

COST INNOVATORS GRANT APPLICATION - PART B

1. INNOVATION POTENTIAL

CONTEXT OF THE CIG PROPOSAL IN RELATION TO INTEGRAPE CA17111

Omics and large-scale phenotyping techniques allow the study of thousands of biological characteristics at once, generating huge datasets manageable only with the support of dedicated informatic tools. However, these datasets are often generated in different and non-standardized formats across research laboratories and often stored only locally, which makes them poorly accessible for further use. Within the [INTEGRAPE COST Action](#) (CA17111: "Data integration to maximise the power of omics for grapevine improvement"; Grant Period from 01/11/2021 to 09/09/2022) **the establishment of an open, international, and representative network was accomplished**, working at different levels insuring that omics and phenotyping data generated in the grapevine research community are being produced in a secure and standardized format, following the F.A.I.R. principles of findability, accessibility, interoperability, and reusability. **Amongst the most significant deliverables of CA17111 INTEGRAPE to highlight:** (1) the elaboration of [Guideline 'cookbooks' and Dictionary of unified grape-sample ontologies](#) (some of which have already being published; e.g., [Savoi et al., 2021](#)), providing recommendations for genomic and metabolomic data acquisition, repository submission and handling of datasets, together with recommended repositories and standard formats for organ naming; (2) the release of the [PN40024 fourth genome assembly and its annotation](#) (Velt et al., in preparation), serving as a platform for researchers to discuss and promote the diverse genomic tools associated to this reference genome, (3) the creation of the [Gene Reference Catalogue](#) ([Navarro-Payá et al., 2022](#)), building a repository for the standard annotation of grape genes and their functions, based on past and current gene/ gene family characterizations, and (4) the enlisting of [Online repositories and tools](#) for omics data exploration and visualization, which to date are not yet interoperable among them.

1.1 PRACTICAL CHALLENGE AND TARGET GROUP THAT WILL BE ADDRESSED IN CIG

Despite the huge amount of work done by the community since the first release of the grapevine genome sequence in 2007, and the recent contributions from INTEGRAPE for the standardization of the data, the scenario right now is that the community still feels overwhelmed by dispersed resources available. There is much room for improvement in the community, as these resources need to be made interoperable and centralized in one single platform, offering customized solutions to the whole community built by research teams, breeders and the wine industry considering the problems now faced by climate transition (practical challenge). The Arabidopsis Information Resource (TAIR) for the *A. thaliana* community represents a successful example in the plant field (Berardini et al., 2015), providing up-to-date biological information on individual genes as well as related resources such as single-gene mutant collections. Other examples of equivalent resources are the SolGenomics Network (SGN) for Solanacea family members such as tomato, potato, and tobacco (Fernandez-Pozo et al., 2015), the Rice Annotation Project Database (RAP-DB) (Sakai et al., 2013), and the Maize Genetics and Genomics Database (maizeGDB) (Woodhouse et al., 2021). Nevertheless, simply mimicking these databases for grape wouldn't be enough as, contrary to all the plant species previously mentioned, grapevine is a perennial crop, grafted, planted for decades and directly facing climate change, its genome is highly heterozygous and several clones may be available for a given cultivar. All these characteristics results in difficulties for creating a standard omics database, and to handle data and metadata, creating a need for the community to develop a database dedicated to grapevine.

The **target group** is the entire scientific community working on grapevine or using grapevine as their model plant for an "orphan" plant species, and also the private sector working on R&D in vitiviculture. There are several companies who have been involved in INTEGRAPE activities and have explicitly manifested their interest in the results of the CIG proposal and their availability to contribute with private funds in accomplishing this platform. Among these is worth mentioning Mercier Nurseries

(France), Concha & Toro Winery (Chile) and Gallo Winery (USA).

