Contribution of near infrared spectroscopy to the assessment of the origin of Amazonian leaves rich in bioactive compounds

V. Baeten,[a,*] E. Martins da Silva,[b] Y. Larondelle,[c] P. Dardenne[a] and H. Rogez[b]

[a]*Walloon Agricultural Research Centre (CRA-W). Chaussée de Namur 24, B-5030 Gembloux, Belgium. E-mail: baeten@cra.wallonie.be*
[b]*Faculdade de Engenharia de Alimentos, Universidade Federal do Pará (UFPA), av. Perimetral s/n, 66.095-780, Belém-PA, Brazil*
[c]*Unité de biochimie de la nutrition, Université catholique de Louvain, Croix du Sud, 2/8, 1348 Louvain-la-Neuve, Belgium*

Introduction

The production of plant extracts rich in new bioactive compounds is a sustainable strategy to give added value to the secondary Amazonian forest largely present in Brazil. Moreover, it will avoid more destruction and fragmentation of the primary vegetation. A secondary forest or "*capoeira*" is a vegetation which (i) was formed following the total anthropogenic destruction (with more than 90% destruction) of the primary forest, (ii) was established on a great surface that presents a structure, species of trees and a dynamics different from the initial settlement, and (iii) that did not yet reach its initial state. It is essential to mention that approximately 30% of the deforested areas in the Amazonian basin are recovered by *capoeira*, predominantly in the North of the State of Pará. There is an urgent need to find new ways of value-assessment of the secondary forest in order to provide additional income for the traditional population, and for the small farmers. Innovative Non-Timber Forest Products (NTFP) should be developed on the basis of the large biodiversity of this ecosystem. In order to analyse a high throughput of samples, analytical strategies based on rapid methods have to be set up. The aim is to check the quality of the raw materials and of the final products, as well as to assess the species origin of the processed products. This tentative study aims at defining the key parameters to take into account in the calibration of NIR spectrometers for the discrimination, at the leaf level, of the species origin of a sample.

Materials and methods

A total of 200 leaves obtained from 20 samples were analysed by near infrared (NIR) spectroscopy. Two species from the secondary forest, and known to be rich in bioactive compounds, were considered: *Inga edulis* (14 samples) and *Byrsonima crassifolia* (6 samples). The fresh leaves were dried and put into plastic bags at the UFPA (Belém, Brazil) and then analysed at the CRA-W (Gembloux, Belgium). For each sample, 10 intact leaves were analysed with the Phazir (Model 1624) instrument from Polychromix (Wilmington, USA). Three NIR spectra of each 200 individual leaves were collected at three different parts of the leaf. A data base of 600 spectra was constructed including 420 from *I. edulis* leaves and 180 from *B. crassifolia* leaves. Data treatment was performed using the Unscrambler v9.2. Spectra were pre-processed using smoothing and 1st derivative (Savitsky-Golay algorithm).

Results

Figure 1 presents the pretreated spectra of the two species respectively.

