AUTHENTICITY STUDY OF *Phyllanthus* SPECIES BY NMR AND FT-IR TECHNIQUES COUPLED WITH CHEMOMETRIC METHODS[#]

Maiara S. Santos*, Edenir R. Pereira-Filho e Antonio G. Ferreira
Departamento de Química, Universidade Federal de São Carlos, CP 676, 13560-970 São Carlos – SP, Brasil
Elisangela F. Boffo
Instituto de Química, Universidade Federal da Bahia, 40170-115 Salvador – BA, Brasil
Glyn M. Figueira
Centro Pluridisciplinar de Pesquisas Químicas, Biológicas e Agrícolas, Universidade Estadual de Campinas, CP 6171, 13081-970 Campinas – SP, Brasil

MODELS OPTIMIZATION

FT-IR data

The parameters used in the models optimization were:

- in the KNN model, K = 1 was selected because with this value there are no prediction errors;
- in the SIMCA model, two PCs were selected for all categories because they obtained more than 85% of information from the data analyzed in all classes: 93.4% for *P. amarus,* 85.0% for *P. caroliniensis,* 88.4% for *P.niruri,* 87.7% for *P. tenellus* and 91.6% for *P. urinaria.*

The PLS-DA loadings for the calibration models were similar to those observed in the PCA analysis. In this model, 5 PCs were used for the *P. amarus, P. niruri, P. tenellus* and *P. urinaria* classes, whereas 4 PCs were used for the *P. caroliniensis* with SEC, SEV and PRESS Val less than 0.153, 0,211 and 1.787, respectively, and R^2 greater than 0.854. The calibration statistics indicated that the model developed could be acceptable to classify new samples.

^{1}H HR-MAS NMR data

The parameters used in the models optimization were:

- in the KNN method, seven prediction errors were obtained with K = 1. These errors are due to the proximity between different classes;
- in the SIMCA model, 5 PCs were selected for the *P. amarus* class (96.5%), 4 PCs were used for the *P. caroliniensis* (95.6%) and *P. niruri* (86.8%) classes, whereas three PCs were used for the *P. tenellus* (84.3%) and *P. urinaria* (86.6%) classes.

Considering the PLS-DA model, 3 PCs were used for the *P. amarus,* 4 for the *P. caroliniensis* and *P. niruri,* whereas 6 PCs were used for *P. tenellus* and *P. urinaria* classes with SEC, SEV and PRESS Val less than 0.118, 0.151 and 1.202, respectively, and R^2 greater than 0.756.

Liquid state NMR - aqueous extracts

The parameters used in the models optimization were:

- in the KNN methods, K = 1 was selected because with this value there are no prediction errors (sets A and B).
- in the SIMCA method from set A, 3 PCs were used for the *P. amarus* (85.3%), *P. caroliniensis* (86.0%) and *P. tenellus* (86.8%) classes and 2 PCs were used for the *P. niruri* (79.4%) and *P. urinaria* (78.2%) classes.
- for the SIMCA method from set B, 4 PCs were used for the *P. amarus* (78.3%) and *P. tenellus* (77.8%) classes, whereas for the *P. caroliniensis* (86.3%)*, P. niruri* (76.6%) and *P. urinaria* (83.5%) classes 3 PCs were used.

Considering the PLS-DA model from set A, 3 PCs were used for the *P. amarus* class, 5 PCs were used for the *P. caroliniensis* and *P. urinaria* classes and 2 PCs were used for the *P. tenellus* and *P. niruri* classes with SEC, SEV and PRESS Val less than 0.072, 0.101 and 0.410, respectively, and R^2 greater than 0.973.

In the PLS-DA model from set B, 4 PCs were used for the *P. amarus, P. caroliniensis, P. tenellus* and *P. niruri* classes and 6 PCs were used for the *P. urinaria* class with SEC, SEV and PRESS Val less than 0.081, 0.135 and 1.347, respectively, and R^2 greater than 0.923.

