# Genetic Variability and Disease Resistance Analysis of Grapevine from Armenia

Author: Emma Hovhannisyan
*BS in Data Science*
*American University of Armenia*

Supervisor: Dr. Hans Binder
*IZBI Affiliation: University of Leipzig*

Supervisor: Maria Nikoghosyan
*Armenian Bioinformatics Institute*

*Abstract*—**The paper studies population structure and genetic variability of Armenian cultivated and wild grapevines. Armenian grapevine genetics is shown to be different from the worldwide population. Further, the study considers powdery mildew disease, as some Armenian wild samples have shown to be resistant to the disease. By conducting GWAS on the wild Armenian samples, the study reveals new genetic markers that are significantly associated with powdery mildew disease resistance.**

*Index Terms*—**grapevine, GWAS, SNP, Armenia**

## I. INTRODUCTION

The cultivated grapevine (Vitis Vinifera) has been used by humans over centuries. It is an important crop not only for its agricultural and economical significance, but also for its socio-cultural impact. Because of grapevine's long standing relationship with humans, studying the crop's genetic structure and diversity can reveal hidden patterns of cultivation and dissemination in the past. According to a recent study, grapevine has been domesticated as early as 11,000 years ago concurrently in Western Asia (today's Lebanon area) and the Caucasus (today's Georgia, Armenia) [1]. This finding has increased the interest of investigating grapevine samples from these areas.

Analyzing the genetics of the grapevine however not only can elucidate past events but also can point to ways of making the future of agriculture more sustainable and of more high quality. Various fungal diseases have been hindering the quality and quantity of the grapevine yield, fruit and wine. The main approach of fighting these diseases has been the usage of fungicides. It is approximated that around 70% of agrochemicals used in the European Union are related to viticulture, with the primary purpose of controlling downy and powdery mildews. However, the overuse of the fungicides, besides being expensive, can negatively affect human health and the environment [2]. Finding resistance-related loci in the DNA of grapevines and applying interbreeding or genetic engineering according to the findings, would be a more beneficial and eco-friendly solution to the problem of fungal diseases. Thus, finding novel resistance loci is of high interest. It can be done by studying genetically distinct populations of the grapevines.

One of the most impactful diseases of Vitis Vinifera is powdery mildew. The biotrophic fungus Erysiphe necator is responsible for causing powdery mildew. It can result in a decrease in cluster weight, delayed fruit ripening, and reduced rates of photosynthesis and transpiration. Moreover, powdery mildew causes metabolic reprogramming in its host, which impacts primary metabolism [3]. Most commercially used cultivated crops are susceptible to this disease which has resulted in the overuse of pesticides and its negative consequences discussed above. Previous studies have found two main gene families that showed association with resistance to powdery mildew, namely Run (Resistance to Uncinula necator) and Ren (Resistance to Erysiphe necator). The resistance provided by these gene families is linked to the presence of the loci, as there are currently no single genes that can offer strong genetic resistance on their own [3]. One of the commonly used methods to identify genomic variants, across the whole genome, which are associated with the desired complex phenotype, e.g. resistance to disease, is genome-wide association studies (GWAS) [4].

In one of the possible regions of origin of the cultivation of grapevine, Armenia, wild varieties of the grapevine show considerable genetic diversity [5]. Interestingly, some wild crops in Armenia have shown resistance to powdery mildew, which indicates the presence of genetic variations responsible for the resistance. In our study, we have used GWAS to identify specific genomic regions or potential genetic markers associated with resistance to powdery mildew. We have described 7 single nucleotide polymorphisms (SNPs), which have significantly higher allele frequency in the resistant samples, compared with the controls. The findings can be useful in future implementations of interbreeding or genetic engineering, aiming to make crops of better quality.

In short, the study analyzes genetic variability of Armenian grapevines in comparison to the worldwide accessions, analyzes genetic structure within the Armenian grapevine population and identifies genetic markers that are associated with powdery mildew disease.

