**Reference Leaf Transcriptomes for Potato Cultivars: Desiree and PW363**

Maja Zagorščak¹, Marko Petek¹, Mohamed Zouine², Živa Ramšak¹, Kristina Gruden¹

¹National Institute of Biology - Department of Biotechnology and Systems Biology, Ljubljana, Slovenia
²Ecole Nationale Supérieure Agronomique de Toulouse, Toulouse, France

Web: [http://www.nib.si](http://www.nib.si), Email: Maja.Zagorscak@nib.si, Phone: +386 (0)59 232 843

**Objectives**

Assemble reference leaf transcriptomes for Desiree and PW363 cultivars to assist with:
- StNIB-v1 potato pangenome expansion with novel genes and alleles
- Improved quantification with RNASeq in further experiments

**Introduction**

Most potato cultivars are highly heterozygous tetraploids with high genetic variability. To bypass polyploidy related sequencing problems, Potato Genome Sequencing Consortium (PGSC, 2011) sequenced a double monoploid derived from *S. tuberosum* group Phureja. While this approach resulted in an acceptable gene model, it does not cover all genes and gene variants present in cultivated potato.

In order to avoid this problem, to discriminate between paralogous genes 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.

**Conclusion**

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.

**Examples of further studies**

- Confocal microscopy
- Network analysis

**Workflow and results**

**Desiree or PW**

- [Image of workflow diagram]

**Hybrid approach**

- CLC Genomics reference-based assembly
  - Desiree: 11251 contigs
  - PW: 10258 contigs
- CLC Genomics de novo assembly
  - Desiree: 11251 contigs
  - PW: 10258 contigs

**De novo approach**

- Trinity de novo assembly
  - Desiree: 183080 contigs
  - PW: 138272 contigs

**Data cleansing**

- Duplicates
- Highly ambiguous contigs (% nt(N) > 25%)
- Artifacts, highly erroneous contigs (length ∈ [200,5000])

**Clustering**

- CD-hit-est algorithm (95%)
  - Desiree: 279777 sequences in 194891 clusters
  - PW: 227042 sequences in 149137 clusters

**Database search**

- BLAST algorithm for rank assignment
  - True positives identified according to sequence weight criteria or high (% nt(N) > 0.50)
  - Missing, dubious sequences assigned with the help of BitScore similarity threshold (weight ∈ [0.00,1.00])

**Confirmed/discovered novel genes and gene isoforms**

- BLASTx against stNIB-v1 or tBLASTx against stNIB-v1
  - Desiree: 202489 hits
  - PW: 176048 hits
- BLASTx against SwissProt
  - Desiree: 759 hits
  - PW: 3071 hits
- BLASTn against PGSC DM v3.4 supercolds
  - Desiree: 71061 hits
  - PW: 38328 hits
- BLASTn against NCBI-nt
  - Desiree: 912 hits
  - PW: 640 hits

**Desiree Reference Leaf Transcriptome**

- Chlorotic sequences → excluded
- De Bruijn graph artifacts → excluded
- Highly repetitive sequences → excluded
- 275221 transcripts and isoforms

**PW Reference Leaf Transcriptome**

- Chlorotic sequences → excluded
- De Bruijn graph artifacts → excluded
- Highly repetitive sequences → excluded
- 218087 transcripts and isoforms

**Potato pangenome**

50% genes verified successfully expanded

**StNIB-v1 potato pangenome**

- [Image of StNIB-v1 potato pangenome]