NIRS (Near Infrared Reflectance Spectroscopy) Potential for Breeding Aroids

Vincent Lebot, CIRAD, PMB 946, Port-Vila, Vanuatu

ABSTRACT:

Aroids corms and cormels are chemically very variable and there is significant variation of their major constituents (starch, sugars, cellulose, proteins, minerals) between genotypes. A fairly common difficulty for breeding programmes is to assess precisely these compounds as chemical analyses are too expensive for routine screening. These programmes are often based on mass recurrent selection and great numbers of hybrids have to be screened to achieve some progress. However, the wrong selection of a parent can cause a serious constraint to the development of new varieties for processing purposes. Likewise, if the table quality is not acceptable for consumers, years of intense and expensive breeding efforts can lead to complete failure. Low-cost methods for rapid evaluation of numerous hybrids are urgently needed. The present paper, assess the potential of NIRS (Near Infrared Reflectance Spectroscopy) as an alternative method for predicting these major constituents and evaluating corms and cormels quality. Models have been developed using data from 642 root crops accessions and their predictive potential has been tested on 100 varieties and hybrid lines of taro, *Colocasia esculenta*. The NIRS calibration results for major constituents, and their practical applications for aroids breeding and genetic improvement of quality traits, are discussed.

Keywords: Aroids, breeding, corms, chemical composition, major constituents, NIRS, quality

Introduction

Aroids are foods of great cultural and economic importance throughout the humid tropics. In many countries, there is a desire to strengthen production by breeding for improved nutritional quality and agronomic performance. All species (*Alocasia macrorrhyzos, Amorphophallus campanulatus, Colocasia esculenta, Cyrtosperma chamissonis, Xanthosoma sagittifolium*) are vegetatively propagated and are highly polymorphic. Morphological characterisation of germplasm collections is usually conducted on hundreds of accessions, each represented by few plants. Once the promising genotypes are identified, agronomic evaluation involves randomized field designs which allow the study of only a limited number of clones and genotypes. Their results are often biased by the heterogeneity of the clonal material (headsets, corms, cormels, suckers) needed to compose a plot treatment.

Breeding programmes start with the evaluation of germplasm followed by crossing the selected accessions. Crosses can be done in polycross plots where natural open pollination is used to produced half-sib progenies and when bulked together, populations for recurrent selection. Crosses are then conducted between parents chosen on their performance *per se* and presenting complementary traits. Controlled crosses done and aroid inflorescences can produce numerous full-sib progenies, often with several hundreds individuals per family. Controlled pollination allows the production of numerous seedlings but seedling growth is very slow and it takes about four years for a new genotype to be properly evaluated because of the time needed to produce enough homogenous clonal propagules for accurate agronomic comparison (Ivancic and Lebot, 2000). Once seedlings are obtained, they must be clonally propagated and multiplied through clonal generations before

accurate assessment of their performance can be conducted. The corm yield of the taro plant, for example, is correlated to the weight of the propagule planted and genotype performance is therefore, difficult to assess in F1s (seminal generation) or even the first (C1) or second (C2) clonal generations. It is necessary to reach a propagule weight comparable to those used in traditional cultivation before an accurate assessment can be made. The screening of large populations for major traits is consequently, a laborious and expensive operation which leads taro breeders to develop visual tools to speed up the process (Ivancic *et al*., 2003) and these traits are efficient for most agro-morphological traits. The significant genotype × environment interactions are limiting the relevance of complex field layouts and adding complexity to the evaluation process. Simple methodological alternatives are therefore needed for rapid and efficient evaluation and screening of numerous genotypes, prior to clonal distribution or use as parents in breeding programmes.

Depending on financial means, there is some variation between the existing programmes but the rationale is the same. Heavy selection pressure is applied at the seedling stage for resistance to diseases. This selection process is visual without any data recording in order to minimize the costs and maximize the number of genotypes assessed. The selected clones are then released as new varieties. A new selection cycle can begin, in which the new selected varieties are used as parents. Unfortunately, chemotypes with very attractive properties can be eliminated at an early stage because of the high selection pressure on other traits (Figure 1).

