Validation of a PLS-DA Model for Biomarker Discovery in *Elaeis guineensis* Leaves Related to Fatal Yellowing using UHPLC-MS/MS analysis

Jorge Candido Rodrigues Neto[1,2], Mauro Vicentini Correia[3], Augusto Lopes Souto[1], Jose Antônio de Aquino Ribeiro[1], Clenilson Martins Rodrigues[1], Patrícia Verardi Abdelnur[1,2].
[1]*Brazilian Agricultural Research Corporation, Embrapa Agroenergy, W3 Norte, PqEB, Brasília, DF, Brazil*
[2]*Institute of Chemistry, Federal University of Goiás, Campus Samambaia, Goiania, GO, Brazil*
[3]*Institute of Chemistry, University of Brasilia, Campus Darcy Ribeiro, Brasilia, DF, Brazil*
patricia.abdelnur@embrapa.br

Elaeis guineensis (oil palm) is the richest source of vegetable oil in the world, but is affected by Fatal Yellowing (FY), a condition that can lead to the plant's death. Its cause has been studied since the 70's and so far remains unknown[1]. Metabolomics has been used as an important tool in the biomarkers discovery. Then, in this work, a metabolic fingerprinting approach based on mass spectrometry was applied to *E. guineensis* leaves[2], and data analysis was carried out by chemometric analysis using Partial Least Square Discriminant Analysis (PLS-DA).

Healthy and affected leaves of *E. guineensis* were collected at Marborges Agroindústria S.A. (Mojú, PA, Brazil), in 8 biological replicates, and analyzed in technical triplicates. Metabolites were extracted[3] and analyzed using a UHPLC-ESI-TOF-MS system (Nexera, Shimadzu and Maxis 4G, Bruker Daltonics). Data pre-processing, including feature detection, retention time correction and alignment of metabolites, was performed using XCMS Online. PLS-DA model was constructed and validated using MetaboAnalyst 3.0, using the 29 most important variables previously selected according to the Variable Importance in Projection (VIP) values, pareto scaling and ampicillin as an internal standard for mass normalization.

Figure 1A shows the VIP scores plot that indicates the most significantions according to PLS-DA of healthy and affected leaves dataset. To prevent model overfitting, permutation test, using group separation distance, with 2000 interactions were performed (Figure 1B). These tests shows that none of the permuted models were as good as original model, so the original model assignment is appropriate. A cross validation was performed using a 10-fold CV method and showed good accuracy, with R^2 of 0.944 and Q^2 of 0.904 for the 3 principal components.

Some metabolites (Figure 1A), as organic acids (chelidonic acid: *m/z* 185.008), alkaloids (glycerylphosphorylcolin: *m/z* 258.109) and flavonoids (schaftoside / isoschaftoside: *m/z* 565.158) were assigned using fragmentation patterns (MS/MS experiments) and metabolic databases comparison (ChEBI, KEGG). These metabolites will be further used to correlate metabolic pathways in the attempt to elucidate the cause of fatal yellowing.

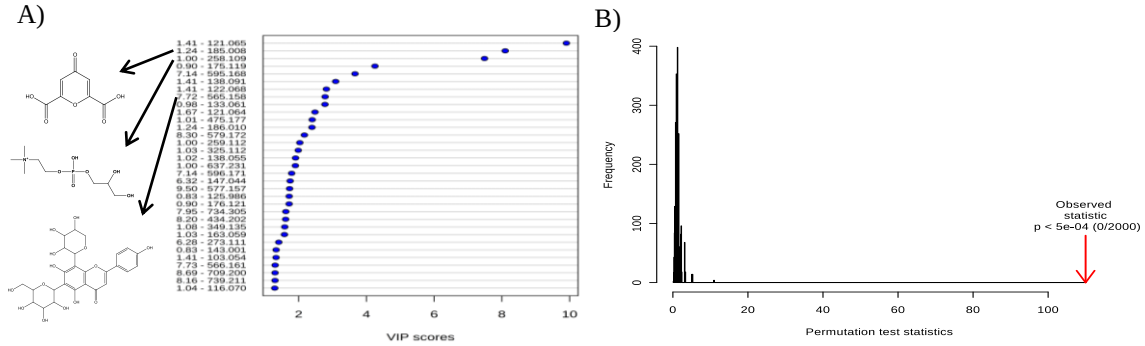

Figure 1. VIP scores plot of potential FY biomarkers and PLS-DA cross validation.

1. Elliot, M. L. Bud Rot of Palm. Electronic Data Information Source of UF/IFAS (Series), 2015.
2. Vargas, *et al*. Metabolomics analysis of oil palm (*Elaeis guineensis*) leaf: evaluation of sample preparation steps using UHPLC–MS/MS. Metabolomics, v.12, p.10, p.153, 2016.
3. Giavalisco, *et al*. Elemental formula annotation of polar and lipophilic metabolites using 13C, 15N and 34S isotope labelling, in combination with high-resolution mass spectrometry. The Plant Journal, v.68, n.2, p.364-376, 2011.