

Brussels, 13 April 2018

COST 022/18

DECISION

Subject: **Memorandum of Understanding for the implementation of the COST Action “Data integration to maximise the power of omics for grapevine improvement” (INTEGRAPE) CA17111**

The COST Member Countries and/or the COST Cooperating State will find attached the Memorandum of Understanding for the COST Action Data integration to maximise the power of omics for grapevine improvement approved by the Committee of Senior Officials through written procedure on 13 April 2018.



MEMORANDUM OF UNDERSTANDING

For the implementation of a COST Action designated as

COST Action CA17111
DATA INTEGRATION TO MAXIMISE THE POWER OF OMICS FOR GRAPEVINE IMPROVEMENT
(INTEGRAPE)

The COST Member Countries and/or the COST Cooperating State, accepting the present Memorandum of Understanding (MoU) wish to undertake joint activities of mutual interest and declare their common intention to participate in the COST Action (the Action), referred to above and described in the Technical Annex of this MoU.

The Action will be carried out in accordance with the set of COST Implementation Rules approved by the Committee of Senior Officials (CSO), or any new document amending or replacing them:

- a. "Rules for Participation in and Implementation of COST Activities" (COST 132/14 REV2);
- b. "COST Action Proposal Submission, Evaluation, Selection and Approval" (COST 133/14 REV);
- c. "COST Action Management, Monitoring and Final Assessment" (COST 134/14 REV2);
- d. "COST International Cooperation and Specific Organisations Participation" (COST 135/14 REV).

The main aim and objective of the Action is to establish an open, international, and representative network that integrates data from existing resources in a cost-effective manner, as well as making interoperable grapevine dataset and tools available in a secure and standardized format. This will be achieved through the specific objectives detailed in the Technical Annex.

The economic dimension of the activities carried out under the Action has been estimated, on the basis of information available during the planning of the Action, at EUR 60 million in 2017.

The MoU will enter into force once at least seven (7) COST Member Countries and/or COST Cooperating State have accepted it, and the corresponding Management Committee Members have been appointed, as described in the CSO Decision COST 134/14 REV2.

The COST Action will start from the date of the first Management Committee meeting and shall be implemented for a period of four (4) years, unless an extension is approved by the CSO following the procedure described in the CSO Decision COST 134/14 REV2.

OVERVIEW

Summary

The COST Action INTEGRAPE will bring together all stakeholders in the grapevine research community (academic, industry, policymakers and consumers) in an open, international, and representative network to develop minimal data standards and good practices in order to integrate data repositories and improve interoperability between datasets. The ultimate objective is to harness and exploit all available data to achieve better management practices and more cost-effective breeding for improved genotypes. Grapevine is grown worldwide to produce fresh berries, processed fruits and wine. The major challenge is to control berry composition and maintain yields while limiting the use of pesticides, water and other inputs, thus adapting the industry to climate change while achieving environmental and economic sustainability. Grapevine research focuses on interactions between the genotype, phenotype and environment, and information must be integrated from heterogeneous datasets including ampelography, environmental biology, genetics, genomics, epigenomics, transcriptomics, proteomics and metabolomics. The data are currently dispersed and difficult to access, hindering meta-analysis (the re-use of grapevine data beyond the original experiments). No institution working in the field of grapevine research has yet taken on the mission to improve data integration and interoperability at the global level, although the grapevine research community is continuously producing large datasets. The concepts described will support stakeholders by developing innovative strategies to integrate grapevine data from existing resources and new experiments in a cost-effective manner, as well as making interoperable grapevine datasets and tools available in a secure and standardised format.

<p>Areas of Expertise Relevant for the Action</p> <ul style="list-style-type: none"> ● Agriculture, Forestry, and Fisheries: Databases, data mining, data curation, computational modelling ● Biological sciences: Systems biology ● Biological sciences: Genomics, comparative genomics, functional genomics ● Biological sciences: Transcriptomics ● Biological sciences: Metabolomics 	<p>Keywords</p> <ul style="list-style-type: none"> ● PHENOTYPE ● GENOMICS ● METABOLOMICS ● SYSTEMS BIOLOGY ● GRAPE
--	--

Specific Objectives

To achieve the main objective described in this MoU, the following specific objectives shall be accomplished:

Research Coordination

- Agree, with input from scientific and industry stakeholder communities, on minimal information and metadata standards (biological material, experimental methods and data provenance) for different data types (e.g. plant material, genotypes, phenotypes, genetic maps, quantitative trait loci (QTLs), genome sequences/annotations, omics datasets and environmental parameters).
- Adopt a community-based curation approach in which working groups will set and apply the standards to integrate new datasets into existing archives.
- Disseminate standards and good practice using classical routes (publications, web pages, conferences) but also hands-on sessions that will (i) provide training, (ii) provide standardised datasets/repositories, and (iii) identify potential improvements in standard specifications.
- Improve the interoperability of existing international and regional infrastructures that manage grapevine data to promote data integration and re-use, including standard web services (RESTful APIs) based on standardised underlying data models. These are being developed by the international initiative Breeding API and could be tested by the grapevine community.
- Develop a common strategy for the publication and sharing of less-structured data that currently lack centralised repositories. This will provide access to broad datasets and reference data from diverse sources with a flexibility that promotes the rapid integration of data from novel and emerging technologies.
- Propose best practices for data analysis to improve reproducibility, transparency and accessibility by

optimising data flow and interoperability among datasets/repositories, different analytical tools and existing data repositories. □

- Promote, with additional input from the grape and wine industries, the adoption open data in grapevine research and its use to develop new tools and strategies to improve the profitability and sustainability of the grapevine industry. □
- Support advanced data modelling and richer semantic integration, providing new biological insights and facilitating their application, ultimately leading to new and better-adapted cultivars through better-targeted, marker-assisted breeding.

Capacity Building

- Engage young scientists and train them to use emerging standards as much as possible in their research. It will also increase awareness, among the public and private grapevine research communities, of data standards that will improve the re-use of data.
- Help the entire community to adopt good practices and a common global data management strategy that will spread through future regional, national and international projects (e.g. EU projects must describe a precise data management plan to promote open data in science) thus improving the transparency and reproducibility of published data.
- Implement and improve data standards endorsed by large international consortia or initiatives. It will also provide better links to current EU infrastructures for data analysis and management such as ELIXIR. The grapevine community will thus improve the FAIR-ness of its data management practices within a distributed information system.
- Address the challenges of omics data integration with other heterogeneous datasets in a transnational scientific community by establishing an information platform based on federated systems distributed initially around Europe and ultimately around the world.
- Promote the adoption of international standards and formats that will facilitate better data sharing not only within the grapevine research community but also in the broader context of building open knowledge resources.
- Enhance the value of existing data and increase the efficiency of future research. By including industry stakeholders, it will also build industrial capacity to engage with technologies of the future.