Each selected offspring individual can represent a new potential cultivar. However, the chances of getting a high yielding hybrid with excellent eating quality are very low and they become much lower when the selection procedure includes resistance against diseases. Corm yield and corm quality appear to be negatively correlated. Soft corms, with high water content, generally characterize high yielding early maturing hybrids. Unfortunately, the physico-chemical characteristics determining the quality of the corms are very expensive and laborious to assess. Low-cost methods for evaluation of numerous accessions need to be developed.

NIRS (Near Infrared Reflectance Spectroscopy) has been used to predict major constituent contents in maize (Berardo *et al.* 2009), rapeseed (Amar *et al.* 2009), sorghum (De Alencar Figueiredo *et al.* 2006), sugar beet (Roggo *et al.* 2004), malt (Marte *et al.* 2009), wheat (Rakszegi *et al.* 2008), potato (Haase, 2006) and tropical root and tuber crops (Lebot *et al.* 2009). NIRS has been intensively used for quality testing of hybrid lines from cereal breeding programmes around the World since the late 1970s with great success (Osborne 2006). More recently, Lebot *et al.* (2011) developed models based on 245 accessions of taro to predict concentrations in a 58 varieties with encouraging results. The r^2pred values were 0.76 for starch, 0.74 for sugars, 0.85 for both proteins and minerals but amylose and cellulose could not be predicted.

These various studies indicate that NIRS could be a useful tool for aroids breeding, selection and quality control. However, the limiting factor for predictive models development remains the cost of chemical analyses (approx 100 euros per sample for five major components) which limit the number of accessions that can be used for comparison between chemical and spectral data. In the present study, we attempt to investigate the potential of new models developed on samples originating from different root crop species (cassava, sweet potato, yam, taro) and to test their predicating values on a set of 100 randomly chosen taro varieties. The results and their practical applications for breeding and improving the quality of Aroids corm and cormels are discussed.

Materials and Methods

Chemical analyses. Overall, 742 accessions representing varieties as well as hybrid lines from four different species and various geographical origins, were chemically analyzed (Table 1). One full corm, root or tuber was peeled and cut. Approximately 0.5-1 kg of fresh weight, corresponding to the central part of the underground organ, were manually sliced into chips and oven dried at 60°C for 48 h. Dry matter samples were split into two sub-samples: one sub-sample was used for chemical analysis and the other for NIRS. Samples of 200g were sent to Laboratoire d'Analyses Agricoles Teyssier, Bourdeaux, France, for chemical analyses. Samples of approximately 50g of dried chips

were milled into flour just after oven drying and dried chips were ground in a stainless kitchen steel mill (SEB, France) prior to NIRS analysis in Vanuatu.

Major constituents (starch, sugars, cellulose, total N and ash) were analyzed according to AFNOR (Association Française, the French standards association) and EEC methods (AFNOR, 2011). Following NF (Norme Française) V 18-109 for dry matter (DM) determination, samples were dried again to remove residual moisture (measured as % of total dry weight) and the powder was analyzed on an oven-dried air basis. Moisture was therefore expressed as a measurement of the sample prior to drying. All measurements were then expressed in %DM and the data were adjusted by the residual moisture following oven drying.

Starch was quantified using Ewers protocol (NF ISO 10-520) corresponding to hydrolysis in HCl, filtration and polarimetric measurement (specific rotation: 185.7°). Total sugars were quantified through the colorimetric method of Luff Schoorl (CEE 98\54\CE). Crude cellulose (total fibers) was measured by Weende method (NF V 03-040) which corresponds to non soluble organic residue obtained by sulfuric acid and alkaline treatments. Total N content (considered as equivalent total proteins) was calculated using the Kjeldahl method (NF V 18-100). Estimation of total minerals content was obtained from ashes produced at 550°C (NF V 18-101). All analyses were performed in duplicate with accepted mean coefficient of variation (SEL) of ± 3% for starch, sugars, cellulose, and residual moisture and ± 2% for proteins (equivalent N), and ashes (minerals).

