific insights, gene expression profiles and DNA-binding details, among many others.

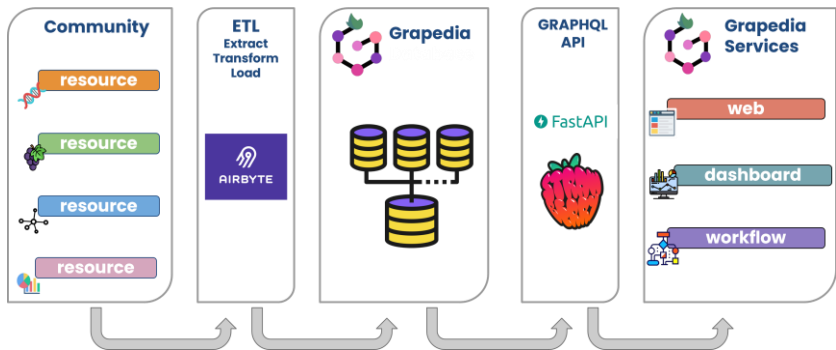
Lastly, GRAPEDIA extends its commitment to transparency and accessibility by providing comprehensive documentation for both developers and users. It also shares guidelines and best practices to support scientific reproducibility. Importantly, all the code developed for GRAPEDIA is open-source, fostering collaboration and community-driven enhancements.

The GRAPEDIA online tools and database will play a pivotal role in harnessing the potential offered by genomic advancements, benefiting both the research and industry sectors. In the realm of plant biology, these resources will be instrumental in aiding crop improvement efforts, whether through selective breeding or biotechnological strategies. Such advancement relies heavily on comprehensive knowledge of a plant species' genome, pangenome, and its constituent genes. Even though this pursuit of knowledge is particularly challenging for grapevine (*Vitis vinifera* L.), one of the world's oldest and most widespread fruit crops, we now live in the best era for standardizing and integrating the large and diverse public datasets being developed by the grapevine community.



GRAPEDIA Portal v1.0: A full-stack prototype to integrate resources and services for the grape community

The GRAPEDIA portal prototype has been developed over the course of last year resulting in a community-driven platform where grapevine-related resources are centralized, interconnected and integrated following a semi-federative database structure. The full stack GRAPEDIA web application is composed of the database Data Lake, the unified data model exposed through the GraphQL interface and the services, such as workflows, dashboard and the web portal.



The portal serves as a central hub that consolidates information from diverse sources. It has a uniform look and feel while implementing robust access controls and standardized procedures across multiple applications and datasets. Functioning as another multi-tier web application, it leverages



NodeJS for backend operations and ReactJS for frontend interfaces. The portal orchestrates the execution of various tools via a queue system and isolated environments, facilitates the presentation of dynamic dashboards, and offers workflows for download or cloud-based execution.

The backend consists of an extract transform and load (ETL) toolkit (Airbyte) which accesses information from heterogeneous sources (collectively referred to as our *data lake*) and transfers it to the SQL database run with PostgreSQL into a more organised scheme.

A unified data model has been designed carefully to deal with the complex SQL queries that must be performed at the frontend's request. The GraphQL interface exposes this unified data model and meets frontend requests to enable querying of both the GRAPEDIA DB itself and other decentralised databases, staying true to the semi-federative notion. The frontend uses React (i.e., ReactJS), an open-source JavaScript library, used to build user interfaces for web applications. An architecture based on components has been pursued with code readability and reusability in mind whilst providing a dynamic and interactive frontend experience. A particular emphasis has been placed on the user experience through the organisation of the different tools and data resources into multi-output dashboards wherever a common input is required.

The GraphQL interface is the gateway to the knowledge base, and enables the user to select and retrieve specific fields of interest allowing them to filter, constrain, and arrange results according to their needs. Many services are built on top of it. These services encompass a spectrum of functionalities, ranging from bioinformatic workflows and interactive dashboards to the



GRAPEDIA web portal. **Bioinformatic workflows** seamlessly amalgamate various programs into a unified pipeline, ensuring portability, reproducibility, and scalability. During execution, each workflow establishes an isolated environment that incorporates all requisite software components through Docker. It proceeds to retrieve datasets from GRAPEDIA and executes each step of the bioinformatic pipeline via Nextflow.

The **Dashboards** represent a data visualization technique, offering quick and concise insights, predominantly focused on gene-related information. These dashboards employ an array of visual elements, including plots, charts and tables. Data is fetched from GRAPEDIA and visualized using ReactJS components, along with additional JavaScript libraries, all embedded within the GRAPEDIA web portal.

Several international scientific collaborations (through STSMs and other means) have addressed the technical and logistical aspects required in this first prototype. Much more work is now needed to integrate additional data sources and to include user feedback. We envisage GRAPEDIA as an outlasting resource beyond the current COST Action. In fact, a beta testing operative portal has been planned to be released by 2024. We hope that the grapevine community will benefit from GRAPEDIA and will fuel the project for years to come.