## II. MATERIALS AND METHODS

### A. Data

In the main study, the whole genomes of 160 Armenian and 943 worldwide grapevine accessions were analyzed, which were taken from the publication of Y. Dong et al. [1]. The Armenian samples were collected from different regions of Armenia including Aragatostni, Ararat Valley, Vayots Dzor, Syunik, Tavush. 72 samples refer to wild vine (V. vitis
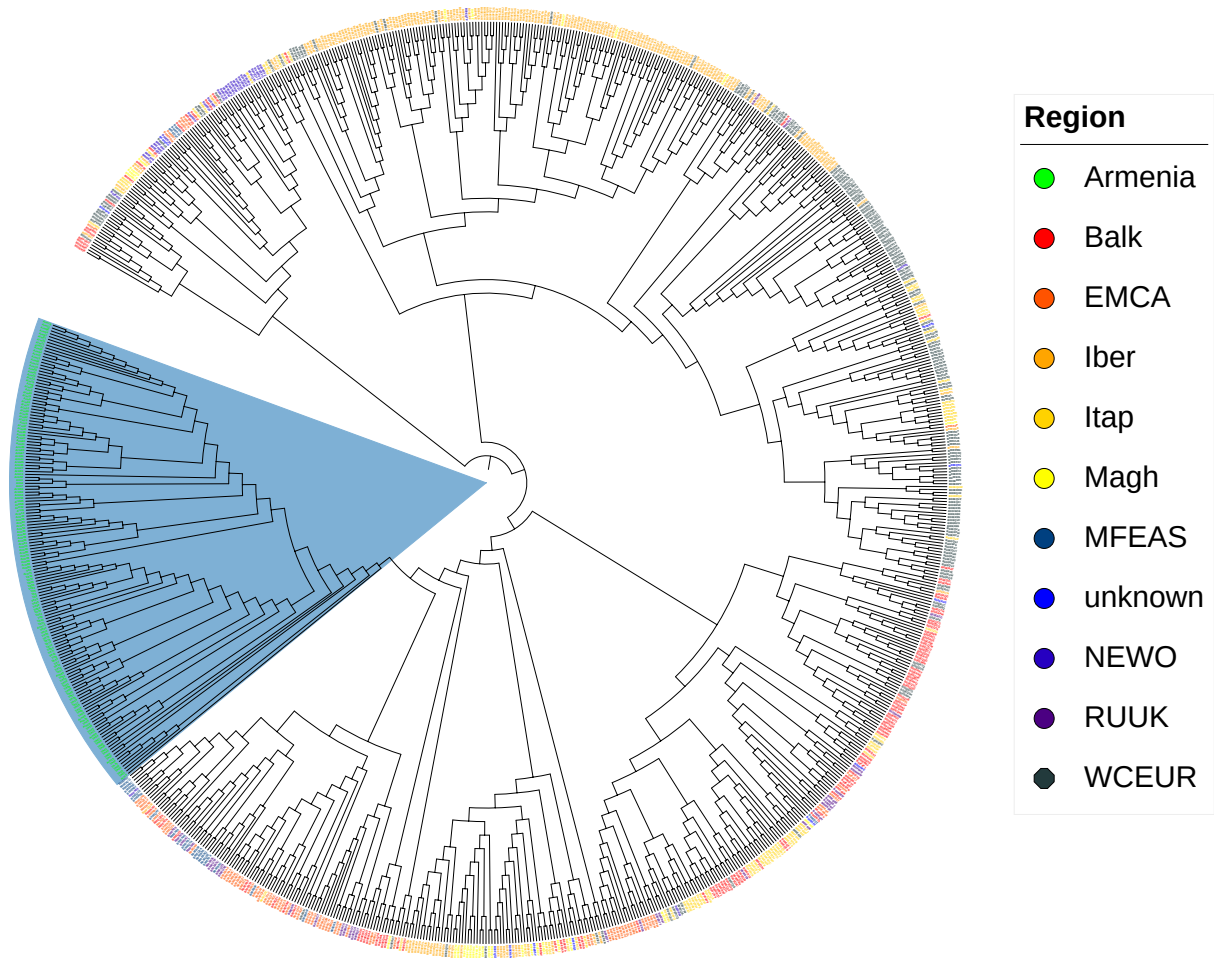
Fig. 1: Phylogenetic tree of worldwide accessions. The blue coloured cluster indicates the Armenian accessions. Label colors indicate different countries, green marks Armenia, etc.