NIRS measurements and data pre-treatment. Dry matter samples were milled into flour and granules size was homogenized using four sieves with decreasing diameters until granules passed through the 106 µm sieve. An ASD LabSpecPro spectrophotometer from Analytical Spectral Devices Inc. (ASD Inc., Boulder, Colorado, USA) fitted with a "muglight" or High Intensity Source Probe (HISP) (ASD Inc.) was used for the measurement of all spectra over the wavelength range of 350–2500 nm (Figure 2). On average, six grams of homogenized taro flour were placed in an individual cell and compacted with a tea spoon to eliminate air voids within the sample. Each spectrum was obtained by averaging three different cells (repetitions) per sample with 25 scans for each. A reference reading (baseline) was taken when starting a session and another every 30 min. All of the spectra were recorded in diffuse reflectance as *log(1/R)* with respect to a Labsphere's Spectralon material reflectance standard (Labsphere Inc., North Sutton, New Hampshire) which is a Lambertian reflective PTFE (thermoplastic resin) with high overall reflectance. For each sample, corresponding to individual accession, three sub-samples were scanned 25 times each and then averaged. The resulting averaged spectrum was recorded for the accession. Overall 742 spectra were recorded and converted to absorbance using the Indico software (ASD Inc.). In order to assess the performance of the calibration, samples were separated in two sets: the calibration and the prediction sets. The prediction set was created by randomly selecting 100 taro (*Colocasia esculenta*) accession numbers (approx. 13% of total 742 acc.) and the calibration set contained 642 samples.

Data analysis. Major constituents chemical data were subjected to multivariate analysis using XLSTAT (version 6.02, 2009). Multivariate analysis (Principal Component) of the spectra was conducted with GRAMS/AI (version 8.0). The spectra and reference data were mathematically modeled using with PLSPlus/IQ spectroscopy software (Thermo Electron Corporation, Ohio, US). Using the values obtained with chemical analyses as the analyte value, a separate calibration was made for each of the six major constituents. Calibration of residual moisture was not attempted because spectra were recorded in Vanuatu, just after oven drying the samples, while residual moisture was measured in France on hygroscopic dry raw material. Partial least-squares (PLS) regression technique was used to develop a predictive model of the near-infrared part of the spectra. The optimum number of PLS factors used for prediction was determined by full cross-validation and PRESS (Prediction Residual Error Sum of Squares). Additionally, light scattering effects due to particle size differences were corrected by multiplicative scatter correction (MSC). The data was mean-centered and the average spectrum calculated from all of the calibration spectra and then subtracted from every calibration spectrum.

As part of the model process a Principal Component Analysis (PCA) was used to check for gross spectroscopic outliers. The Mahalanobis distance of each spectrum to the mean spectrum of the group was calculated and the removal of outliers was based on distance $H > 3$ from the average spectrum of the file. Spectra and concentration outliers were removed and PLS was run again until the highest r^2_{cv} (determination coefficient for cross validation) corresponding to the smallest SECV (Standard Error of Cross Validation) were obtained. At that point, factor loadings were used to determine which wavelengths were important to correlate with concentrations in order to narrow down the spectroscopic region. The loading plots showed which wavelengths were important to correlate with concentrations. The loading weights showed how much each wavelength point contributed to explaining the response along each model component. For starch, proteins and minerals, the regions used were 800-2400nm, while for sugars, and cellulose the region was 1200-2400nm. The PLS analysis was then conducted again on these new regions in order to obtain for each constituent, equations with higher explanation of the total variability in the calibration values without increasing the number of PLS factors used.

Statistical parameters used to evaluate models performances included the standard error of calibration (SEC), the determination coefficient for cross validation (r^2_{cv}), the standard error of cross-validation (SECV), the determination coefficient for prediction (r^2_{pred}), and the standard error of prediction (SEP). SEC and SEP were calculated using Excel spreadsheet by squaring the differences of the actual minus the predicted concentrations for each sample in the calibration (SEC) and test (SEP) sets. These values were then summed and the sum was divided by the number of samples (n). The square root of this value was used for SEC and SEP. SEC describes the calibration set (642 acc.) and SEP describes the test set composed of 100 taro samples not included in the calibration set. The ratio of performance to deviation (RPD= SD/SECV) was also used to evaluate performances of the models (with SD as the standard deviation of the original chemical data in the calibration set (Williams, 2003).