1. S&T EXCELLENCE

1.1. CHALLENGE

1.1.1. DESCRIPTION OF THE CHALLENGE (MAIN AIM)

Grapevine (*Vitis vinifera* L.) is grown worldwide and is one of the most important and valuable horticultural crops. The berries are used as fresh, dried or canned fruit, and for juice or jam, but their main use is wine production. The major challenge for viticulture and oenology is to control berry composition and maintain yields while limiting the use of pesticides, water and other inputs, thus adapting to climate change and achieving environmental and economic sustainability. It is therefore necessary to understand and precisely modulate the complex mechanisms controlling adaptive, yield, quality and sensory traits, based on research to determine relationships between genotype, phenotype and environment (including management). To address these challenges, the grapevine research community must generate and integrate heterogeneous datasets (genomics, epigenomics, transcriptomics, proteomics, metabolomics and ampelography) describing genotypes, phenotypes and the environment. Integrating these multi-omic datasets will reduce the gap between data generation and the ability to analyse and understand the biological mechanisms underlying grapevine responses to the environment. This is achieved by applying high-throughput experimental techniques that generate large datasets (“omics” technologies and multiplex sampling/imaging). The resulting data are currently dispersed and difficult to access, hindering exploitation beyond their initial purpose. International gene data repositories do not store functional data (e.g. regulatory and metabolic networks), detailed plant materials or non-molecular phenotypes. Such data may be stored in regional or local databases but are often inaccessible to the wider research community.

The value of data from individual experiments is much enhanced when considered in a wider context through meta-analysis. In the Arabidopsis community, rich datasets supported by the TAIR (The Arabidopsis Information Resources) and Araport11 portals are used to develop broader hypotheses (<http://www.arabidopsis.org> and <https://www.araport.org/data/araport11>). The meta-analysis of these integrated datasets helps to identify mechanisms underlying the interactions between plants, their environment and plant management techniques. However, the interpretation and re-processing of data requires additional metadata to provide appropriate context. Ideally, data should be formatted in a standardized manner for automated processing to avoid errors caused by manual manipulation, especially in very large datasets (Stephens et al. 2015; Wilkinson et al. 2016). Such automated data re-processing is supported by recent advances in informatics that combine different computational environments (e.g. cloud computing). International consortia are also promoting the use of FAIR principles that ensure data are findable, accessible, interoperable and reusable (Wilkinson et al. 2016). Model system ontologies cannot always be applied directly to grapevine because its unique traits differ from model organisms, which is why current standardisation efforts have been unsuccessful when applied to this species, so the grapevine community must adapt current conventions to build a standardised system for data collection, processing, storage, access and analysis, and develop strategies to integrate them for a full description of the phenotypes.

1.1.2. RELEVANCE AND TIMELINESS

Although the grapevine research community continuously produces large datasets, no institution working in this field has yet taken on the mission to ensure data integration and accessibility at the global level. The concepts in this COST Action will support researchers developing interdisciplinary approaches in grapevine biology. The main challenge is to establish an open, international, and representative network that integrates grapevine data from existing resources in a cost-effective manner, as well as making interoperable grapevine datasets and tools available in a secure and standardised format. Many international initiatives are developing open data platforms, and crop communities should now implement such recommendations to develop interoperable datasets and data management infrastructures that support grapevine research and plant biology. Understanding how the genome, environment and viticulture practices interact to affect fruit quality will allow us to implement agricultural practices that achieve desirable fruit characteristics for every climate/cultivar combination (Jones and Davis, 2000; Fabres et al., 2017) and to develop better adapted cultivars. This will lead to more efficient resource utilisation, better vineyard management, increased profitability and sustainability. In addition, growers can maximise environmental effects on grapevine to highlight the uniqueness of their vineyards, increasing their commercial competitiveness. The development of new planting material and new viticultural techniques take considerable time and effort, but rapid climate change is placing intense pressure on current cultivars and production systems. Powerful omics techniques must therefore be exploited to the fullest extent so that growers have access to new cultivars and practices that are adapted to the changing environment.

1.2. OBJECTIVES

1.2.1. RESEARCH COORDINATION OBJECTIVES

INTEGRAPE will gather grapevine data from producers, managers and users (including academics, breeders and industry representatives) in order to achieve the following objectives:

- Agree, with input from scientific and industry stakeholder communities, on minimal information and metadata standards (biological material, experimental methods and data provenance) for different data types (e.g. plant material, genotypes, phenotypes, genetic maps, quantitative trait loci (QTLs), genome sequences/annotations, omics datasets and environmental parameters). Many standards already exist but better specification is necessary for the community to avoid overlapping interpretations. INTEGRAPE will adopt a community-based curation approach in which Working Groups will set and apply the standards to integrate new datasets into existing archives.
- Disseminate standards and good practice using classical routes (publications, web pages, conferences) but also hands-on sessions that will (i) provide training, (ii) provide standardized datasets/repositories, and (iii) identify potential improvements in standard specifications.
- Improve the interoperability of existing international and regional infrastructures that manage grapevine data to promote data integration and re-use, including standard web services (RESTful APIs) based on standardised underlying data models. These are being developed by the international initiative Breeding API and could be tested by the grapevine community. A common strategy could also be developed for the publication and sharing of less-structured data that currently lack centralised repositories. This will provide access to broad datasets and reference data from diverse sources with a flexibility that promotes the rapid integration of data from novel and emerging technologies. The supervision of data submission by consortium experts will maintain standards, similar to the WheatIS data repository interface.
- Propose best practices for data analysis to improve reproducibility, transparency and accessibility by optimising data flow and interoperability among datasets/repositories, different analytical tools and existing data repositories.
- Promote, with additional input from the grape and wine industries, the adoption of open data in grapevine research and its use to develop new tools and strategies to improve the profitability and sustainability of the grapevine industry.

1.2.2. CAPACITY-BUILDING OBJECTIVES

INTEGRAPE will engage young scientists and train them to use emerging standards as much as possible in their research. It will also increase awareness, among the public and private grapevine research communities, of data standards that will improve the re-use of data. The COST Action will help the entire community to adopt good practices and a common global data management strategy that will spread through future regional, national and international projects (e.g. EU projects must describe a precise data management plan to promote open data in science) thus improving the transparency and reproducibility of published data. INTEGRAPE will implement and improve data standards endorsed by large international consortia or initiatives (e.g. Research Data Alliance, Breeding API). It will also provide better links to current EU infrastructures for data analysis and management such as ELIXIR. The grapevine community will thus improve the FAIR-ness of its data management practices within a distributed information system. INTEGRAPE will address the challenges of omics data integration with other heterogeneous datasets in a transnational scientific community by establishing an information platform based on federated systems distributed initially around Europe and ultimately around the world. This will be achieved by promoting the adoption of international standards and formats that will facilitate better data sharing not only within the grapevine research community but also in the broader context of building open knowledge resources. Such standards will support advanced data modelling and richer semantic integration, providing new biological insights and facilitating their application, ultimately leading to new and better-adapted cultivars through better-targeted, marker-assisted breeding. These initiatives will greatly enhance the value of existing data and increase the efficiency of future research. By including industry stakeholders, it will also build industrial capacity to engage with technologies of the future.