RESULTS

Table 1S. ^{1}H and ^{13}C NMR data for compounds in the *Phyllanthus* aqueous extract

Compound	δ ^{13}C	δ ^{1}H (multiplicity, *J* in Hz)
α-Glucose (α-Glu)		
C^1H	95.0	5.24 (*d*, 3.7)
C^2H	74.0	3.55 (*m*)
β-Glucose (β-Glu)		
C^1H	98.8	4.65 (*d*, 7.9)
C^2H	77.1	3.25 (*m*)
C^3H	72.1	3.50 (*m*)
C^4H	72.6	3.41 (*m*)
Sucrose (Suc)		
C^1H	95.1	5.42 (*d*, 3.8)
C^2H	74.1	3.57 (*m*)
C^3H	75.5	3.78 (*m*)
C^4H	72.5	3.48 (*m*)
C^6H_2	63.0	3.86 (*m*)
$C^{1'}H_2$	64.2	3.68 (*m*)
$C^{2'}H$	106.6	---
$C^{3'}H$	79.3	4.22 (*d*, 8.7)
$C^{4'}H$	75.3	4.07 (*m*)
C^5H	84.3	3.90 (*m*)
Alanine		
α-CH	---	3.79 (*m*)
β-CH_3	19.1	1.48 (*d*, 7.2)
Valine		
γ-CH_3	---	1.00 (*d*, 7.0)
γ'-CH_3	19.3	1.05 (*d*, 7.0)
β-CH	35.4	2.28 (*m*)
Threonine		
α-CH_2	---	3.52 (*m*)
γ-CH_3	24.0	1.33 (*d*, 6.6)
β-CH	---	4.27 (*m*)
4-aminobutiric acid		
γ-CH_2	35.5	3.02 (*t*, 7.5)
α-CH_2	26.7	2.32 (*m*)
β-CH_2	42.6	1.94 (*m*)

Abbreviations: *d* – doublet, *t* – triplet, *m* – multiplet

*e-mail: maiarassantos@yahoo.com.br

[#]Artigo em homenagem ao Prof. Otto R. Gottlieb (31/8/1920-19/6/2011)

Table 2S. ^{1}H, ^{13}C and *g*HMBC NMR data for phyllanthin

Position	δ ^{1}H (multiplicity, *J* in Hz)	δ ^{13}C	^{1}H-^{13}C *g*HMBC*	Literature[37]	
				δ ^{1}H	δ ^{13}C
1 (1')	-	136.1	-	-	135.2
2 (2')	6.56 (*d*, 1.96)	113.8	7 (7'); 6 (6'); 3 (3')	6.59	112.2
3 (3')	-	150.3	-	-	148.7
4 (4')	-	148.6	-	-	147.0
5 (5')	6.78 (*d*, 8.02)	112.9	4 (4'); 1 (1'); 6 (6')	6.73	111.0
6 (6')	6.59 (*dd*, 8.02; 1.96)	122.6	3 (3'); 1 (1'); 7 (7'); 5 (5')	6.61	121.0
7 or (7')	2.56 (*dd*, 15.7; 7.25)	36.0	1 (1'); 2 (2'); 5 (5'); 9 (9'); 8 (8')	2.59	34.9
	2.58 (*dd*, 15.7; 7.25)			2.66	
8 (8')	1.98 (*m*)	41.9	1 (1'); 6 (6'); 9 (9'); 7 (7')	2.01	40.7
9 (9')	3.29	74.0	7 (7'); 8 (8'); 9 (9')-MeO	3.25	72.8
	3.41			3.28	
3 (3')-MeO	3.70 (*s*)	56.3	3 (3')	3.78	55.9
4 (4')-MeO	3.79 (*s*)	56.5	4 (4')	3.82	55.7
9 (9')-MeO	3.30 (*s*)	59.0	9 (9')	3.27	58.7

Abbreviations: *s* – singlet, *d* –doublet, *dd* – doublet of doublet, *m* – multiplet. **g*HMBC data set: the numbers correspond to the correlated carbons.