Annual Meeting Program

Monday 11th September

- | | |
|---------------|--|
| 9:00 – 9:30 | Onsite registration |
| 9:30 – 10:15 | Welcome Speeches by Jardín Botánico Vice Director Olga Mayoral and Grapedia CIG Chair J. Tomás Matus |
| 10:15 -10:45 | Grapedia Portal: Structure & Implementation (Marco Moretto) |
| 10:45 – 11:15 | Coffee Break |
| 11:15 – 11:45 | Grapedia Portal: Scientific Roadmap (Marco Moretto) |
| 11:45 – 12:00 | Scientific mission – <i>Integrating community resources into the Grapedia database with Airbyte</i> (Víctor García-Carpintero) |
| 12:00 – 12:30 | Scientific mission – <i>Towards a new grape gene reference catalogue and the Gene Cards app</i> (David Navarro-Payà) |
| 12:30 – 13:00 | Scientific mission – <i>Visualization of gene co-expression networks (AggGCNs) in Grapedia</i> (Luis Orduña) |
| 13:15 – 14:30 | Lunch Break |
| 14:30 – 15:15 | Scientific mission – <i>A new annotation effort for the T2T genome</i> (Antonio Santiago / Camille Rustenholz) |
| 15:15 – 15:45 | Scientific mission – <i>Implementing the ChromoMap tool</i> (Lakshay Anand – Online) |
| 15:45 – 16:00 | Grapedia tutorial: Dashboards (Luis Orduña) |
| 16:00 – 16:30 | Coffee Break |
| 16:30 – 17:00 | OPEN QUESTIONS AND ROUND TABLE |



17:00 – 17:45 Plenary: Ana Conesa (Institute for Integrative Systems Biology, CSIC-UV, Spain): *"Using long reads technologies to reveal new insights into the functional complexity of the transcriptomes"*

Tuesday 12th September

9:10 – 9:55 Plenary: Doreen Ware (Cold Spring Harbour Laboratory, US Dept Agriculture, USA): *Plant Genomes: Past, Present and Future*

9:55 – 10:20 Talk 1: Pascual Villalba Bermell (Institute of Integrative Systems Biology, Spain): *Genome-wide identification of long non-coding RNAs (lncRNAs): a pilot study using the PN40024 reference genome*

10:20 – 10:45 Talk 2 – Corinna Thoben/Boas Pucker (Institute of Plant Biology & BRICS, TU Braunschweig, Germany): *Automatic approaches for cross-species knowledge transfer based on orthology*

10:45 – 11:15 Coffee Break

11:15 – 12:00 Plenary: Dario Cantu (UC Davis, USA): *Leveraging diploid genome assemblies and pangenomes to advance the science of grapevines*

12:00 – 12:25 Talk 3 – Gabriele Magon (University of Padova, Italy): *Investigating the cistrome landscape of three main transcription factor families in grapevine*

12:25 – 12:50 Talk 4 – Stefania Pilati/Valter Cavecchia (Fondazione Edmund Mach/ CNR-IMEM, Italy): *Using OneGenE and Grapedia Resources to Support Research in Grapevine Physiology*



- 12:50 – 13:15 Talk 5 – Maria Nikoghosyan (Armenian Bioinformatics Institute, Armenia): *Sommelier: Genomic data analysis and functional annotation of grapes with emphasis on Armenian wild and cultivated associations*
- 13:15 – 14:30 Lunch Break
- 14:30 – 15:00 Stakeholder Presentation: Olivier Zekri (Mercier Groupe & Novatech, France): *Cultivating the Future: How GRAPEDIA can support Novatech in Creating Novel Grapevine Genetic Resources*
- 15:00 – 15:30 Stakeholder Presentation: Felipe Gainza (Concha y Toro Winery, Chile): *Shaping the Vineyard of the Future: Viña Concha y Toro Experience, Challenges and the Acceleration through GRAPEDIA's Efforts*
- 15:30 – 16:00 Stakeholder Presentation: Walter Sanseverino/Marco Di Marsico (Sequentia Biotech S.L., Spain): *How Sequentia's services within GRAPEDIA will help the Grapevine community*
- 16:00 – 16:30 Coffee Break
- 16:35 – 17:20 Plenary: Pablo Carbonell (Instituto de Ciencias de la Vid y el Vino, Spain): *Genome assemblies and long-read sequencing to search for functional variation in grapevine*

Wednesday 13th September

- 9:30 – 10:20 Plenary: Silvia Vezzulli/Laura Costantini (Fondazione Edmund Mach, Italy): *The Vitis-QTL Browser: A Dream Coming True (online)*
- 10:20 – 10:45 Talk 6 – Paola Bettinelli (Fondazione Edmund Mach, Italy): *Rgb3: an organ-specific QTL for bunch resistance to black rot identified in the hybrid cultivar 'Merzling'. Hands-on demo of the potential of GRAPEDIA*