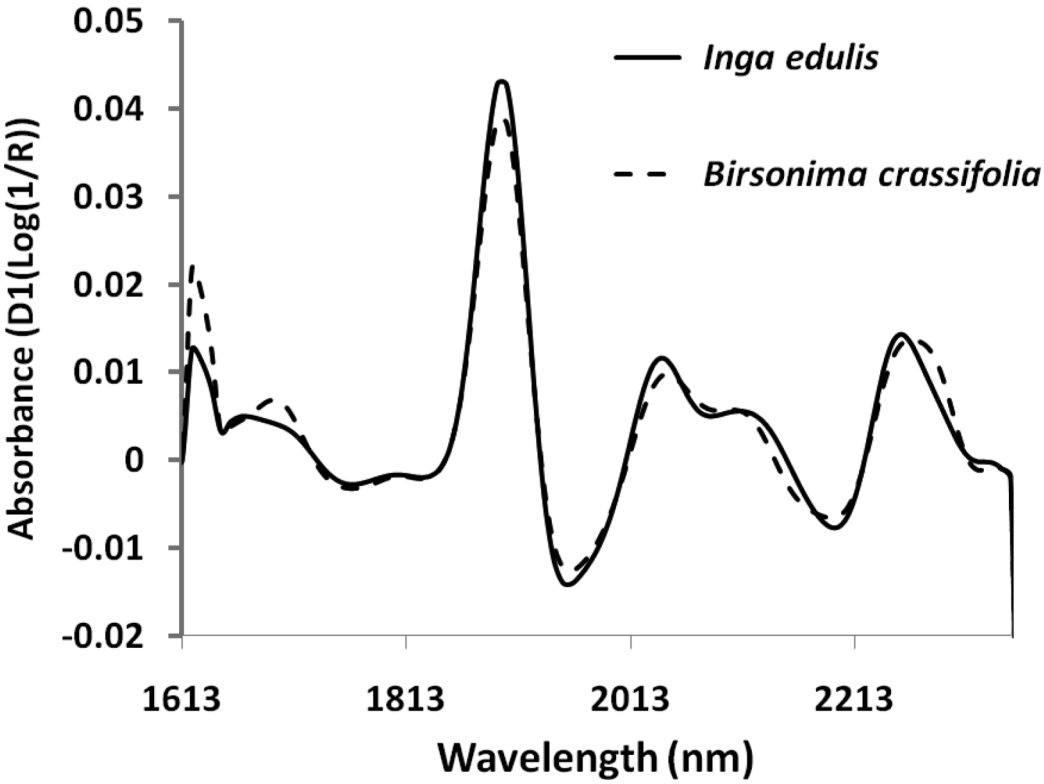

Figure 1. First derivative of *I. edulis* (solid line) and *B. crassifolia* (dashed line).

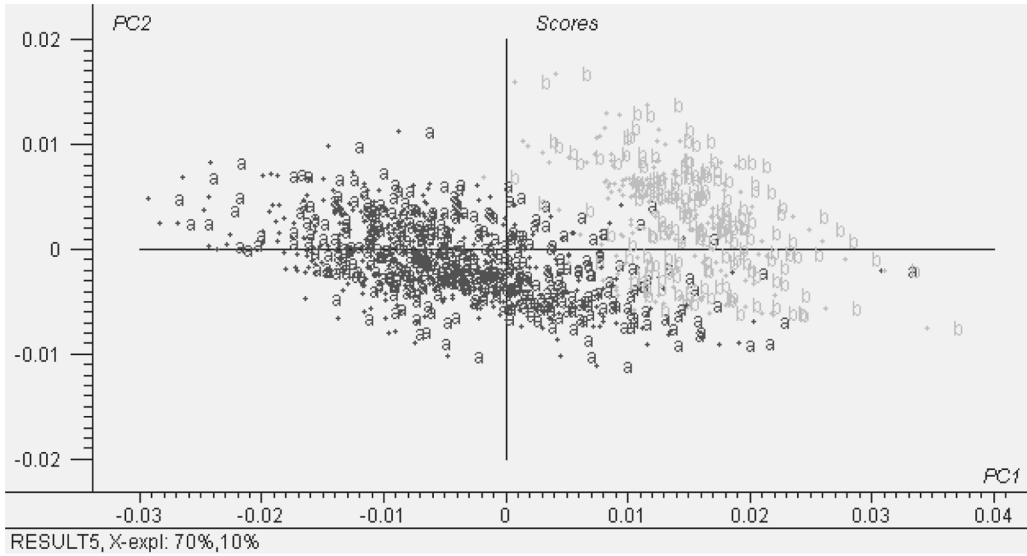

Figure 2. PCA analysis results. PC1 vs PC2 score plot of the 600 leaves analysed. In blue the *I. edulis* leaves and in green the *B. crassifolia* leaves.

PCA analysis of the NIR spectra from individual leaves showed the possibility to discriminate the samples according to their botanical origin (Figure 2).