1.2 PROPOSED SOLUTIONS AND MAIN OUTCOMES OF THE CIG

As part of the legacy intended for the achievements of CA17111 INTEGRAPPE, the goal of this CIG proposal (**proposed solution**) is to integrate them to the resources already generated by the community to allow progress in the creation of a global grape database, named the **GRAPE Genomics EncycloPEDIA (GRAPEDIA)**. In this vision, especially related to the resources gathered, the CIG will seek a place where the whole community can connect and integrate the data generated in their research. It will disseminate and formally invite all researchers to form part of this broad initiative. The world's grape and wine industries will have full access to *up-to-date* information to meet their major challenges: to improve grape berry quality and produce better grape cultivars and wines. The CIG aims to centralize the resources in a single database, allowing data exploration and visualization and adding tools for comparative analyses, with open access to the public but also offering customized services for the research community and industry (**main outcome**), thus giving economic value to this platform. This database (DB) will combine omics, phenotypes and climate data with predictions to also help the scientific community and grape/wine companies with their design of next-generation cultivars that would be well adapted to climate change.

1.3 OBJECTIVES OF THE CIG

The CIG has two main objectives that will be implemented as work packages (WP) in the CIG grant period*: 1) To **build GRAPEDIA** and make it fully available and open access; and 2) To develop a customized service for grapevine **data integration and modelization** that can fit a business plan.

	WP1: GRAPEDIA database buildout.	WP2: Development of Data integration models
Specific	<p>The CIG team (all which are currently producing and handling huge amounts of omics data and many with expertise in database construction) and active participants of the INTEGRAPPE COST Action, will gather grapevine omics data into a single database. This objective will be divided in several subtasks:</p> <ul style="list-style-type: none"> - To build the structure of the database after comparing what has been achieved for other species. - To identify where it should be hosted on a long-term basis, also considering versions/releases. - To implement priority types of data / modules in the database in the time of the project. - To develop an automated way to feed data in this database. - To prepare for the possibility of adding new types of data / modules. - Interconnection to existing resources and catalogues such as 'grapegenomics.com', 'VIVC' or <i>up-to-now</i> unrelated DBs. 	<p>The members of the CIG team with expertise in data integration and modelization and active participants of the INTEGRAPPE COST Action, will use data gathered in GRAPEDIA (also related to phenotyping, metabolomics and climate) to develop models to predict which cultivars or new genotypes will be best adapted to climate change in a given location. This objective will be divided in several subtasks:</p> <ul style="list-style-type: none"> - To clarify the needs of the viticulture industry (which countries, which type of grapes, for how long.). - To select relevant types of data to integrate in the models. - To build models. - To test and validate them with climate data from the past.

Measurable	The deliverable will be to provide the community with a working and accessible GRAPEDIA database with at least some functional priority modules like genomics, pangenomes, QTL and molecular markers, and germplasm collection databases.	The deliverable will be a model that could be the basis of a customized service for the viticulture industries all over the world. In close participation with the relevant companies a business plan will be developed for some degree of commercial exploitation of GRAPEDIA.
Achievable	The objective will be achieved thanks to the expertise in the field of highly skilled scientists in the CIG Team and of the Advisory Board. Some of them have already built similar databases for grape and other species. Also, their respective national research institutions are provided with computing infrastructures to develop an open access database.	The objective will be achieved thanks to the expertise in the field of highly skilled scientists in the CIG Team. Moreover, having a global database like GRAPEDIA providing all the data needed, following the F.A.I.R. principles and in a format ready for integration, represents a unique opportunity to achieve this objective.
Relevant	The objective meets the expectations of the grapevine community after the INTEGRAPPE COST Action because all work done so far paves the way to create this global database with all the guidelines, reference datasets and the list of repositories produced, all following F.A.I.R. principles.	This objective meets the needs of the grapevine community and viticulture industry because it will provide an innovative tool to anticipate the effects of climate transition on grapevine, a perennial crop whose production needs to be quantitatively and qualitatively suitable for decades in a sustainable manner.