- 10:45 – 11:15 Coffee Break
- 11:15 – 11:40 Talk 7 – Diego Intrigliolo (Desertification Research Center, CSIC, Spain): *Phenotyping of 'Bobal' grapevine biotypes using remote sensing, plant physiology and genetic tools*
- 11:40 – 12:05 Talk 8 – David Carrasco (Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, CSIC, Spain): *Modelling the response of drought tolerant and sensitive grapevines shows different strategies at transcriptomic, metabolic and physiological level.*
- 12:05 – 12:30 Talk 9 – Kristina Margaryan (Institute of Molecular Biology NAS, Republic of Armenia): *Armenian Grape Genome Project: from vineyard to innovation*
- 12:30 – 13:15 Plenary: Pietro Liò (Computer Science and Technology of the University of Cambridge, UK): *Multi omic, multi scale models using computational biology and deep learning*
- 13:15 – 13:30 WRAPUP OF MEETING

Organizing committee:

José Tomás Matus Picero

Luis Orduña Rubio

David Navarro Payá

Antonio Santiago Pajuelo

Anne-Marie Digby

Aureliano Bombarely



Using long reads technologies to reveal new insights into the functional complexity of the transcriptomes

Dr. Ana Conesa

Institute for Integrative Systems Biology, CSIC-UV, Spain

Biography:

Dr. Ana Conesa is a Senior Scientist in Bioinformatics and Computational Biology and a professor in bioinformatics at the University of Florida, where she leads the Genomics of Gene Expression Lab. She is also the co-founder of Biobam Bioinformatics, a spin-off company that specializes in user-friendly software for biologists. Dr. Conesa holds an Engineering degree from the Polytechnical University of Valencia (1993) and a PhD in Molecular Microbiology from Leiden University in The Netherlands (2001). She has built a scientific program focused on the development of algorithms and bioinformatics solutions for the analysis of Big Genomics Data, particularly gene expression and its impact on phenotype. Dr. Conesa has conceived and led the creation of over a dozen bioinformatics tools that apply to all kinds of species, including model and non-model organisms, plants, microbiomes, and humans. These tools include Blast2GO, maSigPro, NOISeq, Paintomics, SQANTI, and TAPPAS. She regularly serves on funding awarding committees for Spanish, European, and US research agencies. Dr. Conesa's contributions to the field of bioinformatics have been widely recognized and her work continues to have a significant impact on the analysis of Big Genomics Data. In January 2022, she was elected Member of the Spanish Royal Academy of Engineering.



Plant Genomes: Past, Present and Future

Dr. Doreen Ware

Cold Spring Harbour Laboratory, US Dept Agriculture, USA

Biography:

Dr. Doreen Ware is Molecular Biologist with the USDA ARS and Adjunct Professor at Cold Spring Harbor Laboratory. She is recognized as a leader in plant genomics and bioinformatics. Over the last several years her laboratory has contributed to collaborative projects supporting genomics enabled science, with a focus on understanding plant genome architecture and the impact on complex traits, including grain yield and response to nitrogen. Dr. Ware's group has contributed to the development of reference genomic resources for rice, maize, sorghum and grape and Cyberinfrastructure projects to support access and integration of genome scale data. Dr. Ware serves in several leadership positions within the plant science community, and served as acting CSIO for USDA ARS from 2014-2017.



Genome-wide identification of long non-coding RNAs (lncRNAs): a pilot study using the PN40024 reference genome

Pascual Villalba Bermell

Institute of Integrative Systems Biology, Spain

Abstract:

Long non-coding RNAs (lncRNAs) are regulatory non-coding RNAs longer than 200 nucleotides characterized by low coding potential, low expression, low conservation at sequence level and tissue specificity. Numerous lncRNAs have recently emerged as important regulators of many biological processes, including stress response, reproduction and vegetative growth. As a consequence of the expanding list of lncRNAs and accumulating evidence of their functions in plants, comprehensive databases have emerged for lncRNA research. Here, we present a fast and easy-to-use pipeline to predict potential lncRNAs in plants that creates a database consisting of three different confidence levels. As a pilot study, we executed this pipeline using two unpublished RNA-seq experiments (corresponding to 17 samples, representing different organs and experimental conditions) and the *Vitis vinifera* PN40024 reference genome (V4 assembly). We identified 6818, 990, 1873 and 347 potential and non-redundant lincRNAs, NAT-lncRNAs, int-lncRNAs and SOT-lncRNAs, respectively. Many improvements are still needed, but the results are a good starting point for further analysis and further refinement of the pipeline accuracy.