ssp. sylvestris) , 84 samples to cultivated vine (V. vitis ssp. vinifera) and 4 to cross/hybrid of wild and cultivated vine. Genomic sequences were obtained by pair-end sequencing of the samples on a Illumina NovaSeq 6000 device. The raw reads were aligned to the PN40024 (12X.2 assembly) reference genome. The final genotype data in VCF format consisted of more than 15 million SNPs.

The 63 wild armenian samples were phenotyped empirically to be resistant or non-resistant to powdery mildew. The phenotyping was confirmed by REN1 resistance locus SSR markers. 21 wild samples were phenotyped to be resistant case samples, while 42 were non-resistant and used as control samples.

### B. Data Preprocessing

In order to perform further analysis, genotype data of wild samples were preprocessed. At first, only biallelic SNPs were selected. Then additional filters were applied for the SNPs with the PLINK tool (version 1.9) [6], which allows to manipulate and to analyze thousands of genomic markers. The following parameters were used: SNPs and samples with missing genotype rates (mind and geno) higher than 0.2 were removed, SNPs with minor allele frequency (maf) smaller than 0.05 and

P value larger than 1e-6 for the genotypic chi-squared test of Hardy-Weinberg Equilibrium (hwe) were removed. Then, the samples which deviated more than 3 standard deviations from the heterozygosity rate mean were also removed. The final genotype data of wild Armenian accessions included 12,657,999 SNPs and 63 samples.

### C. Population Structure Analysis

Phylogenetic trees were constructed with the neighbor joining method by VCF-kit tool implemented in python [7]. Principal Component Analysis (PCA) was carried out with the PLINK tool and visualized in R [6]. The genomic structure of the wild grape population was obtained with ADMIXTURE tool and visualized in R [8].

### D. Genome-wide association study

GWAS analysis was performed only for 63 wild samples, with the consideration of 2 groups; 21 case samples, which are phenotypically resistant to powdery mildew and 42 control samples, which can be affected by this disease. GWAS was performed with the GAPIT package in R [9].
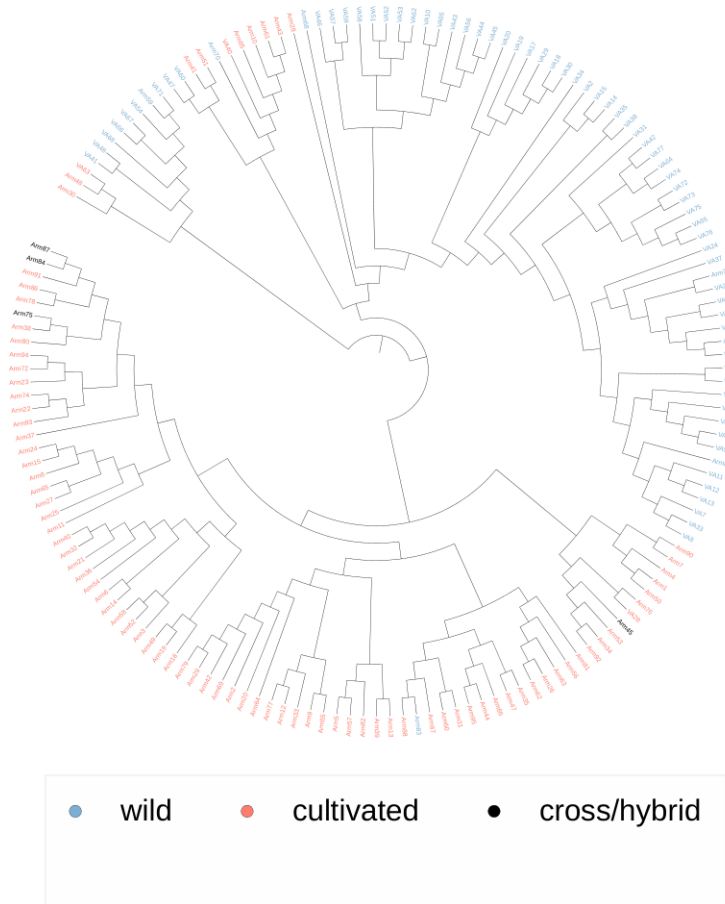
Fig. 2: Phylogenetic tree of the Armenian population of wild, cultivated and cross/hybrid accessions.