1.3. PROGRESS BEYOND THE STATE-OF-THE-ART AND INNOVATION POTENTIAL

1.3.1. DESCRIPTION OF THE STATE-OF-THE-ART

Grapevine is a perennial woody crop, and many of its non-molecular traits can only be evaluated in the field. The grapevine phenotype is determined by complex interactions involving the genotype (G) of the scion and rootstock, multifactorial biotic and abiotic environmental factors (E), and viticulture management strategies (M). Field experiments are complemented by experiments under controlled conditions to study these factors in isolation. Increasingly, this dedicated corpus of knowledge requires the integration of traditional datasets (varietal classifications, phenotypes, environments, viticulture practices) with large datasets derived from high-throughput experiments (genomics, transcriptomics, proteomics, metabolomics). The community already benefits from a large number of such datasets, including expressed sequenced tags (ESTs), simple sequence repeats (SSRs), single nucleotide polymorphisms (SNPs), QTL maps, the grapevine genome sequence, transcriptome and proteome databases, a better understanding of genetic diversity via genotyping/re-sequencing studies or the de novo sequencing of new genotypes, and the detailed metabolomic analysis of different genotypes grown in different environments (Adam-Blondon et al. 2016). The value of this rich data collection is limited by the lack of standardisation, which reduces the interoperability between datasets and restricts them to the context of the original experiments. Much more information could be extracted by meta-analysis and the integration of data from diverse sources.

The interoperability of grapevine phenotyping data is a new challenge raised by the international scientific community (Ćwiek-Kupczyńska et al. 2015). Ampelography is the identification and classification of grapevine species and cultivated varieties based on phenotype, i.e. morphological characteristics, and it is used not only for the selection of vines for wine production, but also for rootstock selection, clonal selection, table-grape production, and other agricultural and industrial applications. Grapevine has been described at the morphological level in more detail than most other species. However, there are many heterogeneous lists of distinctive characteristics, and these have been harmonised by the International Organisation of the Vine and Wine (OIV), the International Union for the Protection of New Varieties of Plants (UPOV), and Bioversity, formerly known as the International Plant Genetic Resources Institute (IPGRI). These efforts to develop shared catalogues of traits and evaluation procedures are being converted into ontologies such as the *Vitis* Trait Ontology (<http://www.croponontology.org/ontology/VITIS/Vitis>) under the Crop Ontology framework (<http://www.croponontology.org/>). Accurate definitions of characteristics necessary for the identification of grapevine species and varieties have been drawn up (OIV, 1997; 2009) resulting in a definitive list of

grapevine varieties and their synonyms (<http://www.vivc.de>). The challenge for the grapevine community is to improve this ontology for new traits or new phenotyping methods (e.g. high-throughput phenotyping methods and remote sensing) and to ensure that all the data are made available, e.g. using the current list of grapevine genetic resources (<http://www.eu-vitis.de>). Recent advances in DNA sequencing (next-generation sequencing, NGS) have revolutionised biology by increasing scalability and throughput, thus accelerating genome projects but also facilitating whole-genome re-sequencing for the analysis of sequence variations at the population level. RNA sequencing (RNA-Seq) is useful for transcriptome and non-coding RNAome analysis, and ChIP-Seq analysis for DNA–protein interactions. Furthermore, whole-genome bisulfite sequencing allows the quantitative detection of epigenomic variations based on DNA methylation. Other approaches have been developed to study proteins and small molecules, including interactome analysis for networks of protein–protein interactions, hormone analysis for phytohormone-mediated cellular signalling, and metabolome analysis for the analysis of dynamic metabolic pathways and networks. These rapidly developing fields use genome-scale resources to model each molecular network as an integrated component of the larger cellular system. Bioinformatics is crucial to every aspect of omics research by processing genome-scale datasets and extracting valuable knowledge.

A decade ago, the publication of the grapevine genome sequence allowed the application of other omics approaches in this species (Jaillon et al. 2007). Multiple transcriptomic, proteomic and metabolomic platforms have been developed to investigate how the information encoded in the grapevine genome is converted into the phenotype, with each step generating massive datasets from series of experiments performed in the open field. NGS has driven down the cost of sequencing and increased its throughput and accuracy, which means it will soon be possible to sequence most of the cultivated elite varieties and wild grapevine species worldwide. Grapevine physiology has also advanced such that the interplay among enzyme functions, structural components, regulatory networks, and flux through metabolic pathways can be investigated in detail, and linked to epigenetic regulatory networks acting at the transcriptional, post-transcriptional, translational and post-translational levels. Metabolomics is particularly relevant in this context because the sharing of metabolomics data would facilitate the development of new varieties with enhanced quality traits important for consumers. The large international repositories such as EMBL, NCBI and SwissProt have accommodated these developments and most of the primary data types (sequences, transcriptomes, proteomes, metabolomes) benefit from internationally agreed formats and standards. However, the metadata are often poor, particularly the precise identification of genetic material, and it is important to overcome this issue. Some datasets corresponding to more integrated or curated data are not managed by these repositories, including QTL maps, association data, and metabolic networks. These data are sometimes held in regional databases (e.g. VitisCyc for metabolic pathways) but often remain inaccessible to the wider community.

1.3.2. PROGRESS BEYOND THE STATE-OF-THE-ART

INTEGRAPE will integrate existing data repositories by promoting data standardisation and interoperability among datasets, with the ultimate objective of harnessing the available data to facilitate biological improvements (cost-effective breeding for improved genotypes, matching genotypes to environments and viticulture management). The COST Action will progress beyond the state of the art by developing a framework to harness new resources made available by large-scale omics experiments, and will use them to provide the first truly integrated platform for the sharing, analysis and interpretation of grapevine data. Repositories that comply with the newly developed standards will be used to bring other repositories into line. The integration of heterogeneous datasets also requires procedures that generate inter-relationships between data and consolidate the data structure for each feature. This COST Action will progress beyond the state of the art by improving the interoperability of existing infrastructures, allowing the heuristic mining of genome-scale datasets representing different omics experiments, as well as morphology, physiology and genetics. The most ambitious step beyond the state of the art will involve the transition to multiscale phenotyping, i.e. the integration of omics data and existing phenotypic data resources as well as new value-added data in an appropriate metadata format to ensure that biological functions can be inferred from complex and diverse datasets. The power of the new omics technologies can in this way be focused on the development of solutions to challenges in the grapevine industry posed by changes in the agronomic, commercial and social environments.

