COST Action
CA17111
INTEGRAPE

1st Annual Meeting

INTEGRAPE 2019

Data Integration as a key step for future grapevine

Book of Abstracts

CIHEAM
MAI CHANIA

Chania, Greece
25 - 28 March, 2019

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the Horizon 2020 Framework Programme of the European Union

VESPUCCI: the integrated gene expression database for grapevine

Moretto Marco[1], Sonego Paolo[1], Pilati Stefania[2], Malacarne Giulia[2], Costantini Laura[2], Moser Claudio[2], Engelen Kristof[1]

[1]Unit of Computational Biology, Research and Innovation Center, Fondazione Edmund Mach, San Michele all'Adige, Italy, [2]Department of Genomics and Biology of Fruit Crop, Research and Innovation Center, Fondazione Edmund Mach, San Michele all'Adige, Italy

VESPUCCI, the Vitis Expression Studies Platform Using COLOMBOS Compendia Instances, is an on-line gene expression compendium that integrates publicly available transcriptomic data for grapevine measured using both microarray and RNA-seq. It was originally developed in 2014 and has since grown from the initial ~1500 samples to the current ~2000 samples. Each sample has been manually annotated using a controlled vocabulary developed ad hoc to ensure both human readability and computational tractability to precisely describe which parameters have changed across different experimental conditions. VESPUCCI was created based on an approach for dealing with the large heterogeneity of data formats present in public databases, and to integrate cross-platform gene expression experiments in one dedicated, coherent database. The expression data in the compendium can be visually explored using several tools provided by the web interface, an exploratory tool meant to assist more dedicated research in grapevine genomics, biology, and physiology. The technology underlying VESPUCCI has recently been completely overhauled in order to simplify the data management. This update has involved both the compendium creation step and the data query step. Concerning the former, the new technology supports different normalization strategies within the same compendium, and thus the legacy normalization data in which samples were treated disregarding the fact that they were measuring the same biological condition will be available together with normalized data in which replicated measurement are taken into consideration. Then, the sample annotation has been revisited to better exploit existing ontologies such as the Plant Ontology, the Plant Trait Ontology, the Plant Experimental Conditions Ontology and the Unit of Measurement Ontology. Finally, the programmatic access to the compendium is going to be easier and more powerful, simplifying the integration of VESPUCCI with other services. All these improvements will be implemented in the new release of the VESPUCCI, which will include all publicly available expression experiment up to 2019, raising the number of collected samples to ~3000.