Open Conference on
Grapevine Physiology
and Biotechnology
LOGROÑO - LA RIOJA
7-11 July 2024

Harnessing whole genome sequencing data to predict protein structure and function variation in grapevine

Dylan Grobler[1*], Justin Lashbrooke[1], Pablo Carbonell-Bejerano[2]

[1] *Department of Genetics, Stellenbosch University. (7600) Stellenbosch, South Africa*
[2] *Instituto de Ciencias de la Vid y del Vino, ICVV, CSIC – Universidad de La Rioja – Gobierno de La Rioja. (26007) La Rioja, Logroño, Spain*

[*]*Corresponding author: dylang@sun.ac.za*

Grapevine (*Vitis vinifera*) is amongst the world's most cultivated fruit crops, and of global and economic significance, producing a wide variety of grape-derived products, including wine, and table grapes. The genus *Vitis*, encompassing approximately 70 naturally occurring inter-fertile species, exhibits extensive genetic and phenotypic diversity, highlighted by the global cultivation of thousands of predominantly *Vitis vinifera* cultivars. Despite the importance of harnessing its naturally occurring genetic diversity to pursue traits of interest, especially considering the continued and growing demand for sustainable high-quality grape production, the systematic characterization of available functional genetic variants remains limited. Such characterizations hold the potential not only as a critical tool for directed breeding, including the identification of molecular markers for genetic selection, but would also enable the functional characterization of genes that may exert influence over key functional traits. This project aims to comprehensively analyze global genetic variation in grapevine cultivars, emphasizing the development of methodologies for large-scale prediction of the impact of genomic sequence variations on protein structure and function. To this end, a curated bioinformatics pipeline was designed to enable the detection of sequence variation present within the coding regions of previously reported *Vitis* Whole-Genome Sequencing (WGS) datasets. Validation of this pipeline included utilizing genomic regions with known Insertions and Deletions (InDels) and Single Nucleotide Polymorphisms (SNPs) to quantify its efficiency. Lastly, comparative analyses and recording of coding region sequence variation among grapevine cultivars were conducted to facilitate future predictions regarding the impact of genetic variations on protein structure and function.

Keywords: grapevine, genetic resources, genetic variation, protein prediction