

## Using PLS to combine LC-MS and gene expression data to identify genes involved in the metabolism of cyanogenic glucosides in Cassava

**Kasper Brink**

University of Copenhagen, Denmark

Cassava is one of the most important tuber crops in the tropics, serving as the main carbohydrate source in some regions. A major complication for the use of Cassava is the content of cyanogenic glycosides, linamarin and lotaustralin.

Traditional identification of genes involved in the production of cyanogenic glycosides has involved “wet-lab” methods of pathway identification, and genetically altering plant material.

Here we propose to identify these genes in a PLS framework using LC-MS spectra of the metabolites, and gene expression data from an array of 13865 Cassava genes. Data was collected for 32 plants, using three different treatments, added water, added  $\text{KNO}_3$  and a control. The resulting datasets are very large and reduction is required before going further. In particular genes were selected according to p-values for differential expression between treatments, and LC-MS spectra were binned and regions of interest selected. The PLS model was able to make good predictions with 2 components, which also gave the lowest error.

From the PLS coefficients belonging to a given metabolite peak, information about the genes involved in the production of this peak, can be extracted by sorting genes according to numeric coefficients. When comparing results from the PLS models there is good agreement with previously discovered genes in the cyanogenic pathway.

Overall, this method is a fast and computationally simple way to combine several types of data for a better understanding of the underlying networks.

### Biography

Kasper Brink completed his MSc in forestry in 2010. He is now a PhD student in biostatistics/bioinformatics. Main research interests are modeling of high dimensional -omics data, and implementation of new analytical methods.