**Marian Groenenboom**[1,3]
e-mail: `marian.groenenboom@wur.nl`
**Laura Astola**[1,3]
e-mail: `laura.astola@wur.nl`
**Victoria Choserot**[2,3]
e-mail: `victoria.choserot@wur.nl`
**Jaap Molenaar**[1,3]
e-mail: `jaap.molenaar@wur.nl`
[1] BIOMETRIS, PLANT SCIENCES GROUP, WAGENINGEN UNIVERSITY AND RESEARCH CENTER, WAGENINGEN, THE NETHERLANDS
[2] BIOSCIENCE, PLANT RESEARCH INTERNATIONAL, WAGENINGEN, UNIVERSITY AND RESEARCH CENTER, WAGENINGEN, THE NETHERLANDS
[3] NETHERLANDS CONSORTIUM FOR SYSTEMS BIOLOGY, AMSTERDAM, THE NETHERLANDS

# Glycosylation Networks in Tomato, Top-down and Bottom-up Inference Combined

Tomato (*Solanum Lycopersicum*) is a common element in human diet. In 2009, more than 140000 million tons of tomatoes were produced worldwide. Tomato fruit contains relatively large amounts of flavonoids. Flavonoids have recently gained growing interest due to their anticipated positive health effects as antioxidants. As is the case for many plant metabolites, flavonoids mainly occur in glycosylated form. Although it is widely accepted that glycosylation is of great importance to maintain metabolic homeostasis, the pathway leading to the diverse glycosides, and the specificity of the involved enzymes is not known. In this study, we combine experiments and mathematical modeling to infer the network governing flavonol glycosylation, and study its functioning in vivo.

Tomato seedlings are grown under different conditions, and flavonoid glycoside concentrations are measured for a number of consecutive days. To infer the flavonoid glycosylation network from the resulting time-series, we combine two different approaches. First, we make use of a top-down approach that has as starting point a priori obtained general biological knowledge of molecular reactions and metabolic pathways in plants. This knowledge leads to a number of candidate structures for the network. In a fitting procedure, we estimate the reaction rates in the model, formulated in terms of ordinary differential equations, by applying an iterative minimization method in order to match the model with the observations. The best fitting network is then selected.

In the bottom-up approach one directly infers the network structure from the data via a statistical approach. We explore a method that involves only simple matrix manipulations and standard statistics. In both frameworks we inherently exploit the time-series structure of the data. Because the data are noisy, it turned out difficult to identify the flavonoid network using either the top-down or the bottom-up approach separately. However, by combining both approaches we were able to obtain a reliable estimate of the network model for flavonoid glycosylation in spite of the presence of considerable noise.