1.3.3. INNOVATION IN TACKLING THE CHALLENGE

Community-based distributed information systems such as TransPLANT and WheatIS are difficult to establish due to social (durably in the community), scientific (future-proof data interoperability) and technical (construction of a functional distributed infrastructure for big data sharing/analysis) challenges. INTEGRAPE will build the foundations of a standardised, open data sharing system by linking its standardised infrastructure to data provided by the scientific community. This is an innovative approach to tackle the challenge of diverse, heterogeneous and dispersed large-scale grapevine datasets by developing and establishing new ways to process existing datasets and integrate them with the large-scale data from current and future omics studies. Sophisticated tools and algorithms will be developed to convert existing heterogeneous datasets into standardised and machine-readable data, including the integration of traditional ampelography data with omics datasets. This innovation will be expressed as a modern phenotyping approach for grapevine cultivars and wild species in the open field, integrating historical, pre-genomic and current largescale omics and systems biology resources.

1.4. ADDED VALUE OF NETWORKING

1.4.1. IN RELATION TO THE CHALLENGE

Recent advances in large-scale biology have expanded the horizons of grapevine research and made international collaborations essential for progress in this field. Although these combined approaches provide new strategies to investigate the influence of genotype, environment and management on grapevine quality, it is beyond the expertise and capacity of a single research group to integrate the large amounts of data from such studies. The partners in this COST Action have realized that the transformation of high-throughput data into knowledge requires close cooperation and coordination among groups with different forms of expertise, experience in different techniques, and local insights. Therefore, the added value of this network is its ability to build a better community-wide foundation, based on consensus, to harness existing knowledge and cutting-edge technologies. This can be used to link genetic and ampelographic data with gene expression, protein expression and metabolomic datasets so that the data can be exploited more effectively for the improvement of grapevine crops in Europe and beyond. The COST Action will establish a dedicated network linking these research groups, ensuring that data are shared among partners throughout Europe, as well as linking the European research community to the most prominent international groups in this field. The COST Action will promote discussion, strengthen and enlarge collaborations among different research organisations and promote the development and distribution of new tools for data sharing and analysis. It will also network with industry to enhance the dialogue between researchers and potential users of improved grapevine products.

1.4.2. IN RELATION TO EXISTING EFFORTS AT EUROPEAN AND/OR INTERNATIONAL LEVEL

Data integration is now recognised as an essential part of the life sciences and agriculture. The EU funded FP7 project STATegra (<http://stategra.eu/>) and the COST Action SeqAhead (http://www.cost.eu/COST_Actions/bmbs/BM1006) addressed the need for bioinformatics in the era of high-throughput data and hosted the “Workshop of Omics and Data Integration” to review current technologies for the production, integration and analysis of omics data, highlighting the need for tools to facilitate the integration of heterogeneous datasets. European ESFRI platforms such as the European Plant Phenotyping Infrastructure (EMPHASIS) and the European Infrastructure of Bioinformatics for Life Science (ELIXIR) are developing and disseminating standards for plant omics and plant phenotyping experiments (including laboratory, greenhouse and field-based phenotyping). The Research Data Alliance, an international organisation that promotes technological and social mechanisms to enable data sharing, has focused on crop data. Ontologies that can be applied as data standards are being developed by the Crop Ontology and Planteome projects (<http://www.cropontology.org/>). The DivSeek project is coordinating resequencing and informatics for the investigation of biological diversity. The Wheat Information System applies these approaches to wheat and is developing technologies and practices that could be transferred to fruit crops. Other platforms such as GnpIS and Gramene (crops and model plants) are also providing strategies for data integration. The COST Action will provide added value in relation to these existing efforts by establishing a common access point for grapevine data,

including genetic and physical genome maps, QTLs/associations, omics data, and integrated data in the context of systems biology. The COST Action will also develop tools to link associated field data (e.g. climatic, agricultural and viticulture data) using data standards, ontologies, common application programming interfaces, tools and training to maximise data interoperability. Grapevine data will be integrated with other relevant datasets, such as the metagenomics of microorganisms interacting with grapevine plants and those used for winemaking. The COST Action adds value by fitting precisely into the initiatives developed by existing projects, and will exploit available infrastructure by establishing a network to manage grapevine ampelographic, genetic, omics and phenotypic data in collaboration with the worldwide activities and standards developed by the OIV.

2. IMPACT

2.1. EXPECTED IMPACT

2.1.1. SHORT-TERM AND LONG-TERM SCIENTIFIC, TECHNOLOGICAL, AND/OR SOCIOECONOMIC IMPACTS

The partners in the COST Action have contributed significantly to the development of the latest generation of high-throughput technologies and biological resources for many fruit species including grapevine, and the COST Action addresses an urgent need to promote access to these and future assets. One of the greatest short-term impacts of the network will be to make integrated datasets and associated tools available to more researchers working in grapevine biology, viticulture and oenology. A lasting legacy of the COST Action will be its longer-term impact on the research community by providing training for young scientists in the use of new methods for data integration and analysis, thus enabling them to develop more sophisticated grapevine research strategies in the future. The COST Action will bring together scientists qualified in viticulture, plant physiology, omics and bioinformatics, and will have a lasting impact by bridging the genotype – phenotype gap and developing multidisciplinary approaches to understand the G x E x M interactions underlying the quality of grapevine berries. On a broader level, the main impacts of the COST Action will include the restructuring of European research in grapevine omics and phenotyping by bringing together different European research groups and building networks with the most prominent international groups. The COST Action will facilitate knowledge transfer between researchers working on different aspects of grapevine phenotyping by combining resources and skills to integrate genomics, epigenomics, proteomics, metabolomics, reverse genetics, bioinformatics, ampelography, viticulture, physiology, ecophysiology and oenology, so grapevine phenotypes can be studied using a holistic approach despite the diverse nature of the techniques, resources, materials and data types. Another key impact will be the provision of access to research infrastructure and facilities for the partners via short-term scientific missions (STSMs) and the diffusion of advanced technologies among different groups. Technological impact will also be achieved by developing new tools to improve data sharing and interoperability, and organizing training schools for the sharing of knowhow and advanced research technologies. This will enhance communication between research groups to avoid the duplication of work in different European countries, accelerate decision making in terms of research priorities, and establish a road map showing how the multidisciplinary approach in this COST Action can be applied to other crops. At the socioeconomic level, the COST Action will also enhance the competitiveness of Europe in basic science, biotechnology and agriculture by providing young scientists with access to the latest resources in the grapevine research field. It will provide advanced graduate and postdoctoral training in state-of-the-art high-throughput genomic, post-genomic, phenomic, and data integration methods, and it will establish tools and standards for grapevine phenotyping that can be applied worldwide. This will improve breeding strategies focusing on the quality, hardiness and adaptability of grapevine cultivars, with a knock-on impact on growers and wine-makers, increasing the prospects for employment in these areas and increasing profitability in the sector. Local knowledge can be spread through the network and applied in new settings, addressing issues caused by climate change and the spread of pests and diseases. This will encourage conversations between scientists and industry stakeholders. The long generation time of grapevine, the high cost of breeding and the exacting requirements for berry composition determined by the sophisticated wine markets mean that the holistic approach discussed above is necessary for the development of new, improved cultivars and management practices in a changing environment.