Automatic approaches for cross-species knowledge transfer based on orthology

Corinna Thoben/Dr. Boas Pucker

Institute of Plant Biology & BRICS, TU Braunschweig, Germany

Abstract:

The significant increase of available plant genome sequences and the development of sophisticated approaches for the prediction of gene models turned the functional annotation into a new challenge. Assuming that orthologs have the same function, knowledge can be transferred between species. An approach for the automated annotation of pathway enzymes is implemented in Knowledge-based Identification of Pathway Enzymes (KIPes). Comprehensive sequence sets and knowledge of functionally relevant amino acid residues are combined to identify enzyme candidates. Putative enzyme functions are assigned based on phylogenetic distance to bait sequences and functionally relevant amino acid residues are inspected in candidates to predict functionality. Imperfect candidates are targets to study sub- and neofunctionalization or pathway blocks. The MYB_annotator and the bHLH_annotator have been developed for the automated annotation of transcription factor families. Initial candidates are identified by sequence similarity to bona fide bait sequences. Non-homolog candidates are filtered by phylogenetic distinct outgroup sequences. Candidates with a diverged domain and candidates forming new phylogenetic clades are preserved to allow investigation of functional diversification. Functional annotations of previously characterized reference sequences are assigned to candidates based on orthology and family-specific characteristics are analyzed. In conclusion, the approaches utilize existing knowledge to provide a detailed investigation of pathway enzymes and transcription factor families. The identification of orthologs facilitates cross-species comparison.



Leveraging diploid genome assemblies and pangenomes to advance the science of grapevines

Dr. Dario Cantù

UC Davis, USA

Biography:

Dario Cantù is professor and Louis P. Martini Endowed Chair at University of California Davis. He earned his PhD in Plant Biology from UC Davis in 2009. He is a pioneer in the development of diploid phased genomes and reference-free pangenomes, which capture the diversity in the *Vitis* genus at nucleotide resolution and facilitate genetic association studies for agronomic and enological traits in wild grape populations. His research has significantly advanced our understanding of grape evolution, flower sex determination, and grape-pathogen interactions. He and his team have established and maintain www.grapegenomics.com, a public web portal providing access to genomic resources for both cultivated and wild grapes.

Abstract:

This seminar will focus on the challenges and opportunities of studying the genomes of *Vitis* species, characterized by exceptionally high levels of heterozygosity and frequent interspecific introgressions. Grape genomics achieved a crucial breakthrough when assembly methods capitalized on long-read sequencing. This led to more contiguous assemblies and phased heterozygosity in diploid haplotypes. We have utilized these tools in conjunction with genetic maps to create additional resources that facilitate the construction of chromosome-scale diploid assemblies. Recently, we developed a high-density map using 19,049 single-copy orthologous genes, enhancing the speed and completeness of diploid chromosome building. A new version of the Cabernet Sauvignon genome, built on HiFi and this new map, will be presented. This version features fully phased chromosomes achieved



by correcting haplotype switches using parental genome assemblies of Cabernet Franc and Sauvignon Blanc. The seminar will also provide examples highlighting the significance of diploid genome references in grape research. Lastly, the shift to pangenomes for the *Vitis* genus will be discussed. This transition will be explored in relation to the genetic association analysis of natural populations that often interbreed, resulting in interspecific introgressions. This is particularly relevant for genes related to both biotic and abiotic stress resistance. A graph nucleotide-scale-based super-pangenome of the North American *Vitis* species from diploid chromosome-scale assemblies will be presented, along with the application of novel panGWAS approaches to study the genetic basis of salt tolerance.



Investigating the cistrome landscape of three main transcription factor families in grapevine

Dr. Gabriele Magon

University of Padova, Italy

Abstract:

In the post-genomics Era, the enormous amount of data generated by omics technologies being stored in public databases considerably exceeds the analytical capacities of humans, making it imperative to use increasingly powerful computational resources to process, analyse and integrate this information. Nowadays, several studies aim at specifically addressing this issue in grapevine (*Vitis vinifera* L.), a plant which is quickly establishing itself as an appealing 'model' species for studying non-climacteric fleshy fruits. A large fraction of phenotypic variation appears to be determined by regulatory rather than coding variation and understanding how gene expression is controlled becomes a prerequisite for the exploitation of the full potential of genome editing techniques. In this regard, to fully decipher the molecular mechanisms at the basis of many biological processes, such as phenotypic plasticity and fruit development, is mandatory to focus efforts on the study of transcriptional regulators activity. About that, DNA-Affinity Purification Sequencing (DAP-Seq) is the ultimate approach for binding sites discovery and provides a scalable alternative for non-conventional model species where genetic transformation is difficult, to rapidly and inexpensively investigate large numbers of TFs. DAP-Seq is a high-throughput experimental procedure that allows to discover transcription factors (TFs) binding sites interrogating genomic DNA with *in vitro*-expressed TFs. In this study, we took advantage of DAP-Seq to investigate the cistrome landscape of a large set of grapevine TFs, belonging to MYB-R2R3, WRKY and NAC families. The DAP-Seq reads were aligned against the PN40024 genome, the reference for the species and the resulting peaks annotated in order to