Results

Overall, 742 accessions were analyzed for the chemical variation of their major constituents. Results of the chemical analyses are presented in Table 1. Significant variation was observed for all major constituents. The least variable constituent was starch (CV%= 10.50) and the most variable was total sugars (CV%= 83.25). Correlation coefficients calculated between major constituents indicate that starch content is positively correlated with %DM but negatively correlated with sugars, cellulose, proteins and minerals contents (Table 2).

Principal Component Analysis conducted on the data matrix (742 acc. x 5 major constituents) reveals the respective contribution of the five variables to the projection, with axes 1 and 2 totalizing 72.49% of the total variance (Figure 3).

The comparison of the NIRS spectra and the chemical values allowed the establishment of equations of calibration for the prediction of starch, sugars, proteins (equivalent N) and minerals. The results are presented in Table 3. For starch, the SECV (2.11%) and SEC (2.11%) values are identical indicating robust fitting. The SEP (2.34%) is not too distant and the r^2_{pred} of 0.79 indicates an acceptable estimation of the equation accuracy on the 100 validation samples. Deviations of single samples are visualized in a scatter plot between measured and predicted starch values of the 58 acc. in the test set (Figure 5A). In terms of predictive performance, the equations for starch could be considered as good with RPD parameters close to 4. Some authors claim that a RPD value of at least 3 is necessary for efficient NIR reflectance predictions with values above 3.5 indicating a very good predictive model (Williams, 2003).

The total sugars model also presents similar SECV (1.30) and SEC (1.32) values but the SEP is not too far (1.40) and the r^2_{pred} is of 0.80. The RPD value of 4.11 indicates a good predictive potential for this equation. Deviations of single samples are visualized in a scatter plot between measured and predicted sugars values (Figure 5B). Cellulose could not be satisfactorily predicted and a poor r^2_{cv} (0.31) was obtained, with very low r^2_{pred} (0.13) and RPD (1.91). Proteins, (measured as total N

equivalent) produce similar SECV, SEC and SEP values (respectively 0.51, 0.53 and 0.63) and a r^2_{pred} of 0.78 indicating good and robust prediction with 78% of confidence. However, the RPD value above 6.0 confirms a very good potential of prediction for this model. If the r^2_{pred} value is not higher than 0.78, this might be due to the fact that the samples selected for the validation test were first of all chosen on the decreasing value of their starch contents, not on proteins, and that therefore they might not represent properly the extent of variation found in proteins. Deviations of single samples are visualized in a scatter plot between measured and predicted proteins values (Figure 5C). Minerals are known to have a poor relationship with NIRS but they presented similar SECV, SEC and SEP values (respectively 0.46, 0.51, 0.38) and could be predicted with 87% of confidence with a good RPD value of 2.52. Deviations of single samples are visualized in a scatter plot between measured and predicted minerals values (Figure 5D).

The r^2_{pred} values of starch, sugars, proteins and minerals are high enough to allow good estimates of their contents. RPD between 3.90 and 4.11 for the starch and sugars models, also allow good quantitative predictions to be made. Values above 2.5 for proteins are considered to be good models (Williams, 2003) and the value here is above 6.0. The number of terms is also relatively low if we consider a general recommendation of 1 factor for every 10 samples in a model (Table 3).

Models for starch, sugars, proteins and minerals present good potential but will need to be further tested on independent samples. When SECV and SEP values differ significantly, this could be an indication that too many samples (HT >3 = 21) were removed during the modeling process.

The models developed in the present study show good accuracy but it remains to be seen whether larger sample sets will improve them to enable more precise prediction. When comparing the performance of the new calibration models (with n=642 from different species), with the values reported by Lebot *et al*. (2011) for taro with 245 accessions in the calibration set and only 58 accessions in the validation set, high r^2_{cv} and RPD values were confirmed. Determination coefficients (r^2_{pred}) generally improve as the working range increases. Consequently, if more range is added in the same model then it could improve coefficient values. Additionally, when different samples are added, a larger spectroscopic diversity is described and, therefore, some samples might actually be better spectrally described as the number of samples in the calibration set increases. However, determination coefficients for the prediction set (r^2_{pred}) cannot reflect the whole situation because the range of the 100 accessions values affects the coefficient values. These values change according to the type and number of validation samples and it is necessary to consider the long term effects. Errors of prediction values have been shown to have uncertainties and it is therefore recommended to be cautious while reporting prediction errors because they may change according to the validation set used (Sileoni *et al*., 2011). SEP is, therefore, a better overall indicator. A better sample selection might be helpful by selecting, for example, on constituent concentration rather than a random selection of numbers. Further work should concentrate on validating the results over different years.