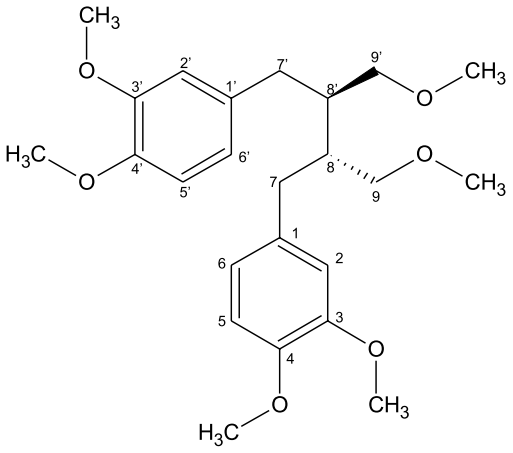

Figure 1S. *Phyllanthin structure*

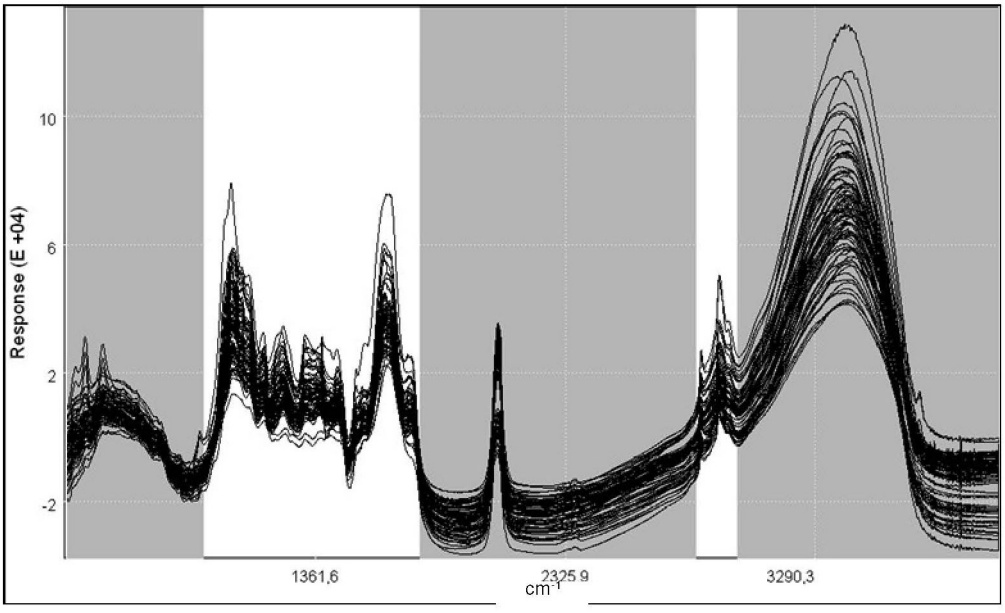

Figure 2S. *FT-IR spectra of all samples analyzed, showing selected regions used in statistical analyses (in white)*