unveil the candidate targets of every TFs. Since wet lab procedures were performed on Cabernet Franc biological material and taking advantage of the newly released phased genome of this cultivar, the procedure was repeated aligning all the DAP-Seq reads on each of the two Cabernet Franc haplotypes. Subsequently, our attention was focused on a specific TF, namely MYB108A. This transcriptional regulator was observed to be highly specific for anther development, and taking advantage of some systems biology *in silico* analysis, a subset of target genes was isolated. These genes were designed as “High Confidence Targets” (HCTs) and represent a subset of multi-level validated genes, which are biologically related to anther, strictly expressed in that tissue and regulated by MYB108A. HCTs were used to generate tissue-specific regulatory networks to infer the function of the TF based on the identification of its target genes and suggesting a direct participation of MYB108A in male organs specific processes. Finally, the putative role of MYB108A was empowered by a *cis* regulatory elements enrichment analysis, finding in this way the most probable binding site, which give back a further confirmation of involvement in anther and pollen developmental specific mechanisms of this TF.



Using OneGenE and Grapedia Resources to Support Research in Grapevine Physiology

Dr. Stefania Pilati

Dr. Valter Cavecchia

Fondazione Edmund Mach, Italy

Abstract:

Causal relation discovery from observational data is a recent research topic that is being actively studied both theoretically and technically. The OneGenE project applied a causal inference method, the PC algorithm, to big homogeneous datasets of transcriptomic data, such as the Vespucci v1 (2016) and the more recent Vespucci v2 (2021) datasets for *Vitis vinifera*. To deal with the high computational complexity and large amount of input data, OneGenE method was applied in the framework of distributed computing, within the BOINC project based on volunteers' availability to run OneGenE on their desktops. The output is represented by a list of causally associated genes for every input gene of the Vespucci matrix, based on the 12X.V1 release of grapevine genome. The next step has been the delivery of this information to the biologists through a website and the development of tools which take it as input for further exploration and analysis. The final aim would be to reconstruct gene regulatory networks, where genes are connected by means of oriented edges representing causal relationships. This model would support gene functional studies and system biology approaches to interpret relevant aspects of plant physiology. The information has been published according to the FAIR principles of data management, and the opportunity to integrate it inside Grapedia is unvaluable to improve its accessibility and usability, as already occurring for the gene annotation integrated from the Gene Ref Catalogue resource. Regulatory networks related to floral initiation and organ specification will be presented as examples.



Sommelier: Genomic Data Analysis and Functional Annotation of Grapes with Emphasis on Armenian Wild and Cultivated Accessions

Maria Nikoghosyan^{1,2} and Hans Binder¹

1- Armenian Bioinformatics Institute, Armenia

2- Bioinformatics Group at The Institute of Molecular Biology; Yerevan Armenia

Abstract:

Comprehensive genome studies of both wild and cultivated grapes have yielded groundbreaking insights into various aspects of this plant. These encompass the detailed characterization of genomic diversity, exploration of geographical connections, investigation of historical origins with comprehensive evaluations, characterization of associations with various phenotypic traits, and the identification of genetic markers. Therefore, there is a pressing need to develop innovative approaches capable of efficiently analyzing vast genomic datasets while providing user-friendly visualization and annotation tools for these genomic markers.

In this context, we used Self-Organizing Maps (SOM), a neural network-based machine learning method, implemented in the “oposSOM” R-package, to analyze whole-genome genotype data obtained from 160 Armenian wild and cultivated grape samples. SOM analysis effectively reduces the dimensionality of genotype data by clustering genomic variants with similar distributions across samples into distinct clusters. These clusters are then visualized as two-dimensional grids known as SOM maps, offering a visual representation of the distribution of genomic variants within each sample and group.

Furthermore, we correlated these SOM maps with various phenotypic traits, including grape utilization, berry skin color, and resistance to powdery mildew (for Armenian wild samples). This correlation provides



valuable insights into the genomic variations highly associated with specific phenotypes, offering a comprehensive visual representation of this intricate relationship.



Cultivating the Future: How GRAPEDIA can support Novatech in Creating Novel Grapevine Genetic Resources

Dr. Olivier Zekri

Mercier Groupe & Novatech, France

Biography:

Olivier ZEKRI is an agronomist and serves as the Research and Development Manager for Mercier Vine Nurseries, a global leader in grapevine plant production. As one of the Deputy Directors of this renowned group, comprising over 250 employees, they oversee the production of more than 20 million vine plants worldwide. At the helm of Novatech Laboratory, a specialized research facility dedicated to viticulture, Olivier ZEKRI spearheads innovative advancements in the field.

With a profound passion for agriculture and viticulture, Olivier's expertise has been instrumental in driving cutting-edge research initiatives that aim to create novel genetic resources and improve young vines cultivation techniques.

