



# Discovery of SNP markers in Onion and Leek

Thomas van Gorp, Sjaak van Heusden, Gerard van der Linden & Olga Scholten

## Introduction

*Allium* species belong to the plant species with the largest genomes. The sequencing power of the Next Generation Sequencing (NGS) Technology enables comparative genetics studies even in this type of species. The aim of this research project is to discover genetic variation in expressed genes of onion (*Allium cepa*) and two related species *A. roylei* and *A. fistulosum*, and leek (*A. porrum*). The genetic variation will then be used for the development of molecular markers (SNPs), genetic mapping of the genes and identification of disease resistance loci in onion and leek.

## Methodology

Large scale NGS sequencing of expressed genes is performed in breeding material and agronomically important onion and leek cultivars. Genetic variation and SNP markers will be identified using state-of-the-art and newly developed bioinformatics techniques. Comparative genetic analyses using the leek and onion sequences and mapping data will be done to assess the micro- and macrosyteny of these two related crops. The interspecific onion genetic map will be used to find associations between markers and genes for resistance to *Botrytis squamosa* and *Fusarium oxysporum* originating from the two related *Allium* species in a mapping population. Disease screenings will be done in close collaboration with breeding companies and scientists from Wageningen UR Plant Breeding.

## SNP discovery

EST sequences have been obtained from two onion cultivars 'Bravo' and 'Jumbo' and the F1-hybrid between *A. fistulosum* and *A. roylei* (RF parent) using 454-Titanium Flex and Solexa

sequencing. For assembly of these sequences a number of software programmes has been tested and compared. Ideally, the resulting contigs contain sequences of orthologous genes which differ in their sequence on single-nucleotide positions. These can then be used for the development of SNP markers. However, among the SNPs detected, false positives can arise due to paralogous sequences with sufficient sequence homology or errors induced by the sequencing procedures. Both of these artefacts were detected in our data.

In addition to the recently renewed SNP validation framework QualitySNPng, some additional filters were built enabling stringent selection of SNPs, e.g. to filter SNPs within and between genotypes. Currently, a pipeline is under development for obtaining reliable SNPs from Solexa sequencing reads.

At this stage, the pipeline consists of the following filters:

- Pairing unpaired paired-end Solexa-reads
- Duplicate read removal
- QualitySNPng
- Paralog filter
- Detect SNPs within and between genotypes

For the SNPs that have passed the quality checks/filtering process a high throughput SNP detection/genotyping system will be used to validate SNP-markers and to obtain SNP marker data. This data will then be used for the generation of a linkage map using progeny plants of the cross between onion and the RF parent and the cross between two leek parents, as well as for estimating genetic relationships between onion and leek cultivars and wild relatives.



### Wageningen UR Plant Breeding

P.O. Box 16, 6700 AA Wageningen, the Netherlands  
Tel: + 31 317 48 08 71 – Fax: +31 317 41 80 94  
E-mail: olga.scholten@wur.nl

The project is part of TTI-Green Genetics and is a strong collaboration between Wageningen UR Plant Breeding and the Dutch breeding companies: Bejo Zaden (Project-coordinator), ENZA Zaden, Nickerson Zwaan, Nunhems Zaden, Syngenta Seeds and Takii.