PLS was used to construct models to predict the origin of a leaf from an unknown sample. Several strategies were tested to calibrate the spectrometer. Table 1 summarises some of these strategies.

The aim was also to define the number of samples from each plant variety to include in the future calibration set, the number of leaves from each sample to analyse and the number of spectra to collect from each leaf. The different strategies used part of the spectra in the calibration stage (between 3 and 180 spectra), while 420 spectra were used for the validation stage. The strategy C based on 9 spectra from 3 different leaves for the calibration provided a model with equivalent efficiency to models including 180, 90 or 30 spectra. Figure 3 presents the prediction results of the model issued from strategy C obtained on the 420 spectra included in the test set.

Conclusion

This work is a first step to specify the optimised strategy in order to construct a data-base to discriminate at the reception plant the origin of raw material rich in bio-active compounds. It has demonstrated the potential of NIR to discriminate two kind of dried leaves rich in bioactive compounds, and to determine the best compromise for the construction of discriminant equations.

Table 1. Results of the different strategies (A–G) tested to calibrate the Phazir instrument for the detection of the origin of unknown leaves.

Strategy	Calibration set			Validation set	PC in the model	*r*	*RMSEP*	Bias	Misclassifcation
	Number of leaves	Number of spectra by leaves	Total number of spectra	Total number of spectra					
G	60	3	180	420	2	0.97	0.23	–0.05	0
A	30	3	90	420	2	0.98	0.19	–0.015	0
B	30	1	30	420	2	0.98	0.2	–0.02	0
C	3	3	9	420	2	0.98	0.19	–0.14	0
D	9	1	9	420	2	0.94	0.43	–0.14	2
F	3	1	3	420	2	0.87	0.61	0.1	23

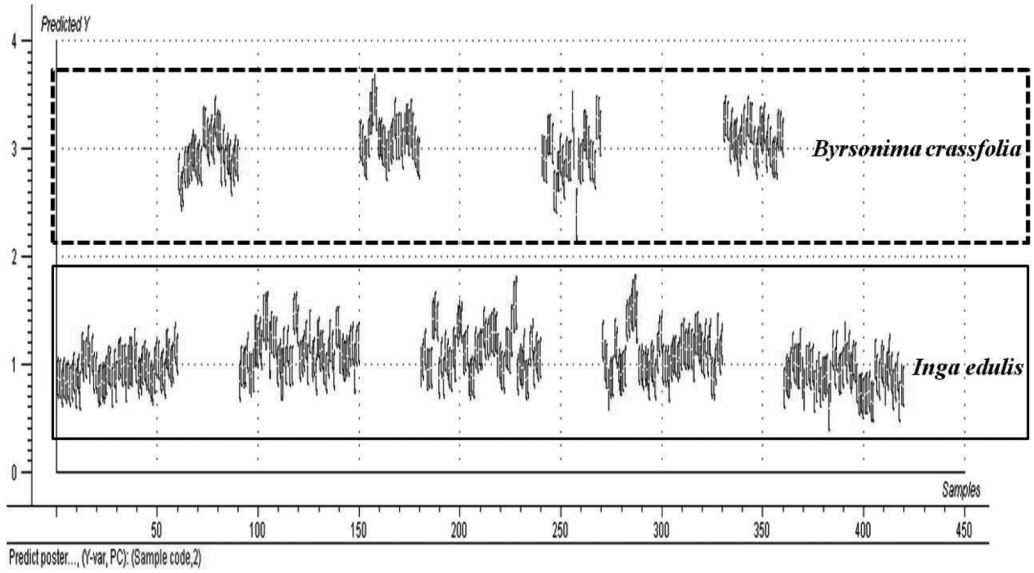

Figure 3. Prediction results obtained on the 420 spectra included in the test set for the model issued from the strategy C.

Acknowledgement

This work has received a financial support through CNPq (Brazil) and FNRS-FRS (Belgium). Part of the research leading to these results has been performed in the framework of Walnut project from the Walloon Region. The authors thank the CRA-W, UCL and UFPA teams for their technical and scientific support in the realisation of this work.