2.2. MEASURES TO MAXIMISE IMPACT

2.2.1. PLAN FOR INVOLVING THE MOST RELEVANT STAKEHOLDERS

Basic and applied research scientists studying grapevine biology and phenotyping. The structure of the grapevine biology research landscape will be moulded by the COST Action because it will prevent the duplication of initiatives or experiments in different laboratories and countries (wasting time and resources) or will facilitate the integration of such initiatives to achieve added value (e.g. similar approaches applied to different cultivars, or under different environmental conditions). Basic and applied research scientists will therefore be involved in all aspects of the COST Action, including opportunities to attend consortium meetings, give oral presentations, present posters and discuss their work and ideas with a wider range of stakeholders including industry representatives. They can also participate in hands on-training, webinars and STSMs.

National government and European-level policymakers. Documents and guidelines prepared during this COST Action will be valuable to policymakers at the national and European levels, and members of the INTEGRAPPE consortium will be available to discuss the resulting issues with policymakers. The guidelines will be used for decision making in terms of research funding in different member states, and to plan or direct breeding programs in the European viticulture community.

Industry stakeholders: Breeders, growers, winemakers and the food-processing industry. The resources and knowledge generated during this COST Action will facilitate the development of innovative breeding strategies and will improve current technologies. INTEGRAPPE will also facilitate the development of new varieties better adapted to the needs of growers (e.g. disease resistance, drought/frost tolerance) and also the aggregation of knowledge on the behaviour of known varieties to guide selection and management. The resources and knowledge from the COST Action will promote research to improve the sensory and nutritional properties of fresh and dried fruits, juices and wines, which will benefit producers and consumers. To achieve these aims, key industry representatives will be invited to participate as an **Advisory Council** representing the industry perspective. The Advisory Council will also include consumer representatives to provide input from consumers of grape-derived fruit products and wines.

Research scientists taking part in the COST Action will enable the transfer of new knowledge and technologies to industry stakeholders, who will have open access to relevant information and will be able to provide input at meetings, conferences and training sessions. This approach will also create a larger platform for the dissemination of results generated during the COST Action.

2.2.2. DISSEMINATION AND/OR EXPLOITATION PLAN

Three levels of dissemination activities are envisioned. First, activities within the COST Action will include STSMs, meetings, workshops, mailing lists, and an internal COST Action website linked to or referenced by the International Grapevine Genome Project (IGGP) web site. The purpose of these activities will be to disseminate information rapidly within the INTEGRAPPE consortium and to get feedback on guidelines and recommendations. Second, activities outside the COST Action will include publications, communication via regional organisations, targeted mailing lists, a public website, and workshops at international conferences. The purpose of these activities will be to disseminate the consortium activities to different external audiences, including academic researchers (e.g. presentations/posters at conferences), as well as representatives of the breeders, nurseries, fruit processing and winemaking industries, and viticulturists who are not partners in the COST Action (e.g. presentations/posters at conferences with mixed academic/industry participants). European and national government policymakers and regional planners will also be targeted in this second group, as will the general public, i.e. the consumers of grapevine products interested in knowing how improvements in viticulture using existing and novel varieties will affect their choices. Press releases and articles in wine-related magazines will also be used to disseminate information to the general public. The third level of dissemination concerns the products of the COST Action activities, i.e. the tools, data formats and databases developed during the lifetime of the COST Action. One strategy is to provide links to such resources on the IGGP web site. INTEGRAPPE will also develop a decentralized resource for data and software tools, which will ultimately be searchable through a centralized portal so that participants share the burden of hosting while users can access all resources via a common interface. An inspiring example is ExpASy, a bioinformatics portal operated by the Swiss Institute of

Bioinformatics, which integrates many resources, databases and software tools for omics, phylogeny/evolution, systems biology and population genetics hosted by different groups on different servers, but accessed through a single site including links to external groups/institutions. ExPASy also provides a search function, user statistics, instructions and publications so that both experts and non-experts are accommodated. The tools and resources will be provided as cloudbased systems that can be used over the internet and as local systems that can be downloaded and run independently on non-networked computers, i.e. without uploading data. STSMs targeting young scientists (especially those from developing regions) will facilitate the exchange of ideas and technology. Scientific communication within the COST Action will be facilitated by group meetings, a dedicated website and mailing lists. To ensure interactions with plant breeders and nursery production researchers, at least one of the workshops will be organised as a satellite event for plant breeders and nurseries. Dedicated e-mail networks will be set up at different levels, e.g. to include specific stakeholders/Working Groups/committees or all members of the COST Action. The other mailing list will be for the general public, policy makers and scientists not directly involved in the COST Action but willing to be informed and kept up to date with further developments. This will involve biannual summaries in the form of text and hyperlinks to inform subscribers about the latest developments and direct them to relevant articles or information sources.

A dedicated public website will provide information to the international scientific community and will facilitate communication between the partners. The website will be maintained by a webmaster appointed during the kick-off meeting and will be accessible to a non-specialist audience. This website will initially provide access to the ExPASy-like decentralised resources developed during the COST Action, as well as activities and meetings, publications and contact details for participants. Later it will also include (i) online courses, proceedings of meetings, talks and posters from meetings, and scientific articles published by the participants; (ii) STSM calls and reports; (iii) teaching tools; (iv) links to the websites of participating institutions, websites related to fleshy fruit development and quality traits; and (v) job announcements. To increase visibility, the proceedings of key meetings/conferences will be published in highly-ranked journals, preferably in a special issue dedicated to the information gained within this COST Action. The databases and tools would also be ideal to feature in the special annual “databases” issue of *Nucleic Acids Research*, and the COST Action would pursue this with the editor at the earliest opportunity.

A Technology Transfer Platform will be established as a forum to optimise the diffusion and exploitation of results from the INTEGRAPPE consortium. The activities will be tailored for academic partners that lack sufficient resources to exploit their discoveries. The research published during the final years of this COST Action and the application of guidelines developed by the partners will provide a benchmark for success in the dissemination of the results and outcomes. The representative members of each country will be responsible for disseminating INTEGRAPPE activities to research groups, industrial partners and other stakeholders in their countries. We will therefore use the innovative INTEGRAPPE approach as the basis for a social sciences investigation of the impact of omics on society. The results will be made available for internal consultation, to policymakers, and for publication in journals in fields such as innovation studies and the social sciences. This will facilitate dissemination to technical specialists working in other fields of technology development, and associated end-users. Expertise will also be available on the wider dissemination of findings from the COST Action. The Dissemination Plan will be updated during the COST Action to take recent progress into account.