The models for starch, sugars, proteins and minerals present potential for improvement if more samples could be added. The protein content calibration is particularly interesting as it can be further improved. Proteins content is usually estimated by multiplying the total N content by a standard conversion factor of 6.25. However, the nitrogen to protein ratio does vary according to the species considered and change with amino acid content and mineral nitrogen and non-protein nitrogen. For the present study, we decided to present our results measured as total N as proteins. In the future, it would of interest to improve the calibration models on the real protein content of taro which vary according to amino acids. Once known, the values obtained by the Kjeldahl method could be converted into more accurate measurements for NIRS calibrations on taro.

In taro breeding programs, mass selection results in the rapid accumulation of suitable genes but has to be complemented with efficient screening techniques of hundreds of hybrids generated in controlled crosses. Correlation coefficients between major constituents indicate that breeding for increased DM and starch contents will reduce sugars, proteins and minerals. These correlations do not present practical problems as “poor” quality varieties have been shown to present low DM and starch and high sugars, cellulose, proteins and minerals (Lebot *et al*., 2011). Obviously, NIRS could assist taro breeders in their choice and selection of the best genotypes, based on the chemical

composition requested by consumers by predicting simultaneously starch, sugars, proteins and minerals on a single sample. As starch is significantly negatively correlated with the other three major constituents, the simultaneous prediction of all four constituents allow for rapid estimation of the variety chemotype and therefore its quality.

Taro is a diploid species but nothing is known on the segregation of these major constituents. The problem is rather complex as these constituents are most likely controlled by many sets of different genes and molecular tools can hardly be used for markers assisted selection and conventional selection of parents for breeding or selection. NIRS offer interesting perspectives for spectra assisted selection.

Abbreviations Used

SEC: standard error of calibration,
SECV: the standard error of cross-validation,
SEP: the standard error of prediction,
SEL: the standard error of the laboratory analysis,
r^2_{cv} *: the determination coefficient for cross validation,*
r^2_{pred} *the determination coefficient for prediction,*
RPD: the ratio of performance to deviation.
H: Mahalanobis distance limit.
HT: number of outliers removed.
%DM: percentage of dry matter.
CV%: coefficient of variation.

Acknowledgements

This research was financially supported by the Food Security Thematic Programme of the EU (grant no DCI-FOOD/2010/230-267) and the International Network for Edible Aroids (INEA, www.EdibleAroids.org). We thank S. Stephens for laboratory assistance in preparing samples and taking spectra.

References

AFNOR. 2011. http://www.boutique.afnor.org/NEL1AccueilNormeEnLigne.aspx.

Amar, S.; H.C. Becker; C. Möllers. 2009. Genetic variation in phytosterol content in winter rapeseed (*Brassica napus* L.) and development of NIRS calibration equations. *Plant Breed*. *128,* 78-83.

Berardo, N.G.; Mazzinelli, P.; Valoti, P.; R. Laganà Redaelli. 2009. Characterization of Maize Germplasm for the Chemical Composition of the Grain. *J. Agric. Food Chem*. *57,* 2378-2384.

De Alencar Figueiredo, L.F.; Davrieux, F.; Fliedel, G.; Rami, JF.; Chantereau, J.; Deu, M.; Courtois, B.; C. Mestres. 2006. Development of NIRS Equations for Food Grain Quality Traits through Exploitation of a Core Collection of Cultivated Sorghum. *J. Agric. Food Chem. 54,* 8501-8509.

Haase, N.U. 2006. Rapid Estimation of Potato Tuber Quality by Near-Infrared Spectroscopy. *Starch/Stärke. 58,* 268-273.

