

Analysis of Croatian wild and cultivated grapevine diversity by genotyping by sequencing

Marinov, Luka¹; Zdunić, Goran^{1*}; Di Gaspero, Gabriele²; Magris, Gabriele^{2,3}; Morgante, Michele^{2,3}; Lukšić, Katarina¹; Hančević, Katarina¹; Maletić, Edi⁴; Bubola, Marijan⁵; Pejić, Ivan⁶

¹Institute for Adriatic Crops and Karst Reclamation, Split, Croatia

²Istituto di Genomica Applicata, Udine, Italy

³University of Udine, Department of Agricultural, Food, Environmental and Animal Sciences, Udine, Italy

⁴University of Zagreb, Faculty of Agriculture, Department of Viticulture and Enology, Zagreb, Croatia

⁵Institute of Agriculture and Tourism, Poreč, Croatia

⁶University of Zagreb, Faculty of Agriculture, Department of Plant Breeding, Genetics and Biometrics, Zagreb, Croatia

*goran.zdunic@krs.hr

Abstract

The traditional Croatian grapevine varieties (*Vitis vinifera* L. subsp. *vinifera*) represent a significant part of the germplasm of the so-called Balkan grapes. Over time and the rich history of this area, numerous varieties have accumulated and are still grown commercially. There are also several populations of wild grapevine (*Vitis vinifera* subsp. *sylvestris* Hegi Gmel) in natural sites in this region. Both methods, morphometric and SSR markers, have previously been used to validate identities and investigate genetic variability of Croatian germplasm, but without a detailed characterization of their genomes. Genotyping-by-sequencing (GBS) is a recent approach to effectively characterize individuals at thousands of variant sites based on next generation sequencing technology. The aim of this study was to use GBS to: (i) define a panel of single nucleotide polymorphisms (SNPs) and determine the identity of Croatian cultivated and wild grapevine accessions, (ii) verify the presence of homonyms and synonyms, (iii) determine the genealogical relationships and the varieties that acted as the main progenitors of the Croatian germplasm, (iv) determine the distance between wild and cultivated accessions. We successfully genotyped 192 grapevine accessions, including 132 cultivated (*vinifera*) and 60 wild (*sylvestris*) individuals. Approximately 50,000 genomic intervals in the gene space were analyzed using a targeted sequencing system based on single primer enrichment technology. The filtered DNA reads were aligned to the PN40024 v0 reference genome and more than 500,000 high quality SNPs were identified using the Genome Analysis Toolkit (GATK) software. Twelve synonyms were found within Croatian varieties, while eleven varieties were matched to international varieties, mostly from neighboring countries, indicating historical migration of cultivars. Distance- and model-based cluster analysis showed a clear genetic separation between cultivated varieties and wild accessions but also spotted the presence of rare feral individuals in natural populations of wild grapevines. For the prediction of relatedness, an identity-by-descent (IBD) analysis identified 'Plavac Mali', 'Bombino bianco', 'Bljuzgavac', 'Heunisch weiss' and 'Tribidrag' as varieties with highest numbers of first-degree relationships among the genotypes studied.

Keywords: single nucleotide polymorphism, native grapevines, Croatian germplasm, first degree relationship, genetic diversity