Reference Leaf Transcriptomes for Potato Cultivars: Desiree and PW363

Maja Zagorščak[1], Marko Petek[1], Mohamed Zouine[2], Živa Ramšak[1], Kristina Gruden[1]

[1]National Institute of Biology - Department of Biotechnology and Systems Biology, Ljubljana, Slovenia

[2]Ecole Nationale Supérieure Agronomique de Toulouse, Toulouse, France

Web: 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:
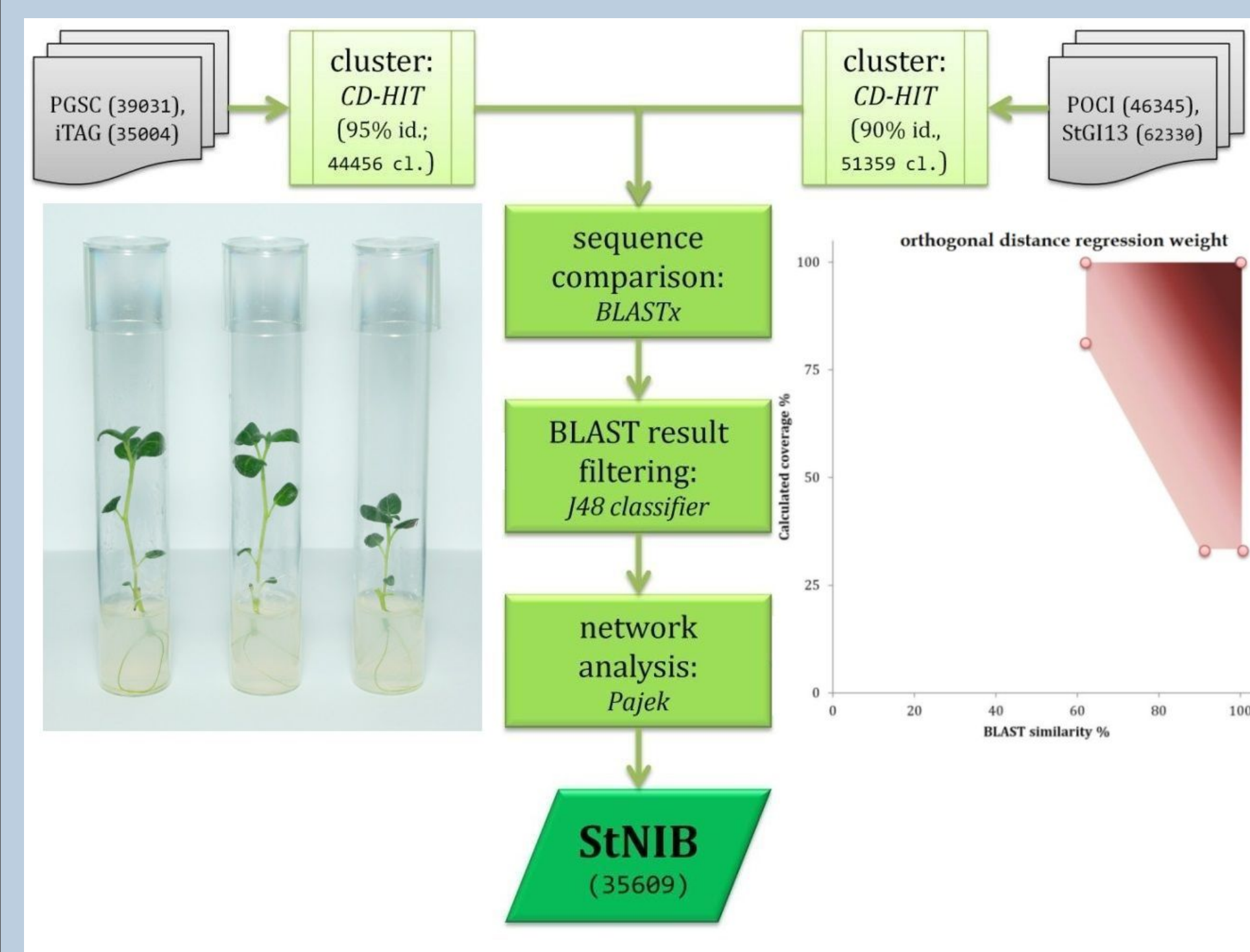
- StNIB-v1 potato pangenome expansion with novel genes and allels

- 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 a acceptable gene model, it does not cover all genes and gene variants present in cultivated potato.