Abstract:

NOVATECH, the R&D department of the Mercier Group, a leading producer of vine plants, stands at the forefront of innovation in grapevine genetics. Our mission is to create genetic resources for grapevines that push the boundaries of quality and resilience. To achieve this, we leverage the power of the Grapedia (Grapevine Genomics Encyclopedia) Initiative, a central open-access portal that offers a wealth of agronomic and genetic data. Grapedia not only facilitates data collection and integration but also provides us, as a private enterprise, with unprecedented access to the wealth of knowledge within the global scientific community.



With Grapedia as our trusted ally, NOVATECH aspires to pioneer a new range of grapevine varieties. These varieties will not only exhibit enhanced disease resistance but will also be finely attuned to our unique environmental conditions. Our partnership with Grapedia represents a significant leap toward achieving this ambitious goal and advancing the field of viticulture



SHAPING THE VINEYARD OF THE FUTURE: Viña Concha y Toro Experience, Challenges, and Acceleration through GRAPEDIA's Efforts

Dr. Felipe Gainza Cortés

Concha y Toro Winery, Chile

Biography:

Felipe Gainza holds a Bachelor's degree in Biotechnology Engineering from the Universidad Tecnológica de Chile (2005) and earned his Ph.D. in Plant Genetic Engineering from the Institute of Plant Biology and Biotechnology at the Universidad de Talca (2011). From 2012 to 2016, he led the genetic improvement program for *Prunus* spp. rootstocks at the Center for Advanced Studies in Fruit Cultivation (CEAF) in Chile. Currently, he serves as the R&D Leader in Molecular Biology at the Research and Innovation Center of Viña Concha y Toro, heading the Genetic Resources and Plant Health Program through the use of biotechnological tools that add value and seek applied solutions for the industry.

Abstract:

Viña Concha y Toro, through its Research and Innovation Center (CRI), will delve into the transformative work undertaken by the Genetic Resources and Plant Health Program. This program has spearheaded advances in plant health by harnessing state-of-the-art molecular biology techniques, including qPCR and LAMP, to develop innovative diagnostic systems. Moreover, the program has harnessed techniques like meristem rescue and thermotherapy to purify plant material. Our endeavors have extended into the realm of varietal and clonal identification, making use of cutting-edge molecular markers. Furthermore, we have embraced precision breeding technologies within the genetic resources domain. We are confident that a pivotal force in our journey will be the GRAPEDIA platform, poised to



significantly influence our strategies. This platform, a repository rich in grapevine genetic information, will play a pivotal role in identifying candidate genes and unraveling their functions. Designed as a federated system, GRAPEDIA's database is anticipated to streamline our research processes and expedite the pursuit of the vineyard of the future. By hosting comprehensive data and insights, it will pave the way for a swifter and more informed decision-making process. The potential collaboration between Viña Concha y Toro and GRAPEDIA underscores the fusion of innovation and data-driven research, propelling us closer to our vision of an advanced and resilient vineyard ecosystem.



How Sequentia's services within GRAPEDIA will help the Grapevine community

Dr. Walter Sanseverino

Dr. Marco Di Marsico

Sequentia Biotech S.L., Spain

Biography:

Marco is an expert bioinformatician at Sequentia Biotech and he holds a PhD in plant genomics from the University of Perugia, Italy. Related to Grapedia, Marco is the coordinator of the Grapedia project within Sequentia Biotech and one of the organizers of the Grapedia Training School done in Barcelona.



Genome assemblies and long-read sequencing to search for functional variation in grapevine

Dr. Pablo Carbonell-Bejerano

Instituto de Ciencias de la Vid y el Vino, Spain

Biography:

Pablo Carbonell Bejerano is a CSIC permanent researcher at the Institute for Grape and Wine Sciences (ICVV, La Rioja, Spain) since 2022. After completing his PhD thesis at the IBMCP (Valencia, Spain, 2008) studying the molecular control of fruit set capability in the model plant *Arabidopsis thaliana*, he performed several postdoctoral stages focused on grapevine research. Working at the CNB (Madrid, 2009-2011) he applied transcriptomics to understand grapevine berry ripening and its interaction with abiotic and biotic factors. Then he got introduced into genetics and genomics research at the ICVV (2012-2018) to identify the genetic origin of functional diversity in grapevine, including the origin of grape seedlessness and somatic variation in fruit colour. More recently, he moved to the Max Planck Institute for Biology (Tübingen, Germany, 2018-2022) to develop skills in long-read sequencing (PacBio and Nanopore) and bioinformatics to study plant genomic and epigenomic diversity, including the production of genome assemblies for relevant grapevine cultivars. The current research of Pablo Carbonell at ICVV comprise the genomic screening of improved natural somatic variants with increased agronomic quality and climate change adaptable traits.