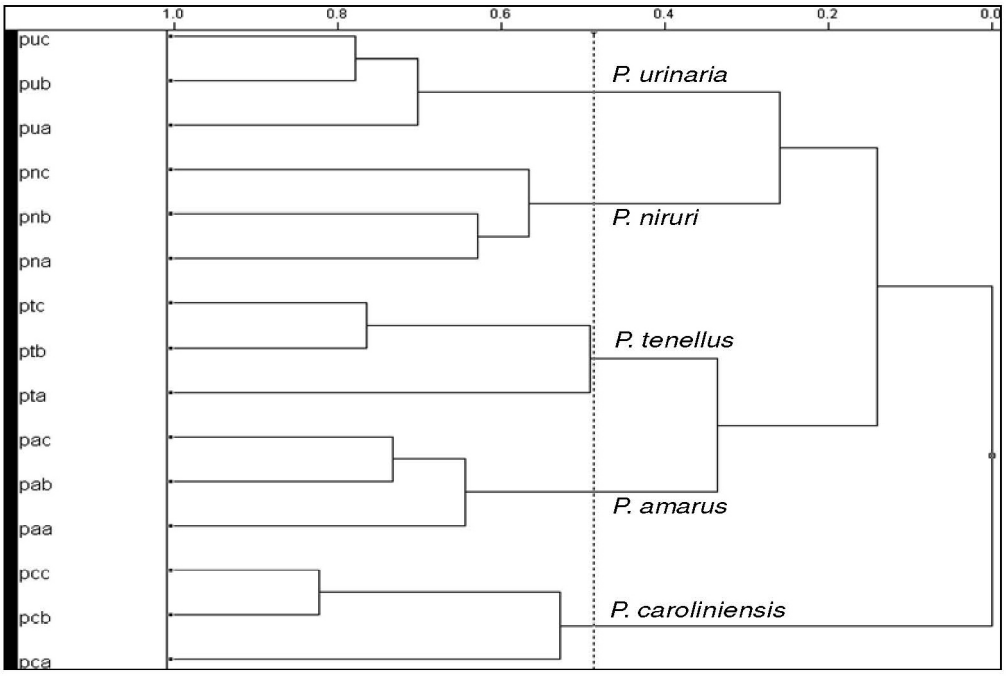

Figure 3S. *HCA dendrogram obtained from FT-IR data of five standard samples of Phyllanthus species (similarity index: 0.487)*

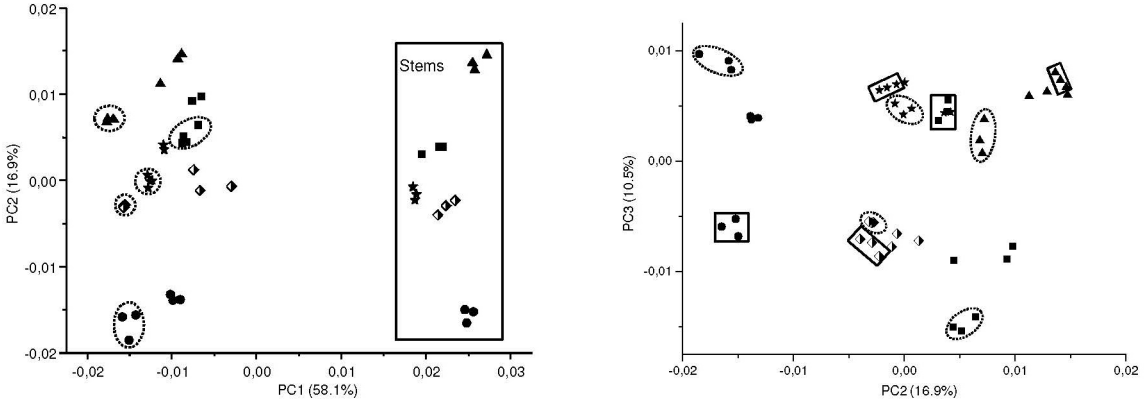

Figure 4S. *PCA score plots of five standard samples of Phyllanthus species (aerial parts, leaves and stems separately) analyzed by FT-IR: (a) PC1 x PC2 (b) PC2 x PC3. The samples composed of only leaves were circled with a dashed line and samples composed of only stems were circled with squares*