Ivancic A. and V. Lebot. 2000. *Taro (*Colocasia esculenta*): Genetics and Breeding.* Collection "Repères", CIRAD, Montpellier 194 p.

Ivancic A., Quero Garcia J., V. Lebot. 2003. Development of visual tools for selecting qualitative corm characteristics of taro (*Colocasia esculenta* (L.) Schott). *Australian Journal of Agricultural Research, 54,* 581-588.

Lebot, V.; Champagne, A.; R. Malapa; D. Shiley. 2009. NIR Determination of Major Constituents in Tropical Root and Tuber Crop Flours. *J. Agric. Food Chem. 57,* 10539–10547.

Lebot, V., Malapa, R. and M. Bourrieau. 2011. Rapid Estimation of Taro (*Colocasia esculenta*) Quality by Near-Infrared Reflectance Spectroscopy. *Journal of Agricultural and Food Chemistry* doi.org./10.1021/jf202310w.

Marte, L.; Belloni, P.; Genorini, E.; Sileoni, V.; Perretti, G.; Montanari, L.; O. Marconi. 2009. Near-Infrared Reflectance Models for the Rapid Prediction of Quality of Brewing Raw Materials. *J. Agric. Food Chem. 57,* 326-333.

Osborne, B.G. 2006. Applications of near infrared spectroscopy in quality screening of early generation material in cereal breeding programmes. *Journal of Near Infrared Spectroscopy* 14, 93-101.

Rakszegi, M.; Boros, D.; Kuti, C.; Làang, L.; Bedō. Z.; P.R. Shewry. 2008. Composition and End-Use Quality of 150 Wheat Lines selected for the HEALTHGRAIN Diversity Screen. *J. Agric. Food Chem. 56,* 9750-9757.

Roggo, Y.; Duponchel, L.; Huvenne, JP. 2004. Quality Evaluation of Sugar Beet (*Beta vulgaris*) by Near-Infrared Spectroscopy. *J. Agric. Food Chem. 52,* 1055-1061.

Sileoni, V.; van den Berg, F.; Marconi, O.; Perretti, G.; P. Fantozzi. 2011. Internal and External Validation Strategies for the Evaluation of Long-Term Effects in NIR Calibration Models. *J. Agric. Food Chem. 59,* 1541-1547.

Williams, P. 2003. Near-Infrared Technology: Getting the Best out of Light. A short course in the practical implementation of near-infrared spectroscopy for the user. PDK Projects, Inc.: Nanaimo, Canada.

Table 1. Descriptive statistics on 742 accessions

Spp	Stats	DM	starch	sugars	cellulose	proteins	minerals
Dioscorea	min	12.35	58.78	0.40	0.10	4.40	1.58
n = 135	max	52.44	90.40	18.30	6.30	21.00	8.10
	mean	31.26	77.14	3.62	2.68	10.39	4.36
	std	5.89	6.09	3.56	0.98	3.13	1.20
	cv%	18.84	7.90	98.19	36.61	30.16	27.56
Manihot	min	28.20	60.64	1.52	1.73	1.33	1.24
n = 66	max	55.46	91.21	18.32	11.84	7.42	3.96
	mean	37.82	85.23	5.07	3.78	2.77	2.53
	std	5.08	5.76	3.06	1.96	1.15	0.54
	cv%	13.43	6.76	60.35	51.81	41.55	21.33
Ipomoea	min	11.62	52.15	1.49	2.39	2.67	1.35
n = 227	max	53.29	83.83	25.29	15.80	9.97	8.22
	mean	30.94	69.17	10.93	5.11	5.70	3.45
	std	5.50	5.95	4.77	2.81	1.19	0.90
	cv%	17.79	8.60	43.62	55.09	20.89	26.21
Colocasia	min	19.51	44.96	0.21	1.40	2.13	1.47
n = 314	max	53.74	89.46	28.05	7.30	14.79	8.85
	mean	35.81	78.29	4.23	3.09	4.67	4.18
	std	7.01	6.66	4.12	0.92	1.71	1.10
	cv%	19.56	8.51	97.28	29.77	36.64	26.41
All spp.	min	11.62	44.96	0.21	0.10	1.33	1.24
n = 742	max	55.46	91.21	28.05	15.80	21.00	8.85
	mean	32.21	75.91	6.25	3.69	5.85	3.84
	std	6.42	7.97	5.20	2.06	2.96	1.16
	cv%	19.93	10.50	83.25	55.70	50.56	30.22

