

## Supporting Information

### *Brachiaria plantaginea* as a Potential (New) Source of Shikimic Acid.

#### Quantification by NIR and PLS Regression

Maria Cristina A. Costa<sup>1</sup>, Marcus B. Matallo<sup>2</sup>, Márcia M. C. Ferreira<sup>1</sup>, Sonia C. N. Queiroz<sup>3</sup>, Sydnei Dionísio B. de Almeida<sup>2</sup>, Daniel A. de Siqueira Franco<sup>2</sup>

#### Affiliations

<sup>1</sup>Theoretical and Applied Chemometrics Laboratory (LQTA), Institute of Chemistry, University of Campinas - Unicamp, Campinas - SP, Brazil

<sup>2</sup>Instituto Biológico, Jardim das Palmeiras, Campinas - SP, Brazil

<sup>3</sup>Embrapa Meio Ambiente, Jaguariúna - SP, Brazil

#### Correspondence

*Dr. Maria Cristina Andrezza Costa*

Theoretical and Applied Chemometrics Laboratory (LQTA)

Institute of Chemistry

University of Campinas – Unicamp

P.O. Box 6154

Campinas - SP, 13084-971

Brazil

Fax: +55 19 3521 3023

[cristina.costa@iqm.unicamp.br](mailto:cristina.costa@iqm.unicamp.br)

**Table 1S** Comparison between the measured values of shikimic acid concentration ( $\mu\text{g} \times \text{g}^{-1}$ ) and the predicted ones by PLS regression with seven factors on the complete set of forty samples, including training and validation sets. Bias for the test set is -10.55, showing a tendency of overestimating the shikimic acid concentration.

Sample	Measured Y	Predicted Y	Residual Y	RE%
smp1*	523.69	552.31	-28.62	-5.46
smp2	502.73	430.79	71.94	14.31
smp3	785.93	758.07	27.86	3.54
smp4*	739.88	685.31	54.57	7.38
smp8	755.46	723.74	31.72	4.20
smp9	1027.88	936.16	91.72	8.92
smp11	475.33	577.70	-102.37	-21.54
smp12	1062.01	1091.43	-29.42	-2.77
smp13*	2185.90	1908.65	277.25	12.68
smp14	1922.15	1954.76	-32.61	-1.70
smp15	1059.35	1085.34	-25.99	-2.45
smp16	3592.45	3561.94	30.51	0.85
smp17	2703.11	2674.46	28.65	1.06
smp18	3273.33	3318.27	-44.94	-1.37
smp19	1983.19	2059.83	-76.64	-3.86
smp20*	1582.40	1679.42	-97.02	-6.13
smp21*	1771.15	1795.97	-24.82	-1.40
smp22*	1547.79	1698.19	-150.40	-9.72
smp23*	1852.87	1610.54	242.33	13.08
smp24	2635.18	2508.84	126.34	4.79
smp25	745.60	970.26	-224.66	-30.13
smp26	795.89	865.47	-69.58	-8.74
smp27	590.95	515.18	75.77	12.82
smp28*	793.40	1044.22	-250.82	-31.61

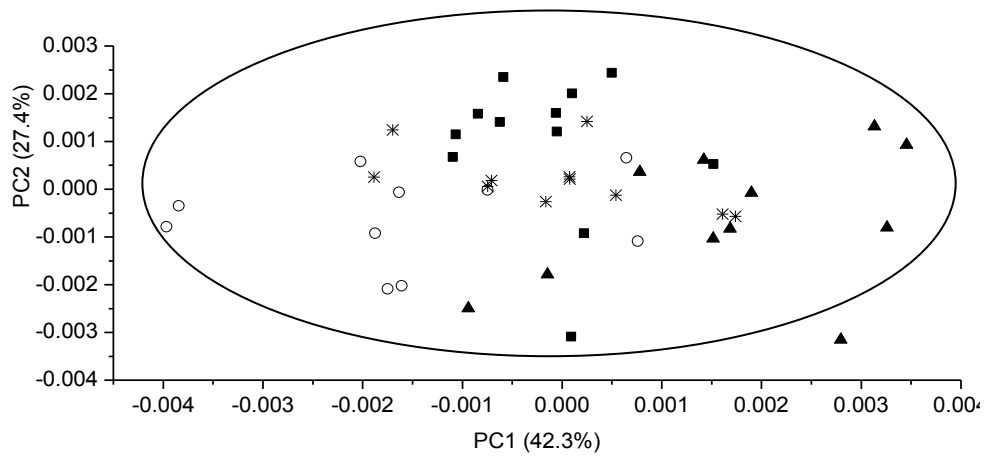
smp30	760.73	741.67	19.06	2.50
smp31	1114.25	1120.95	-6.70	-0.60
smp32	924.46	828.75	95.71	10.35
smp35	333.90	365.71	-31.81	-9.53
smp36	452.00	381.23	70.77	15.66
smp37	658.52	650.36	8.16	1.24
smp38	833.23	923.55	-90.32	-10.84
smp39	1093.28	1080.51	12.77	1.17
smp40	477.23	454.90	22.33	4.68
smp41*	501.43	449.33	52.10	10.39
smp42	646.92	632.48	14.44	2.23
smp43	881.58	817.94	63.64	7.22
smp45*	814.13	973.23	-159.10	-19.54
smp46	669.43	735.86	-66.43	-9.92
smp47*	761.50	793.00	-31.50	-4.14
smp48	674.89	664.81	10.08	1.49

\*Test set with 11 samples used in the external validation. RE% is the relative error.

