Reference leaf Transcriptomes for potato cultivars: Desiree and PW363

Maja Zagorščak¹, Marko Petek¹, Mohamed Zouine², Kristina Gruden¹
¹National Institute of Biology - Department of Biotechnology and Systems Biology, Ljubljana, Slovenia
²Ecole Nationale Supérieure Agronomique de Toulouse, Toulouse, France

The rapid development of modern life science technologies, such as Next Generation Sequencing (NGS) techniques, allows the generation of biological data with increasing speed and precision. Most potato cultivars are highly heterozygous tetraploids with high genetic variability while being susceptible to pathogens, pests and inbreeding depression. To bypass polyploidy related sequencing problems, Potato Genome Sequencing Consortium (PGSC, 2011) sequenced a double monoploid derived from S. tuberosum group Phureja.

In order to avoid problems, such as discriminating between paralogous genes, divergence and expression bias between the reference genome and potato cultivars, and to identify traits that are not present in initially sequenced genotype, RNA-sequencing for cv. Desiree and cv. PW363 leaves was conducted on Illumina NGS platform. In house generated raw reads were complemented with data already deposited in the NCBI database and stNIB-v1 S. tuberosum gene groups, which included two genome assemblies and two EST datasets (Ramšak et al., 2014). The preliminary transcriptomes were assembled using both hybrid and de novo assembly approaches. The hybrid approach, combining genome-guided and de novo RNA-Seq assembly, was implemented using the pipeline available from CLC Genomics Workbench 7.0.3¹. De novo assembly was performed using Trinity (Grabherr et al., 2011). The resulting two sets of preliminary transcriptomes were then merged using the CD-HIT clustering algorithm (Limin et al., 2012) and merged with existing gene models with BLAST against stNIB-v1. The presumed novel clusters were annotated using SwissProt, PGSC DM v3.4 super-scaffolds and NCBI-nt databases. Initial potato pangenome containing 35609 genes was expanded with 24999 potential new transcripts, and will serve to further expand knowledge on the potato pathogen interactions.

References

¹ www.clcbio.com