In order to avoid this problems, 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.
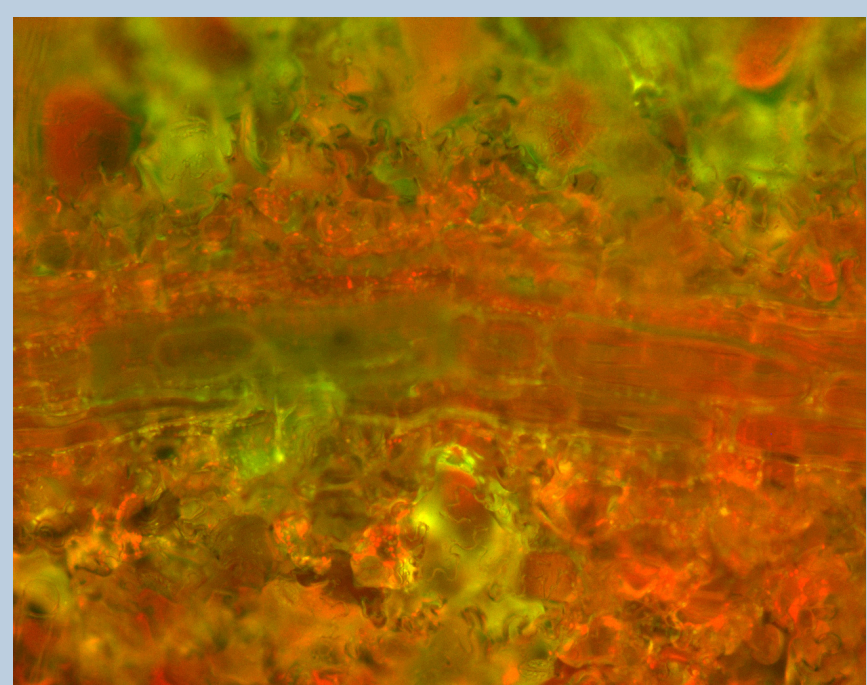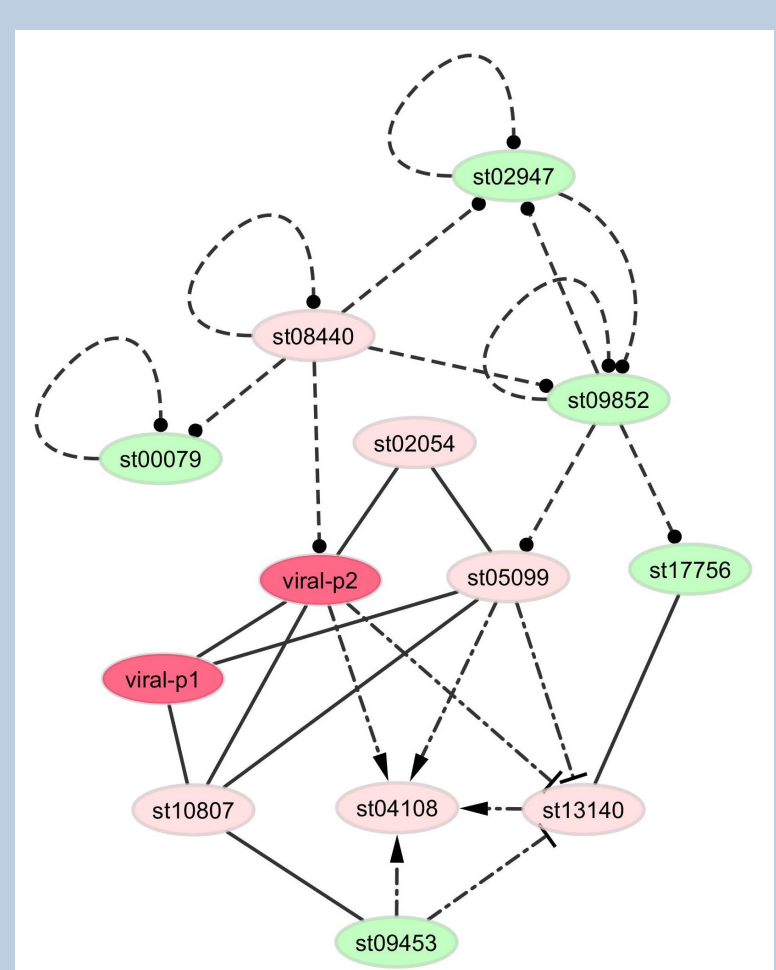

StNIB-v1 potato pangenome

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

Illumina Hiseq DSN PE90 complemented with deposited data

Desiree: 26949735 raw reads
PW: 52171015 raw reads

Hybrid approach

- *CLC Genomics reference-based assembly*
  Desiree: 94111 contigs
  PW: 84092 contigs
- *CLC Genomics de novo assembly*
  Desiree: 9981 contigs
  PW: 11050 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 ($length \notin [200,5000]$)*

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 (weight > 0.00)*
- *missing descriptions assigned with the help of BitScore similarity threshold $[1001,+\infty) \Leftrightarrow$ nearly identical,.. $[50,100] \Leftrightarrow$ very weakly similar*

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 superscaffolds*
  Desiree: 71061 hits
  PW: 38328 hits
- *BLASTn against NCBI-nt*
  Desiree: 912 hits
  PW: 640 hits

Desiree Reference Leaf Transcriptome

- *chimeric sequences → excluded*
- *de Bruijn graph artifacts → excluded*
- *highly repetitive sequences → excluded*
  275221 transcripts and isoforms

PW Reference Leaf Transcriptome

- *chimeric sequences → excluded*
- *de Bruijn graph artifacts → excluded*
- *highly repetitive sequences → excluded*
  218087 transcripts and isoforms

Potato pangenome

59 % genes verified successfully expanded