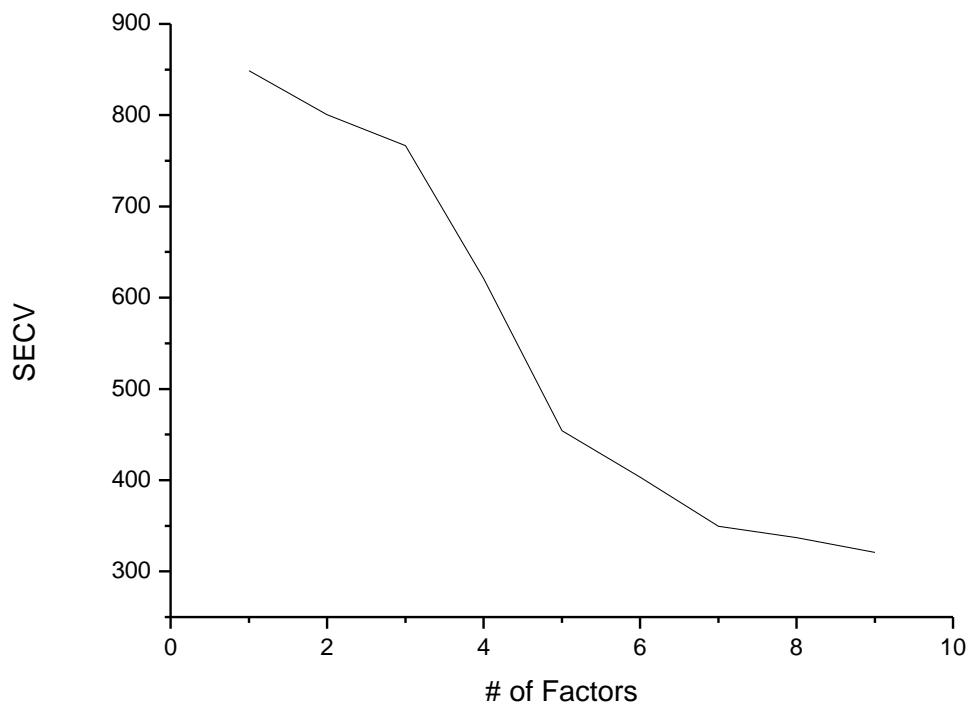
**Table 2S** Equations for validation parameters employed in PLS regression.

Parameters <sup>a</sup>	Equations
R <sup>2</sup>	$1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$
SEC	$\sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - k - 1}}$
SECV	$\sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$
SEP	$\sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i - bias)^2}{n - 1}}$
RE%	$\frac{(y_i - \hat{y}_i)}{y_i} 100$
RSD%	SEP × 100/mean <sup>b</sup>

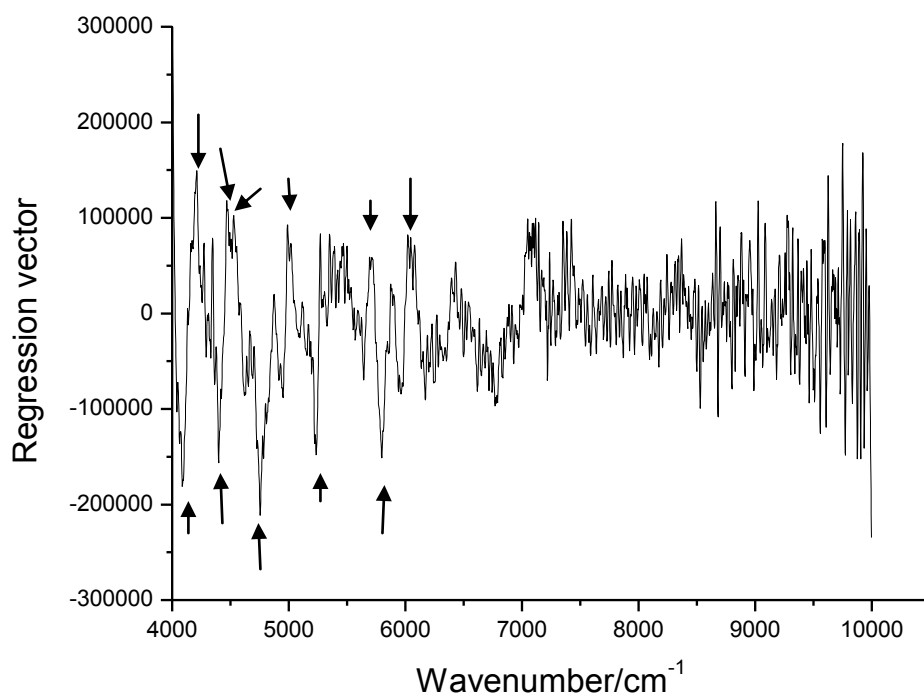
<sup>a</sup>R<sup>2</sup> (coefficient of determination), SEC (standard error of calibration), SECV (standard error of cross validation), SEP (standard error of prediction), RE% (relative error), RSD% (relative standard deviation); <sup>b</sup>mean of the reference values in the validation set.



**Fig. 1S** PC1 vs. PC2 scores plot of mean-centered NIR spectra of samples collected at 3 (■), 6 (○), 9 (\*), and 12 (▲) days.



**Fig. 2S** SECV versus number of factors plot. The optimum number of factors is indicated by the local minimum in the curve.



**Fig. 3S** The regression vector obtained by the final model of PLS regression. The positive signals at 4200, 4470, and 4430 and the negative signals at 4080, 4400, and 4760  $\text{cm}^{-1}$  correspond to the combination bands of lignin (C-H stretching). The peaks at 4760 (negative), 4990, and 5230 (positive) are in the region of combination bands of O-H stretching found in the structure of shikimic acid. The signal at 5230 can also be attributed to the C=O stretch of carboxylic acids. Other signals, positive at 5710 and 6050 and negative at 5800 are attributed to the C-H stretching first overtone.