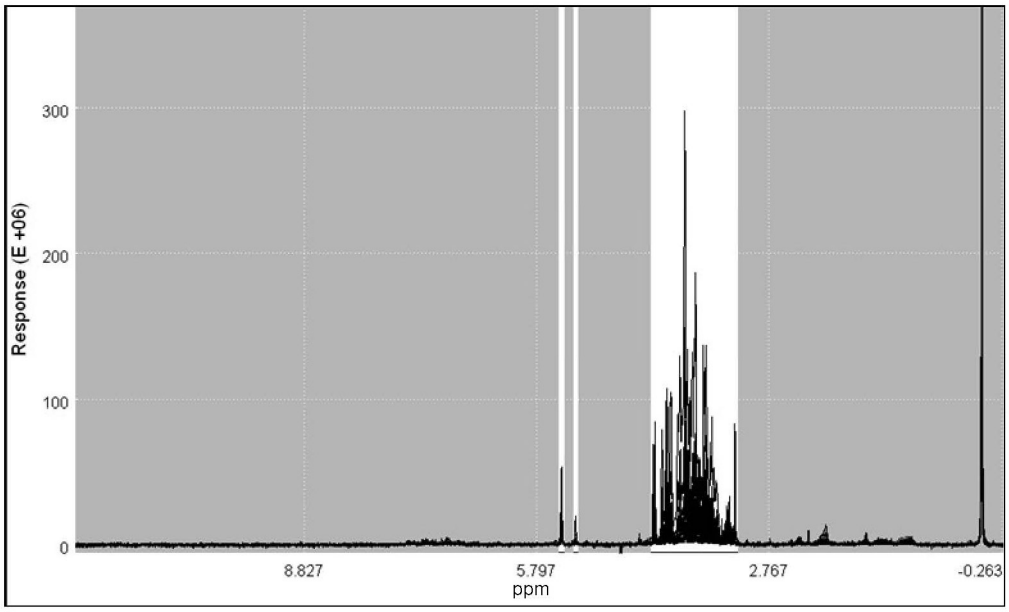

Figure 5S. *[1]H HR-MAS NMR spectra of all samples analyzed, showing the selected regions used in the statistical analysis (in white)*

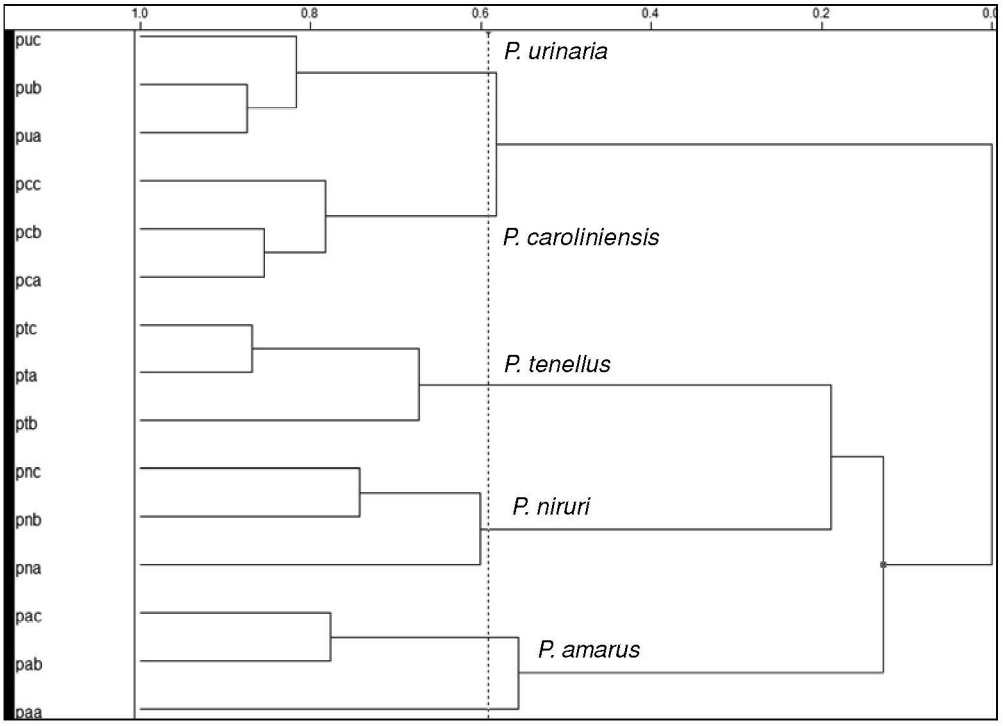

Figure 6S. *HCA dendrogram obtained from [1]H HR-MAS NMR data of five standard samples of Phyllanthus species (similarity index: 0.591)*