Abstract:

Domesticated grapevines exhibit a high level of genome diversity among cultivars. In addition, grapevine cultivars are usually highly heterozygous, with great diversity between the two haploid complements comprising their diploid genome. Therefore, to increase the possibilities of capturing the genome variants responsible for



phenotype variation in traits of agronomic interest, we have developed cultivar-specific and diploid genome assemblies for relevant grapevine cultivars. To overcome the highly repetitive and heterozygous nature of grapevine genomes, we either combined noisy PacBio and Nanopore long reads with trio binning of haplotypes or used accurate HiFi PacBio long reads. The obtained genome assemblies, combined in specific cases with long read-based gene annotations and methylation analysis, were useful for detecting genome and epigenome variations associated with intra-cultivar variants displaying diversity in fruit colour or decreased bunch compactness.



The Vitis-QTL Browser: A Dream Coming True

Dr. Silvia Vezzulli

Fondazione Edmund Mach, Italy

Biography:

Dr. Silvia Vezzulli studies the genetics of disease resistance to support marker-assisted breeding in grapevine, with emphasis on downy and powdery mildew as well as black rot. Given her interest in new genetic material from wild species, she is adopting innovative strategies from both genotyping and phenotyping points of view. Her research interests include breeding and genetics, structural genomics, molecular biology, plant-pathogen interactions and botany. She has also been involved in the Grapevine Genome Sequencing project, where she studied the grape genome structure with a special emphasis on transposable elements.

Dr. Laura Costantini

Fondazione Edmund Mach, Italy

Biography:

Dr. Laura Costantini is currently investigating the molecular regulation of anthocyanin esterification, the developmental and genetic alterations associated with novel types of seedlessness in somatic variants, as well as the physiological and genetic mechanisms of resilience to heat stress in grapevine. She has previously focused on the identification of candidate genes for Muscat flavor through QTL and transcriptomic analyses and also contributed to gene functional validation with association studies and genetic transformation. Moreover, she has participated in the development of the Vespucci compendium of transcriptomic experiments.



Abstract:

Collection of results from different QTL (Quantitative Trait Loci) studies has been implemented both in model plants and in crops, mainly cereals, to facilitate the access to information and enhance the impact of existing data on plant improvement activities. To date, cataloguing, summarizing, and making FAIR the plethora of increasing QTL information are the next challenge for the grapevine (*Vitis* spp.) community. To ensure we do not lose valuable data, the multitude of identified QTLs - along with the related raw data like genotypes, phenotypes, and environmental information from published experiments - must be stored in dedicated public databases for effective utilization. Within the GRAPEDIA initiative, this wealth of QTL information is planned to be anchored firstly to the grapevine reference PN40024.v4 genome and secondly to the published and upcoming diverse genomes. The generated “Vitis-QTL browser” would (i) enhance our understanding of the genetic regulation of different phenotypes, (ii) reveal consistent QTLs across studies (consensus genomic intervals), which are particularly valuable for marker-assisted breeding deployment, (iii) assist the QTL cloning process and facilitate the integration of candidate genes and superior alleles into biological and biotechnological applications.



Rgb3: an organ-specific QTL for bunch resistance to black rot identified in the hybrid cultivar 'Merzling'. Hands-on demo of the potential of GRAPEDIA

Dr. Paola Bettinelli

Fondazione Edmund Mach, Italy

Abstract:

Crops are continuously exposed to the onset of emerging diseases and the exploitation of resistance donors in breeding programs is a widely used valid strategy. In the last two decades black rot (BR), caused by the ascomycete *Phyllosticta ampellicida* (syn. *Guignardia bidwellii*), has begun to threaten European viticulture in humid continental areas due to the advent of higher temperatures which favor the pathogenic cycle of the fungus. For this reason, a program has been established for the dissection of BR resistance trait and its introgression both in *Vitis vinifera* varieties and mildew resistant hybrids. A preliminary study was carried out to improve germplasm screening through different approaches. Historical phenotypic and pedigree information of resistance donors were gathered to support the decision-making process in breeding. A new propagation and inoculation strategy was developed to optimize and fasten inoculation experiments. Isolates of *P. ampellicida* were genetically characterized and combined for the assessment of resistance independent of race-specificity. Finally, these improvements were employed first for the identification of new BR resistant parental lines and breeding selections, and then for the quantitative trait locus (QTL) analysis in a segregating population derived from the cross 'Merzling' (hybrid, resistant) × 'Teroldego' (*V. vinifera*, susceptible). The screening of this progeny, under greenhouse and field conditions, allowed the discrimination between two distinct organ-specific QTLs on chromosome 14. The previously identified *Resistance to G. bidwellii* (Rgb)1 locus was confirmed associated with leaf/shoot resistance, while upstream a new QTL



designated *Rgb3* was discovered linked to bunch resistance. Driven and inspired by the INTEGRAPE community effort that led to GRAPEDIA (GRAPEvine -omics encyclopDIA), all the available tools and resources have been exploited for the exploration of the physical region of the two QTLs, providing a hands-on demo of the potential of this portal.