2.3. POTENTIAL FOR INNOVATION VERSUS RISK LEVEL

2.3.1. POTENTIAL FOR SCIENTIFIC, TECHNOLOGICAL AND/OR SOCIOECONOMIC INNOVATION BREAKTHROUGHS

The activities and results of this COST Action will support innovation through an efficient network of cooperation. The outcomes in terms of data integration, data accessibility and new tools for data analysis will be of interest to researchers, grape producers, the food industry and producers of planting material. The main risks identified concern (i) the COST Action ability to fully engage the community, with measurable results in terms of correctly standardised datasets and RESTful-API infrastructure, and (ii) the possibility of not centralising enough resources to develop and maintain a common portal that hosts and integrates grapevine data (see contingency plan). The wide availability of the COST Action tools and resources will allow the grapevine community and industry partners to make use of the resources for translational research and thus achieve scientific and technological breakthroughs. Therefore, both

these risks can be addressed by maximising the visibility and accessibility of available resources via the COST Action broad multimedia communication strategy. Appropriate licensing structures will be used such as open source coupled with limited IP protection, to ensure that the software developed is adequately protected from commercial exploitation but nevertheless allows for further development that benefits the community.

3. IMPLEMENTATION

3.1. DESCRIPTION OF THE WORK PLAN

The primary objective of this COST Action is to assemble an integrated network of research groups to build the foundations of a bioinformatics platform for the collection and storage of grapevine data in standardised formats, thus facilitating data access, exchange and analysis. Four **Working Groups (WGs)**, described in more detail below, have been identified as priorities to achieve the overall COST Action objectives, and these will report to the management team as part of the internal monitoring procedures. Strong interactions and close collaboration among the four WGs is necessary for the success of this COST Action. Efficient integration among research groups working on different aspects of data production will be ensured by the diverse nature of the WGs, which aim to abolish the borders between scientific communities. Therefore, all WGs will mix researchers and industry representatives spanning the spectrum of knowledge encompassing grapevine genetics, genomics, phenotypic analysis, breeding, industrial applications and end-user demands.

3.1.1. DESCRIPTION OF WORKING GROUPS

WG1. Data interoperability and definition of minimal contextual data standards

One of the main objectives of this COST Action is to improve the capacity of the grapevine community to capture and integrate knowledge from diverse sources and scientific disciplines. Community-wide standards for metadata will be established along with guidelines for their proper use. An important challenge in this context is that data sources tend to lack contextual data, defined as additional, nonexperimental data that provides information necessary for data re-use in a new study. Contextual data may identify key concepts or objects such as samples, genes, traits or markers. It is important to progress towards precise, unambiguous identification and shared vocabularies based on shared dictionaries of synonyms. The cyberinfrastructure developed by the grapevine community should seamlessly integrate heterogeneous data types, such as environmental parameters and phenotypes, with large-scale molecular data from genome sequencing and resequencing projects, epigenomics, transcriptomics, proteomics and metabolomics. The challenge is to ensure the global consistency of the data models and vocabularies across disciplines and datasets.

Task 1.1: Definition of minimal contextual standards

Minimal contextual standards can be defined as any kind of meta-information that needs to be provided alongside the experimental results to ensure the data are useful and can be interpreted by others. Initiatives such as the Genomic Standards Consortium (<http://gensc.org/>) are currently establishing minimal contextual standards for genomic and metagenomic data to facilitate the storage of contextual data in genome catalogues. Several initiatives are also developing Minimal Information About Plant Phenotyping Experiments (MIAPPE) standards (www.miappe.org). Task 1.1 will identify optimal standards for the datasets produced by the grapevine community, contributing to their improvement, filling gaps and providing guidelines to avoid heterogeneity. This work will be guided by real datasets and case studies in collaboration with WG3 and with output from WG4.

Task 1.2: Plant materials and phenotypes

Sample identification (genotypes or varieties) is needed for data re-use in the grapevine community. The COST Action will build on achievements such as the Vitis International Variety Catalogue (www.vivc.de) which assigns a variety number (VIVC code) as an unambiguous tag for each variety and a list of associated synonyms. However, the provenance of samples is often poorly described and rarely includes the VIVC code or even accession numbers from the source germplasm collection. Guidelines for sample identification will combine plant agnostic standards and grapevine-specific resources. The

latest MIAPPE standards and Vitis Trait Ontology (<http://www.croponontology.org/ontology/VITIS/Vitis>) will be challenged with real datasets, improved where necessary, and updated guidelines will be issued. Task 1.2 will be coordinated with other species agnostic international initiatives such as DivSeek.

Task 1.3: Genes, variations, gene expression, proteomics and metabolomics

High-throughput omics data are now widely used to understand the molecular basis of grapevine traits (QTL detection and/or association mapping, genomics, transcriptomics, proteomics and metabolomics). Despite the existence of international repositories for these data, many of the datasets that need to be integrated remain highly dispersed and difficult to access, and rely on the use of non-standardised terms of reference. The grapevine community has published a nomenclature for genes that should facilitate standardisation (Grimplet et al. 2014). Task 1.3 will develop a strategy to generate and update shared dictionaries of synonyms for genes and markers and will deliver a preliminary version in close collaboration with WG3. This will improve the standards and ontologies, in the genetic and genomic domains, and will provide recommendations to the grapevine community for their use based on real datasets. Standards will be developed for the shared annotation of grapevine proteomic and metabolomic data, particularly datasets from different genotypes grown under different environmental conditions and management practices.

Milestones

M1.1. Agreement with the IGGP consortium on a modality to validate work on grapevine data standards carried out by the INTEGRAPE consortium (M12)

Deliverables

D1.1. Guidelines for proper sample identification published on the IGGP web site (M24)

D1.2. Guidelines for data standards in relation to M1.1 published on the IGGP web site (M48)

D1.3. Control vocabularies and synonym lists for markers, genes, proteins and metabolites (M48)

WG2. Interoperability of infrastructures and web services

WG2 will develop a bioinformatics infrastructure to facilitate the discovery, acquisition, storage, processing and integration of diverse grapevine datasets, and this will be managed by distinct nodes. The exposition of data through standard APIs (e.g. Breeding API; <http://docs.brapi.apiary.io>) will be promoted through guidelines and software tools allowing data discovery in distinct repositories (e.g. TransPLANT integrated search tool; <http://www.transplantdb.eu/node/4322>), standard file formats combining data and metadata (e.g. GenBank format for annotations, MIAPPE-compliant ISA-Tab format for phenotypes) and repositories endorsing these recommendations. WG2 will test its recommendations using case studies defined in collaboration with WG3. Finally, WG2 will identify data exchange bottlenecks and will propose mechanisms to overcome them.