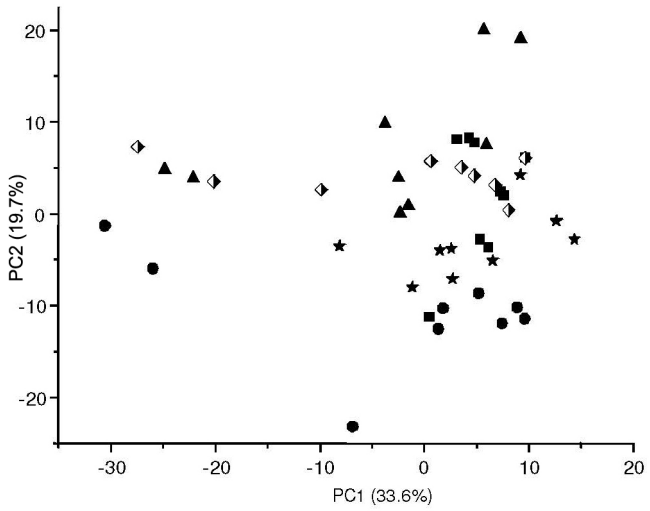

Figure 7S. *PCA score plots of five standard samples of Phyllanthus species (aerial parts, leaves and stems separately) analyzed by [1]H HR-MAS NMR*

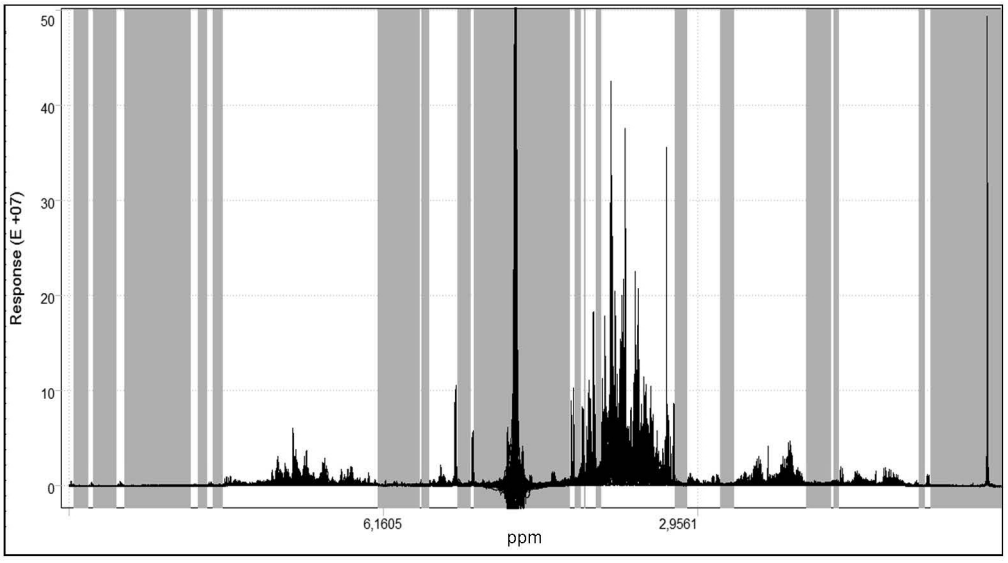

Figure 8S. *[1]H NMR spectra of all samples analyzed (aqueous extracts), showing the selected regions used in the statistical analyses (in white)*