*The Gantt Chart considered for the implementation of these objectives can be found in Section 3.

1.4 WHAT IS INNOVATIVE ABOUT YOUR SOLUTION

The innovation here is that **GRAPEDIA will be the largest network hub resource with directed services for the community**, implementing innovative technologies such as deep learning and modelization methods to provide a service to the companies in order for them to adapt their production whilst being aware of the impact of climate transition. Also seen as an innovative aspect of the proposal the CIG will count on the direct involvement of stakeholders (several world-renowned nursery and wine companies) and a bioinformatic service company (Sequentia Biotech) that are ongoing participants of the INTEGRAPPE COST Action and that will benefit and contribute to the accomplishment of GRAPEDIA.

2. EXPECTED IMPACT

2.1 DESCRIPTION OF BENEFITS (WHO, HOW, WHEN AND TO WHICH EXTENT).

The **end-users** who will benefit from the CIG deliverables will be the entire grapevine community along with communities working on other fruit tree species using grapevine as a model plant. In addition, grapevine breeders will benefit from the GRAPEDIA database itself to implement omics data in their breeding programs. Nurseries, breeding and viticulture/wine industries will also benefit from the service regarding modelization of the effects of climate change on cultivars. No competition to GRAPEDIA is expected as the intention is to obtain the interest of the entire community. The INTEGRAPPE COST Action has already been largely open to the actors worldwide and there has been sound coordination with overseas partners; the same is expected here. Regarding overlapping, the American VITIGEN project shares some of the same goals but there has been constant interaction sharing tools, and some of the participants are indeed part of the CIG Team.

The end-users will benefit from the CIG deliverables because the data won't be spread in several public or laboratory websites / personal databases, some of them not being maintained over time. By these means, datasets will be fully interoperable and users won't have to deal with different data formats and metadata that are not properly handled regarding grapevine plant and genome characteristics. Therefore, end-users **will connect online to a single, centralized and Open Access database.**

The end-users will benefit from the tools and resources of the GRAPEDIA database as soon as a first version will be released, i.e., **within the first 6 months of the CIG project.** The release will be advertised as much as possible to the community through conferences, social media, and publication. At the end of the CIG period, it is expected that the Work Package 2 will deliver a first proof-of concept for a modelization of data.

The GRAPEDIA database will make grapevine data exploration **much easier and faster.** It will guarantee an access to **up-to-date information** of genome assemblies and their annotations in all versions, for the reference and other sequenced cultivars, of gene catalogues, QTLs and metabolomic/phenotyping data of grapevine cultivar accessions.

Having all the data accessible and centralized in one dedicated database will greatly **facilitate meta-analyses and data integration** as proposed in the CIG proposal to provide the **first decision support system** for grapevine breeders and viticulture industries allowing anticipation of the effects of climate change on their cultivars. The decision support system could provide a comprehensive metric to allow growers to assess the actual level of sustainability of their vineyard as well as solutions to reach a maximum level based on historical data on their *terroir*. Specifically, they could have support to decide if they opt for more adapted cultivars would provide better economic and environmental outcomes on a long-term basis.

2.2 PLAN FOR EXPLOITATION

User's take-up and commercial exploitation: GRAPEDIA will be an Open Access database but will also offer customized services to research groups and the industry through internal and externalized services, in alliance with bioinformatic companies who are already active partners of the INTEGRAPPE COST Action (e.g., Sequentia Biotech). This model would perfectly fit a business plan which the CIG intends to adjust and execute during the CIG grant period.

Investment Funding: In addition to the capacities given by the CIG Team members own research groups, the CIG expects to be able to find public funding (EU-based and non-EU based) and private investment opportunities. In tune with this, Mercier Nurseries (France) and Sequentia Biotech (Spain) are part of the CIG Team and will not only contribute towards decision making but may also contribute with funds, infrastructure and human resources for the development of the database, including the modelization stage. Concha & Toro Winery (Chile) has already stated their interest to participate and provide funding when needed. Gallo Winery (USA) has been contacted through one of its consultants, which is part of the CIG Team, and we expect they will also join the initiative.