## III. RESULTS AND DISCUSSION

### A. Population Structure Analysis

First, a phylogenetic tree of the whole genome sequences of accessions collected worldwide was constructed (Fig. 1). It showed that accessions from Armenia form a distinct cluster, meaning that Armenian accessions are genetically similar to each other and clearly differentiate from the worldwide samples. These results imply that it is important to look for genetic markers in Armenian samples, as they will most probably be unique and novel.

The zoomed in phylogenetic tree of only Armenian samples was constructed, which showed that wild and cultivated varieties strongly fall into different clusters (Fig. 2). This shows that the cultivation process strongly modified the genome of the wild ancestor grapevines. Thus, GWAS should be restricted to the wild samples as one distinct population and the wild samples shouldn't be used with cultivated samples in the association models.

Further, we have assessed the population structure of the wild Armenian samples using PCA and ADMIXTURE methods, to account for hidden substructures that can influence GWAS analysis. The PCA analysis indicates four distinct clusters (Fig. 3.A). ADMIXTURE captured more of the population's structure and showed 7 clusters that were consistent over k=3,4,5,6 (Fig. 3.B). The K=6 solution divides the population into clusters dominated by red, yellow, green, mixture, blue, dark blue and pink components (from the left to the right), which represent the different ancestral populations that contribute to the sample's genetic makeup. Interestingly, the three clusters at the left enrich resistant accessions while the four clusters to the right enrich non-resistant ones. Note that we identified overall seven clusters where one is a composite one.