Phenotyping of ‘Bobal’ grapevine biotypes using remote sensing, plant physiology and genetic tools

D. S. Intrigliolo¹, J.M. Ramirez-Cuesta^{1,2}, R. López-Urrea^{1,3}, R. Ferrer¹, C. Gisbert⁴, R. Peiró⁴ and I. Buesa⁵

1 Desertification Research Center (CIDE) (CSIC-UV-GVA), Valencia, Spain

2 University of Catania, (Di3A), Catania, Italy

3 Instituto Técnico Agronómico Provincial Albacete, Albacete, Spain

4 Universitat Politècnica de València (COMAV), Valencia, Spain

5 Universidad de las Islas Baleares (PlantMed-UIB), Palma, Spain.

Abstract:

Bobal is the second Spanish red variety in extension, widely cultivated in the Valencian Community and especially in the Utiel-Requena region (about 67% of the 34,000 hectares). Despite this, in this variety, clonal selections were only carried out decades ago with the simple main goal of increasing grapevine productivity. Our research team is therefore carrying out a new clonal selection in an attempt to improve the Bobal's oenological potential. To this end, we are screening the agronomic and physiological characteristics of up to 98 biotypes selected in 2018 and 2019 and now planted in a vineyard under the 110-R rootstock. The results of the first two harvests with very young vines reflect a wide range of variation among biotypes for important agronomic and oenological parameters. Starting from this season, we are also determining the eco-physiological and genetic marker diversity for some of the most interesting biotypes. In addition, high-resolution phenotyping tools are being employed using UAV flights equipped with thermal, RGB, multi-spectral and Lidar sensors. Remote sensing determinations are being compared with on-the ground measurements of leaf area, plant water status and leaf gas exchange. In an attempt to



carry out genomic association studies, the established experimental platform is therefore an opportunity to begin new collaborative efforts in order to relate physiological, agronomic and oenological traits to the genomic diversity that may be encountered.



Modelling the response of drought tolerant and sensitive grapevines shows different strategies at transcriptomic, metabolic and physiological level.

Carrasco, D.^{1†}, Rodriguez-Izquierdo, A.^{1†}, Revilla, M. A.², López-Peredo, E.², Bota, J.³ & Arroyo-García, R.^{1*}

1 Centro de Biotecnología y Genómica de Plantas (CBGP-INIA), CSIC-Universidad Politécnica de Madrid, Campus Montegancedo, Madrid, Spain.

2 Department of Organisms and Systems Biology, Institute of Biotechnology of Asturias, University of Oviedo, Oviedo, Spain.

3 Research Group on Plant Biology under Mediterranean Conditions, Departament de Biologia, Universitat de les Illes Balears (UIB) – Agro-Environmental and Water Economics Institute (INAGEA). Carretera de Valldemossa Km 7.5, 07122 Palma, Illes Balears, Spain.

*Corresponding author: rarroyo@inia.csic.es

† Authors contributed equally

Abstract:

The reduction of water availability in vineyard caused by climatic change on grapevine is a fact that becomes higher along the years, causing several damages at crop level, but also at economical and ecological level in the world. However, not all the grapevine varieties have the same response to drought stress, becoming more tolerant determinate genotypes (near-isohydric) than others (anisohydric). Commercial varieties like Merlot are more sensitivity to water privation than some local varieties like Callet, with a genotype more related to drought stress events in the Mediterranean area. To understand the differences among these cultivars in front of drought stress tolerance, we explored these differences across transcriptomic, metabolic and physiologic analysis at three different levels of drought in the plant (mild,



high and extreme drought), using 110 Richter as rootstock in both varieties. Furthermore, a Differential Expression Gene Analysis were performed using RNA-seq data from roots and leaves. The results revealed that ABA sensitivity in grapevine is key to tolerate the drought stress, showing that Callet developed modulated more gene expression patterns in response to ABA, in order to avoid water losses compared to Merlot, and this modulation changed the metabolic pathways in order to increase cell turgor o produce molecules in response to oxidative stress, such as polyphenol or flavonoid molecules. The study also revealed a double response coming from the scion to the rootstock produced by that modulation, that could influence the rootstock behaviour too. For that, the combination of the selection of genotypes more ABA-sensitive and the choose of the appropriate rootstock is key to manage the vineyard in possible and the future of the grapevine in future scenarios caused by climate change.



Multi omic, multi scale models using computational biology and deep learning

Dr. Pietro Liò

Computer Science and Technology of the University of Cambridge,
UK

Biography:

Pietro Liò received the PhD degree in complex systems and nonlinear dynamics from the School of Informatics, dept of Engineering, University of Firenze, Italy and the PhD degree in Genetics (Genomics) from the University of Pavia, Italy. He is currently professor of computational biology with the Department of Computer Science and Technology, University of Cambridge and a member of the Artificial Intelligence Group.

He is also a member of the Cambridge Centre for AI in medicine, ELLIS (European Laboratory for Learning and Intelligent Systems), Academia Europaea. His research interests include graph representation learning: methods and biological applications.