Task 2.1: Data file standard formats, repositories and dataset identification

Standard formats exist for many types of data, such as General Feature Format (GFF3) for genomic data, Variant Call Format (VCF) for nucleotide sequence variants, various reporting standards for proteomics experiments (MIAPE) and corresponding data exchange standards (e.g. mzML), nmrML for nuclear magnetic resonance spectra, BioPAX and Systems Biology Markup Language (SBML) for pathways and networks, and ArMet for metabolomics experiments. They aim to facilitate exchange among databases as well as data analysis or visualisation using different tools. Task 2.1 will provide recommendations and tools for data submission and data extraction to facilitate the use of commonly-adopted formats and to keep pace with the field. INTEGRAPE will thus play an important role in the coordination of multiple resources, promoting the use of large international repositories that require the submission of standards-compliant data. Task 2.1 will define priority datasets using case studies of data flow in WG1 and WG3. The framework will propose and test tools that format files and control metadata, will make recommendations to the community, and will establish a list of repositories compliant with these recommendations for data storage.

Task 2.2: Data models, data discovery and automatic data retrieval

Task 2.2 will provide the basis and initial implementation of interlinked data management systems based on API and data indexing with common data models. A large, species-agnostic international initiative is developing a global data model for genetic analysis as well as standard APIs based on this data model (<http://docs.brapi.apiary.io>). INTEGRAPE will build on this initiative and other linked projects. Task 2.2 will test and prioritise APIs that retrieve key objects or concepts in collaboration with WG4. In addition, the framework will check the case studies defined in WG3 for extant APIs and will improve the data models if necessary. Finally, guidelines will be provided to information system managers in the grapevine community. A registry for the community web services will be identified.

Milestones

M1.1. Identification of important repositories for grapevine data (M12)

M1.2. Identification of a registry for web services (M24)

Deliverables

D2.1. List of repositories compliant with data standards and formats

D2.2. Guidelines on data formats disseminated using the IGGP website

D2.3. Catalogue of web services useful for the grapevine community

WG3. Data analysis and best practices

The generation and analysis of omics data is largely automated, so quality control and consistent processing parameters are critical requirements. Accordingly, guidelines for best practice focusing on the promotion of standard protocols for data processing and analysis will be developed and recommended. These guidelines will help researchers to produce high-quality data and carry out informative analysis in the context of a standard framework, which will further contribute to data integration. The COST Action will take advantage of experience gained during the development of standards for other technology platforms, e.g. initiatives such as MAQC (MicroArray Quality Control) which is now in its third phase (MAQC-III for the assessment of NGS platforms) and the latest standards for proteomics and metabolomics data. Another important requirement for both data integration and data analysis is gene annotation. This COST Action will build a centralised framework for the standard annotation of grapevine genes by integrating and disseminating the recently-introduced IGGP guidelines which propose a standard nomenclature for locus identifiers and define a convention for a gene naming in the context of grapevine genomics. Such a framework will facilitate the integration of different omics data types by relying on standard formats and ontologies, such as Plant Ontology (<http://www.plantontology.org>).

Task 3.1: Best practices

Task 3.1 will develop guidelines for the processing and analysis of sequence variation, transcriptomics, proteomics, metabolomics, and phenotypic datasets, using whatever information is available about the samples and analytical protocols. These metadata will be housed alongside the processed data as potential indicators of data quality. Task 3.1 will use case studies of data integration (WG1 and WG2) to facilitate the improvement data and infrastructure interoperability.

Task 3.2: Functional annotation

The IGGP has recently a standard nomenclature for locus identifiers and defined conventions for grapevine gene naming. One of the main goals of the INTEGRAPE initiative is to adopt these guidelines and promote a common, standard functional annotation system for grapevine genes. When possible, links will be provided to functional experimental validation. A new version of the annotation of the grapevine reference genome will be developed for this purpose.

Milestones

M3.1. First data types and data integration or analysis case studies sent to WG1 and WG2 (M12)

Deliverables

D3.1. Guidelines for grapevine data integration, analysis and quality, including the identification of gaps, documented on the INTEGRAPE web site (M36)

D3.2. New version of the grapevine reference genome annotation

WG4. Dissemination and user community assessment of guidelines and recommendations

INTEGRAPE will create a strong link between the grapevine community and large international species-agnostic initiatives currently building open data standards for plants. WG4 will disseminate the recommendations made by WGs 1–3 and will train the grapevine community to follow these recommendations, providing feedback from trainees to each WG.

Task 4.1. Web dissemination and newsletters

Task 4.1 will establish the COST Action web site and links to the IGGP consortium. The web site will be populated with the recommendations from each WG and links to relevant data resources.

Task 4.2 Organisation of webinars and hands-on sessions

Webinars will be organised to demonstrate the scope and use of software tools and information systems. They will be made accessible on the COST Action website to facilitate the dissemination of COST Action outputs and recommendations. At least two hands-on sessions will be organised to test the recommendations and tools on real datasets. These will be organised with WGs 1–3 and will train the wider community to comply with the recommended standards. Feedback provided to each WG will allow their recommendations to be refined, thus facilitating their application.

Deliverables

D4.1. Establish the INTEGRAPE web site (M6)

D4.2. Webinars available on the web site (M36)

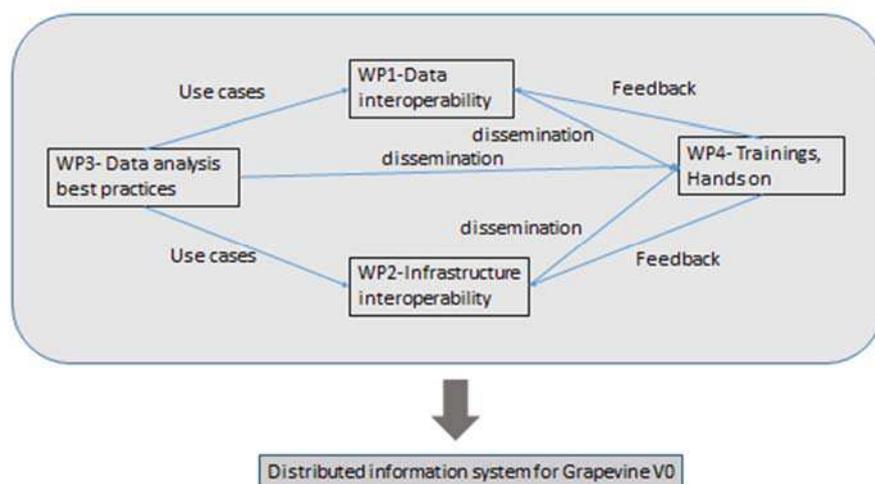
D4.3. First hands-on session (M24)

D4.4. Second hands-on session (M36)

3.1.2. GANTT DIAGRAM

Tasks	Y 1		Y 2		Y 3		Y 4	
	S 1	S 2	S 3	S 4	S 5	S 6	S 7	S 8
1.1 Definition of minimal contextual standards								
1.2 Plant materials and phenotypes								
1.3 Genes, variations, gene expression, proteomics and metabolomics								
2.1 Data file standard formats, repositories and dataset identification								
2.2 Data models, data discovery and automatic retrieval								
3.1 Best practices								
3.2 Functional annotation								
4.1 Web dissemination								
4.2 Organisation of webinars and hands-on sessions								