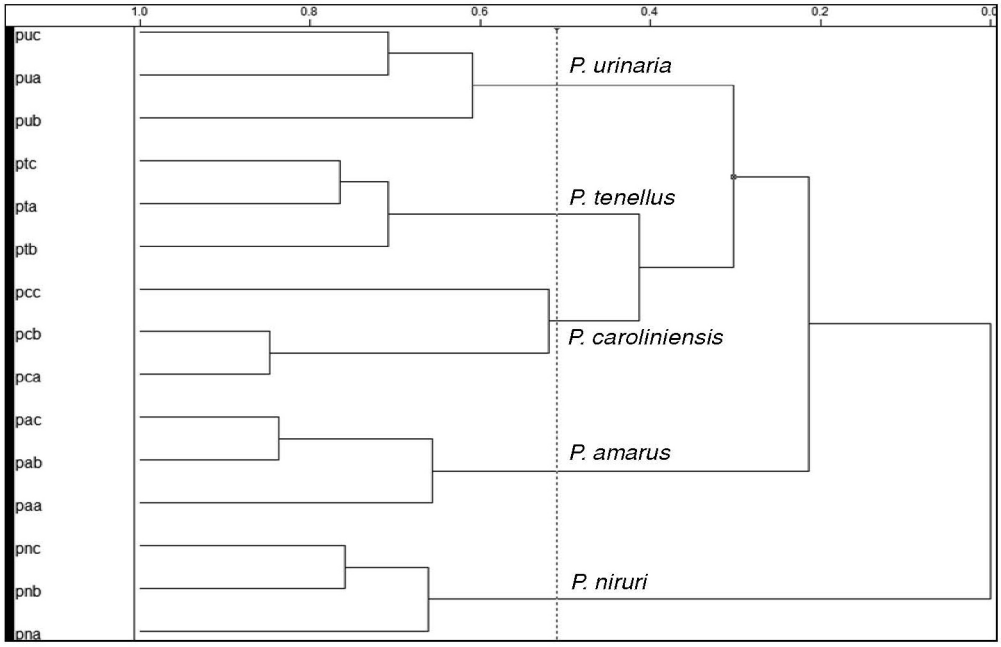

Figure 9S. *HCA dendrogram obtained from* ^{1}*H NMR (aqueous extracts) data of five standard samples of Phyllanthus species (similarity index: 0.510)*

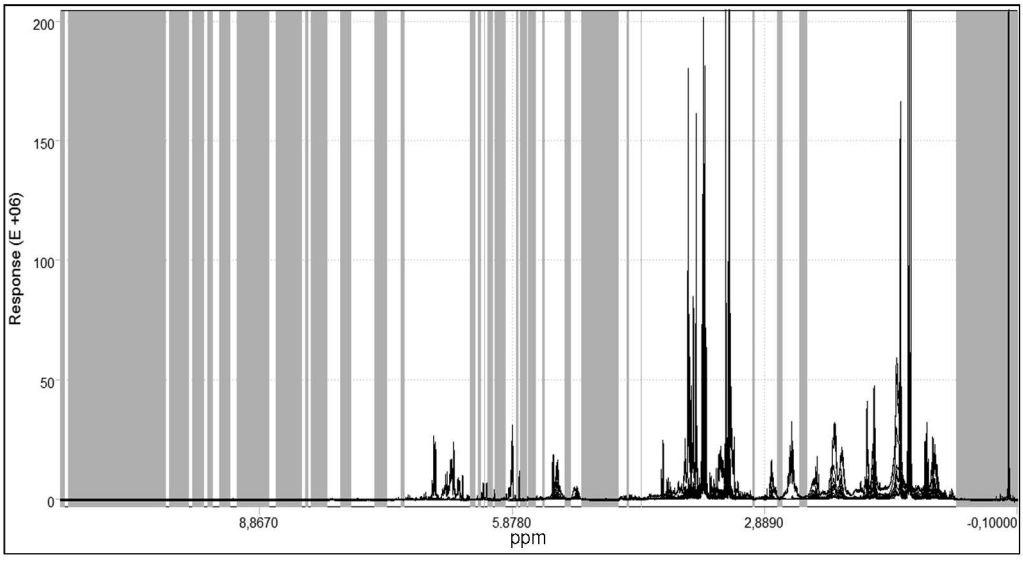

Figure 10S. ^{1}*H NMR spectra of all samples analyzed (ethanolic extracts), showing the selected regions used in the statistical analysis (in white)*