

CRISPR-CAS9-BASED REPEAT DEPLETION FOR HIGH-THROUGHPUT GENOTYPING OF COMPLEX PLANT GENOMES

ROSSATO M.*, MARCOLUNGO L.*, DE ANTONI L.*, LOPATRIELLO G.*, BELLUCCI E.**,
CORTINOVIS G.**, FRASCARELLI G.**, NANNI L.**, BITOCCHI E.**, BETT K.***,
RAMSAY L.***, KONKIN D. J.****, DELLEDONNE M.*, PAPA R.**

*) Dpt. Biotechnology, University of Verona

**) Dpt. Agricultural, Food and Environmental Sciences, Polytechnic
University of Marche

***) Dpt. of Plant Sciences, University of Saskatchewan

****) National Research Council Canada

complex genomes, repeats, Genotyping by sequencing, CRISPR/Cas9

High-throughput genotyping is increasingly required for large-scale diversity analysis, as those foreseen by the INCREASE H2020 EU project (www.pulsesincrease.eu), focused on the characterization of large genetic resources of four food-legumes: chickpea, common bean, lentil and lupin. Thanks to the reduced ascertainment bias, genotyping methods based on sequencing are progressively replacing traditional ones, exploiting PCR and arrays. However cost-effective sequencing approaches allowing a dense and genome-wide genotyping are still missing, especially for large -and highly repetitive- genomes as those of some legumes (e.g. lentil). In this work, conducted in the framework of the EU project INCREASE, we exploited the CRISPR-Cas9 system to deplete un-informative repetitive elements and concentrate data on protein-coding and regulatory regions. The Cas9-depletion system was tested on the lentil genome (3.7Gb with 85% repeats) in combination with a very large array of guideRNA (~600K), to refine the content of different types of sequencing libraries. The CRISPR-Cas9-based depletion showed good efficiency, reducing both coverage depth and breadth (down to 50%) on repeats, while improving the same parameters on the coding fraction (2.8x and 1.6x increase, respectively). Thanks to such “repeat-to-coding” shift of sequencing data, the approach allowed a net increase in the number of genotyped bases (up to 12x), as compared to standard whole genome sequencing. Similarly, it significantly outperformed exome-

sequencing when the methods were compared at equivalent starting budget. The method showed similar performances on different multiplexing levels, type of libraries and genotypes, including different cultivars as well as a close-related species (*L.orientalis*), thus demonstrating its potential for large genotyping studies. Overall, the CRISPR-Cas9-driven repeat-depletion demonstrated an effective approach to focus sequencing data on meaningful genomic regions, thus providing an innovative opportunity to improve high-density and genome-wide genotyping in large and repetitive genomes.