3.1.3. PERT CHART (OPTIONAL)



3.1.4. RISK AND CONTINGENCY PLANS

Potential problem	Proposed contingency actions	WPs
Failure to agree with the IGGP consortium, to develop WGs in this context and disseminate the results of the COST Action through it.	Organize a workshop with IGGP to identify the problem/bottleneck, work with other international initiatives (e.g. RDA) and publish peer-reviewed articles.	All WGs
Failure to develop appropriate and comprehensive controlled vocabularies and lists of synonyms for markers and genes.	Identify a simple case study that will demonstrate the suitability of the approach, and identify a funded project to support the goal.	WG1
Difficulty in identifying and engaging stable repositories for grapevine data.	Organize interactions with the ELIXIR and OpenAIRE infrastructures.	WG2
Failure to identify case studies using public data.	Identify case studies by preparing joint publications or funding applications.	WG3
Failure to mobilize sufficient resources to update the genome annotation	Develop a strategy to find adequate resources and include these as part of a new project.	WG3

3.2. MANAGEMENT STRUCTURES AND PROCEDURES

This COST Action will provide a framework to coordinate activities in different laboratories that are already funded by national governments and industries and will offer tools that allow the development of a standardised approach for data collection and analysis. The coordination of research and data exchange between laboratories in different Member States will be achieved by realising the common milestones/deliverables planned by the different **WGs** described in Section 3.1. The organisation of this COST Action will follow the stipulations set out in the COST guidelines and the corresponding decision-making bodies will be formed as follows. The COST Action will be coordinated by a **Management Committee (MC)** composed of one member from each signatory country. At the kick-off meeting, the MC will designate a **MC Chair**, a **MC Vice-Chair** and **WG coordinators**. The MC Chair will fulfil the following specific management activities: i) to act as the overall coordinator of the Action (the COST Action Coordinator); ii) to act as an intermediary between the consortium and the European Commission; iii) to coordinate the overall legal, contractual, ethical, financial, and administrative

management of the COST Action; iv) to oversee the promotion of gender balance in the COST network; and v) to chair MC meetings.

WG coordinators will be responsible for achieving the goals and milestones of each WG and for reporting their achievements to the MC, thus providing a well-controlled process for internal monitoring and ensuring the appropriate distribution of financial resources. After the kick-off meeting, the MC will meet again later in the same year during the general conference. Thereafter, the MC will meet biannually, once during the annual general conference and again 6 months later. During MC meetings, WG coordinators will also meet to discuss the progress and coordination of their specific WG objectives. WG coordinators will report on the progress of their WG based on the predefined timetable. In the first MC meeting, the location of the first general conference will be decided. The MC will elect a **Conference Organising Committee (COC)** each year comprising the local organiser and representatives of each WG. The COC will be responsible for the organisation of the annual general conference. The local organiser will chair the COC and report progress to the MC. WGs will contribute to the scientific programs of each general conference through their representatives in the COC. The COC will meet annually during each general conference to decide the location and COC of the next conference, with the approval of the MC. During the kick-off meeting, the MC will also appoint a website coordinator whose role is to develop and maintain the website. In addition, dedicated mailing lists will be established to keep MC members informed about the progress of the COST Action and to communicate with the other MC members and the WGs. If necessary, decisions regarding the approval of STSMs and other matters related to the COST Action will be made through votes delivered via these mailing lists.

The COST Action will provide advanced training for researchers and other key staff in order to facilitate technology transfer and to promote the exchange of knowledge and strengthen collaborations between research partners. It will favour the mobility and training of young researchers, especially at the PhD and early postdoctoral levels. Several STSMs will be financed each year to promote technology exchange and training in different Member States, according to budget limitations and COST guidelines. STSM recipients will be selected by a dedicated STSM Committee, prioritising younger members and paying attention to gender equality. The STSM Committee will be appointed by the MC and will include one Chair and a representative of each WG. STSM applications will be sent to the STSM Committee Chair for presentation to the STSM Committee and consideration under COST guidelines.

A special issue of an internationally recognised journal such as the *Journal of Experimental Botany* will be published in the final year of the COST Action detailing its achievements. The COST Action will also provide an article on the INTEGRAPE portal, tools and database resources for the *Nucleic Acids Research* database issue.

3.3. NETWORK AS A WHOLE

To maximise synergy within the INTEGRAPE COST Action, the research tasks outlined in Section 3.1 will be distributed among different participating research groups. This will prevent the duplication of work and will significantly increase the network's ability to achieve the objectives outlined in Section 1.2.1. The WG coordinators will play a pivotal role in reaching the objectives of the COST Action and their geographical location will be important to facilitate the development of training programs that support diverse research techniques and the integration of partners representing different grapevine growing regions and viticulture practices. The WG coordinators will (i) participate in the plenary and restricted meetings of the MC; (ii) coordinate the activities within each WG to meet the objectives of the COST Action; (iii) propose to the MC the organisation of scientific activities (workshops, summer schools, STSMs) related to their specific WG; (iv) promote interactions and collaborations among different researchers within the WG and with the other WGs; and (v) report WG progress to the COST Action Coordinator and the rest of the MC.

The INTEGRAPE COST Action will seek active collaborations with other COST Actions and other European and international research programmes because scientific data standards must be developed and agreed at the international level in order to be effective. Many members of this COST Action are already members of other European programs and international consortia, and have already contacted a number of non-EU competitive research groups from the USA, Canada, Australia, New Zealand, South Africa, Chile, Argentina and China, which have expressed interest in the objectives of the INTEGRAPE consortium and have agreed to coordinate their efforts. The MC will promote the expansion of contacts with these international research programs by inviting leading scientists investigating topics dealing with

data standards, fruit development and fruit quality to participate in the general conferences and focused workshops. When appropriate, the COST Action will also seek interactions with research groups from Mediterranean countries outside Europe, given the importance of grape production in these regions.

Gender equality will be addressed in this COST Action by ensuring an appropriate gender balance in all its activities and its decision-making bodies. The MC will place this as a standard item on all its agendas. Likewise, the COST Action is also committed to involve Early Career Investigators. Moreover, female scientists have already played a major role in the conception of the COST Action and its objectives. The COST Action will maintain gender and age balance in the elected leading team members; young female scientists will also play leading roles in the COST Action management. Gender balance will be also taken into account in the workshops, teaching activities and STSMs where Early Career Investigators are expected to form the majority of participants. This COST Action will also explore every opportunity to encourage and improve the involvement of young scientists at all levels including their active participation in WG meetings, STSMs, conferences, symposia and INTEGRAPE workshops.