### B. GWA Studies

To identify genetic variants that are associated with the binary trait of being or not being resistant to powdery mildew disease Genome-Wide Association Study (GWAS) was conducted with 63 Armenian wild grapevine samples that have passed genotype and phenotype quality checks. GWAS is a type of study that analyzes a large number of genetic variants across the entire genome to identify which variants are more common in samples with the trait than in those without it. There were different statistical models used to find the SNPs associated with resistance, namely General Linear Model (GLM), Mixed Linear Model (MLM) and Settlement of MLM Under Progressively Exclusive Relationship (SUPER). Association results are visualized by Manhattan Plots (Fig. 4).
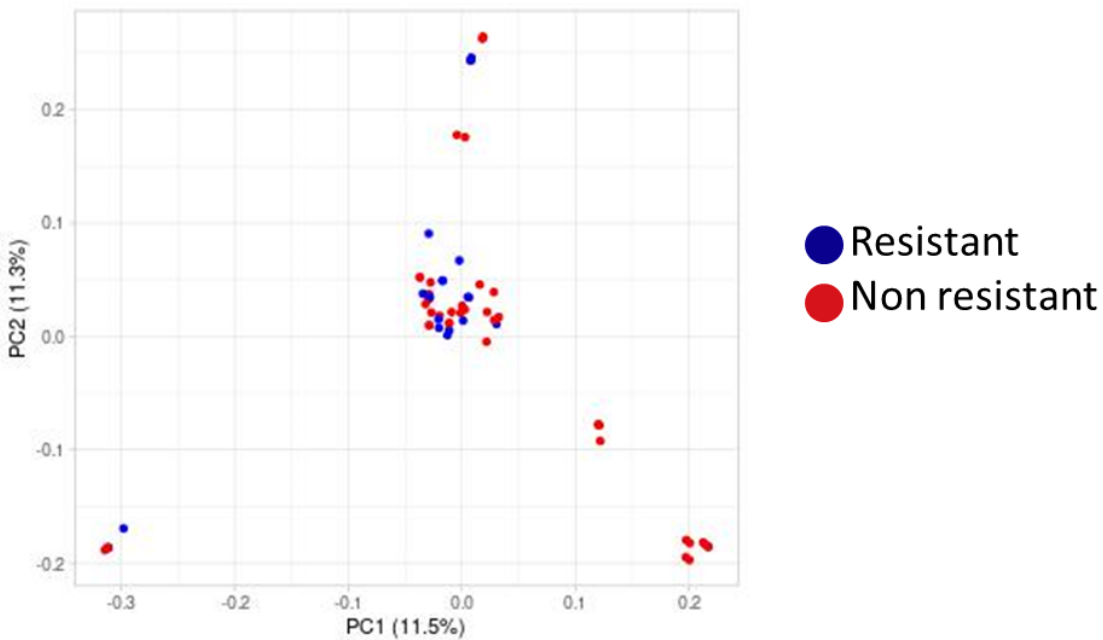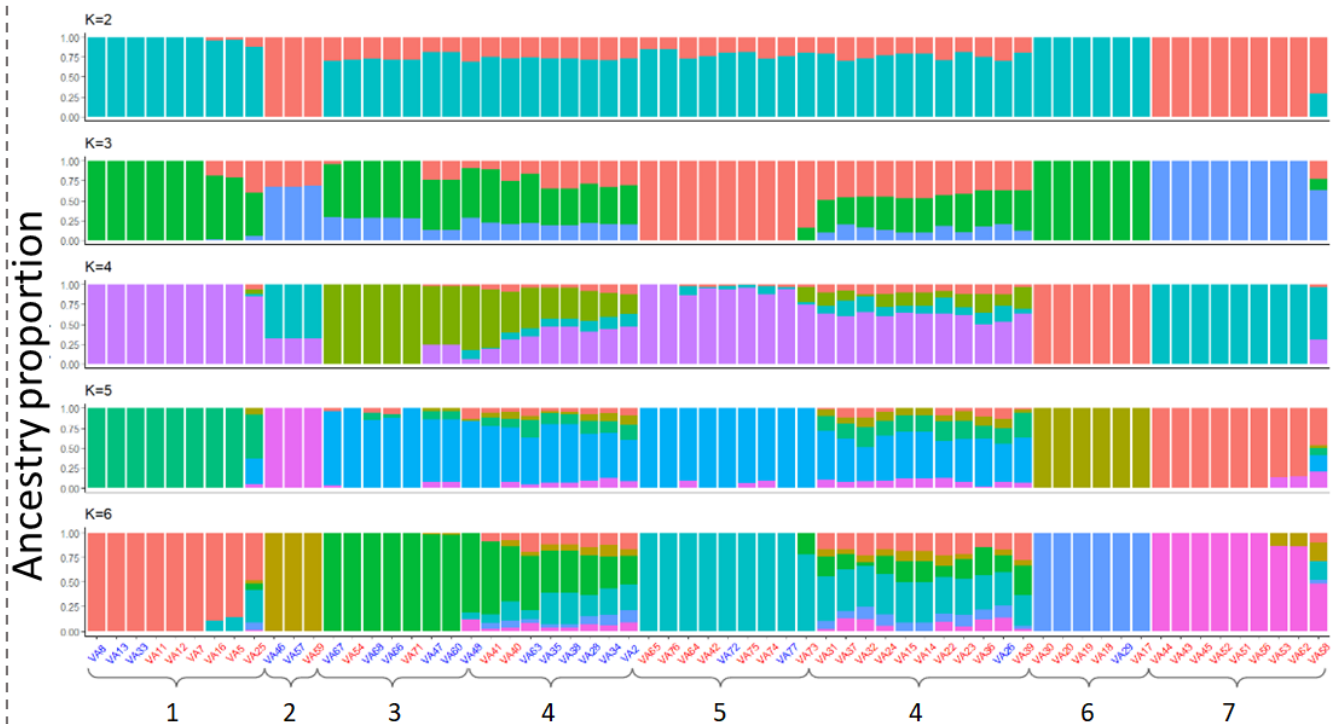
Fig. 3: Genomic variability of Armenian wild grapes. A) PCA plot of Armenian wild grapevine genomes. By blue are coloured resistant samples against powdery mildew, by red: non-resistant ones. B) Admixture plot of Armenian wild grapes. Each row corresponds to k=2...6. Each bar corresponds to one sample, the labels of the samples are given on the bottom of the plot. Labels of resistant samples are coloured with blue, while non-resistant ones' are colored with red.
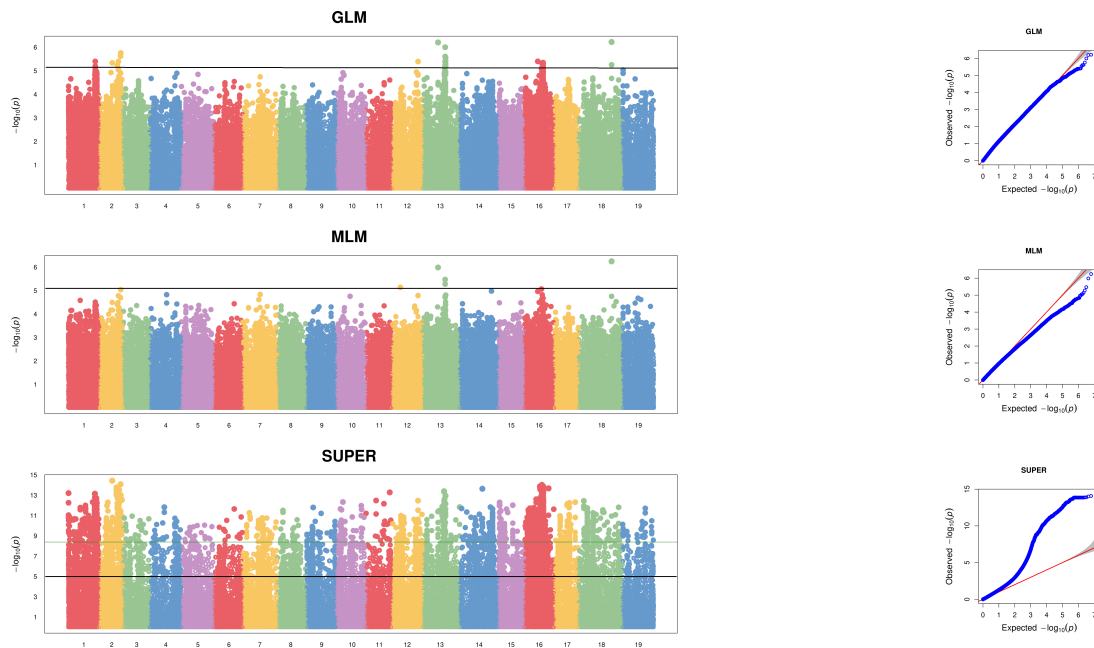
Fig. 4: Manhattan (left part) and QQ (right part) plots of GLM, MLM and SUPER association models. Black horizontal lines indicate Bonferroni adjusted $-log_{10}$ suggestive significance threshold of 5.