3. PLAN FOR IMPLEMENTING THE CIG

The implementation has three main axes, which are **(1) to discuss and decide**, through Working Group Meetings, on the best database structure based on the community needs and the resources already available in and out of INTEGRAPPE (administrative decisions will also be taken here), **(2) to implement** it through the tools available in COST for the execution of the CIG proposal, including Training Schools (conducted in the form of jamborees and hackathons), and Short-Term Scientific Missions, where students and researchers will be nurtured by researchers and laboratories that have expertise in biological database construction, data integration and web site platforms, and **(3) disseminate** the database through Dissemination tools and a Final Annual meeting, to increase visibility, awareness, collaborations and cross-cultural exchanges within the CIG actors and the world-wide community at public level.

The CIG will be coordinated by an active generation of scientists who will make the most of the entire portfolio of new technologies available. Their expertise is shown in Part A of this proposal, being in all cases related to at least one of the following topics: *omics or phenotyping data generation/analysis, data integration, database construction, genome sequencing and assembly, and gene annotations*. In addition, the CIG will count on an Advisory Board with experience in management, Intellectual property rights (IPR) and expertise on databases and tools offered to the community, and will also include the Chair/ViceChair of the CA17111 INTEGRAPe Action. The Board will be crucial in decision making as their opinions and suggestions will be considered in the Working Group Meetings. An important part of the CIG implementation will be promoted by the activities done with the Stakeholders. Up to this moment, the CIG counts on support letters from Concha & Toro Winery, Sequentia Biotech and Mercier Nurseries.

Following the objectives scheme (Section 1.3), the CIG will be implemented in two Work Packages (WP, Figure 1): WP1) Decision making and build-up of the GRAPEDIA database and WP2) Integration of omics and phenotyping/climate data for further modelization. Different CIG Team members will coordinate these working packages and but will make decisions in the Working Group Meetings for all topics of consideration.

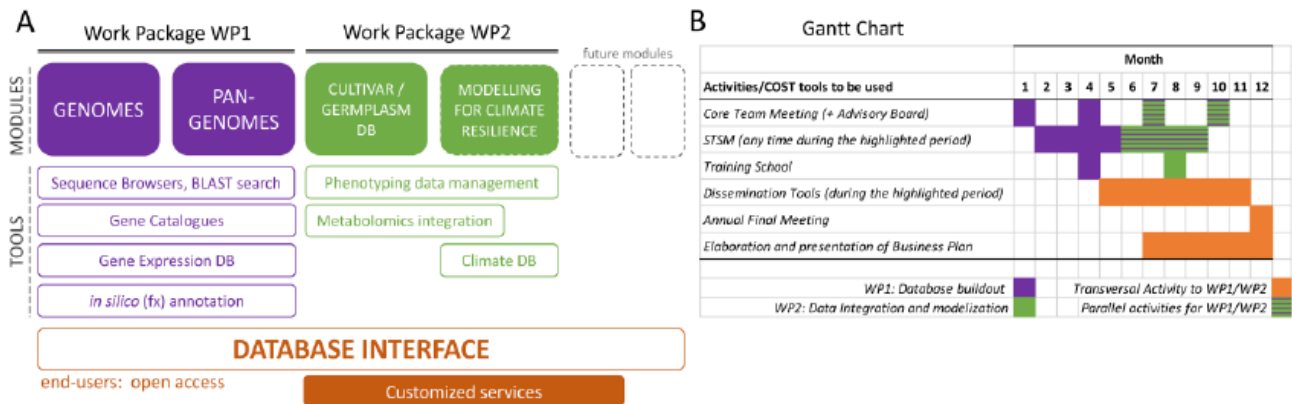


Figure 1. General scheme for the CIG Implementation of GRAPEDIA (A) and time schedule (B).