To account for population stratification, Q-matrix for K=6 from ADMIXTURE analysis was used as a covariate in the models, as the matrix captured more clusters of structural differences (7 clusters) in the wild population than the Principal Component Analysis (4 clusters).

QQ plots were used to evaluate the models, which showed that SUPER model is performing poorly and is giving many false positive results (Fig. 4).

SNPs were identified to be significantly associated with Powdery Mildew resistance in the models, using the $-log_{10}$ p-values that are greater than the widely accepted Bonferroni-adjusted suggestive significance threshold of 5. There were 7 common significant SNPs in the models. Two most significant SNPs common in the models, with both p-values equal to 5.7e-07 in MLM and 6e-07 in GLM, are located on chromosome 18 at positions 26,090,136bp and 26,090,162bp respectively. Next three common significant SNPs, with p-values equal to 1.0e-06, 3.4e-06 and 5.3e-06 in MLM and 6.3e-07, 1.0e-06, 2.5e-06 in GLM, are located on the 13th chromosome at positions 12,090,103 bp, 17,657,771 bp and 17,657,967 bp, respectively (Tab.1).

### C. Gene Mapping

Two significant SNPs on chromosome 13 are located in the intergenic region, which corresponds to the previously identified resistance REN1 locus on chromosome 13 ranging from 16,690,324 bp to 18,130,936 bp. Moreover, 21 SNPs in MLM and 31 SNPs in GLM in the top 100 significant SNPS are in the REN1 locus. These results suggest novel markers in the form of the significant SNPs for the REN1 resistance locus.

TABLE I: SNPs significantly associated with resistance

| SNP | chr | position | p-value(GLM) | p-value(MLM) |
|-----|-----|----------|--------------|--------------|
| 1 | 18 | 26090136 | 6.015895e-07 | 5.699882e-07 |
| 2 | 18 | 26090162 | 6.015895e-07 | 5.699882e-07 |
| 3 | 13 | 12090103 | 6.255398e-07 | 1.0385e-06 |
| 4 | 13 | 17657771 | 1.000623e-06 | 3.396800e-06 |
| 5 | 13 | 17657967 | 2.482198e-06 | 5.322627e-06 |
| 6 | 16 | 13425481 | 8.231431e-06 | 8.637895e-06 |
| 7 | 2 | 16974620 | 1.759501e-06 | 9.159964e-06 |

The significant SNPs on chromosome 18,16 and 2 suggest possible novel resistance loci.

### IV. CONCLUSION

The study showed that Armenian grapevines are distinct from the worldwide grapevine population and present an isolated population. Taking account the population structure of armenian grapevines, GWAS analysis of wild armenian grapevines found 2 new SNP markers for previously found resistance locus on chromosome 13 and identified possible new resistant loci on chromosomes 18, 16 and 2. Future work may include other models to verify the results, affirming the results experimentally and finding new biological functional interpretations of these loci. The study highlights the importance of analyzing the genetic information of Armenian grapevine accessions and how they can offer novel and unique genetic markers.

### REFERENCES

[1] Dong, Y, et al. (2023). Dual domestications and origin of traits in grapevine evolution. *Science*. 379,892-901.10.1126/science.add8655

[2] Hollomon, D. (2015). Fungicide resistance: 40 years on and still a major problem. *Fungicide Resistance in Plant Pathogens: Principles and a Guide to Practical Management*, eds H. Ishii, and D. Hollomon (Cham: Springer), 3–11. 10.1007/978-4-431-55642-8_1

[3] Sosa-Zuniga, V., Vidal Valenzuela, Á., Barba, P., Espinoza Cancino, C., Romero-Romero, J. L., Arce-Johnson, P. (2022). Powdery Mildew Resistance Genes in Vines: An Opportunity to Achieve a More Sustainable Viticulture. *Pathogens* (Basel, Switzerland), 11(6), 703. https://doi.org/10.3390/pathogens11060703

[4] Uffelmann, E., Huang, Q.Q., Munung, N.S. et al. (2021). Genome-wide association studies. *Nat Rev Methods Primers* 1, 59. https://doi.org/10.1038/s43586-021-00056-9

[5] Margaryan, K., Melyan, G., Röckel, F., Töpfer, R., Maul, E. (2021). Genetic Diversity of Armenian Grapevine (Vitis vinifera L.) Germplasm: Molecular Characterization and Parentage Analysis. *Biology*, 10(12), 1279. MDPI AG. http://dx.doi.org/10.3390/biology10121279

[6] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 81(3):559-75. 10.1086/519795. Epub. PMID: 17701901; PMCID: PMC1950838.

[7] Cook, D, Andersen,E . (2017). VCF-kit: assorted utilities for the variant call format, *Bioinformatics*, Volume 33, Issue 10, May. Pages 1581–1582, https://doi.org/10.1093/bioinformatics/btx011

[8] D.H. Alexander, J. Novembre, and K. Lange. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19:1655–1664.

[9] Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399. 10.1093/bioinformatics/bts444