Table 2. Correlation coefficients (Pearson (n-1)) for 742 accessions:

Variables	DM	starch	sugars	cellulose	proteins
starch	*+ 0.425***				
sugars	*- 0.286***	*- 0.798*			
cellulose	*- 0.312***	*- 0.469*	*+ 0.451*		
proteins	*- 0.183*	*- 0.298*	*- 0.058*	*- 0.038*	
minerals	*- 0.426***	*- 0.252*	*- 0.147*	*+0.010*	*+0.336*

Table 3. Statistical Parameters of the New Calibration and Validation Sets.

Constituents*	Calibration *n = 642*									Validation *n = 100*	
	Mean % DM	SD	SEL ±	HT H>3	PLS terms	r^2_{cv}	SECV	SEC	RPD	r^2_{pred}	SEP
starch	75.39	8.21	2.62	36	20	*0.93*	2.11	2.11	3.90	*0.79*	2.34
sugars	6.68	5.33	0.20	37	17	*0.93*	1.30	1.32	4.11	*0.80*	1.40
cellulose	3.82	2.17	0.11	43	10	*0.31*	1.14	1.14	1.91	*0.13*	0.78
proteins	6.09	3.07	0.12	44	18	*0.97*	0.51	0.53	6.03	*0.78*	0.63
minerals	3.78	1.16	0.08	50	12	*0.80*	0.46	0.51	2.52	*0.87*	0.38

**starch, proteins, minerals in 800-2400, sugars in 1200-2400*

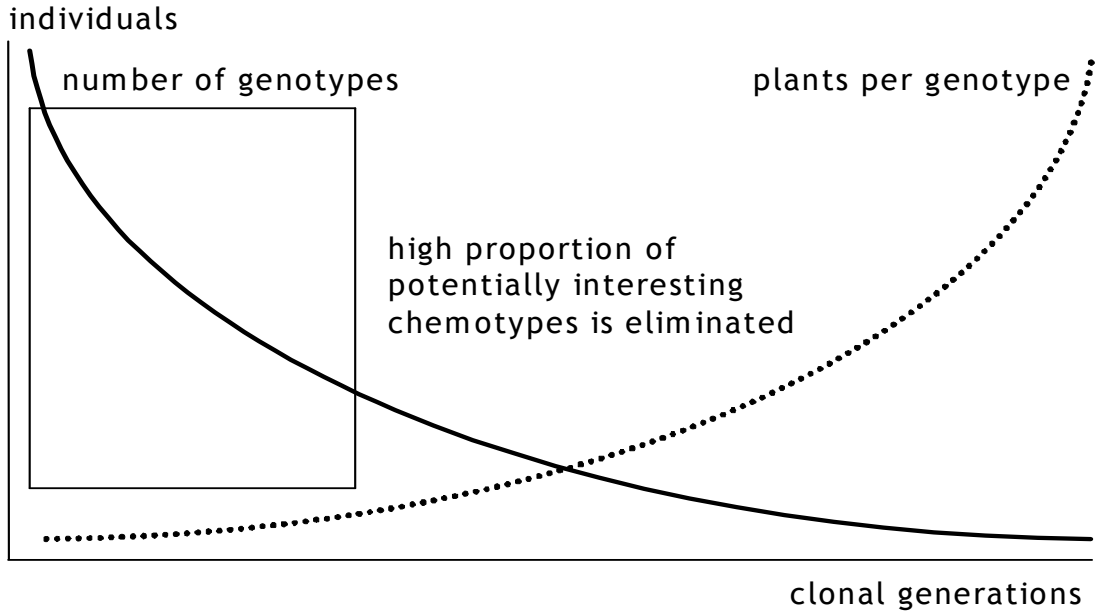


Figure 1:

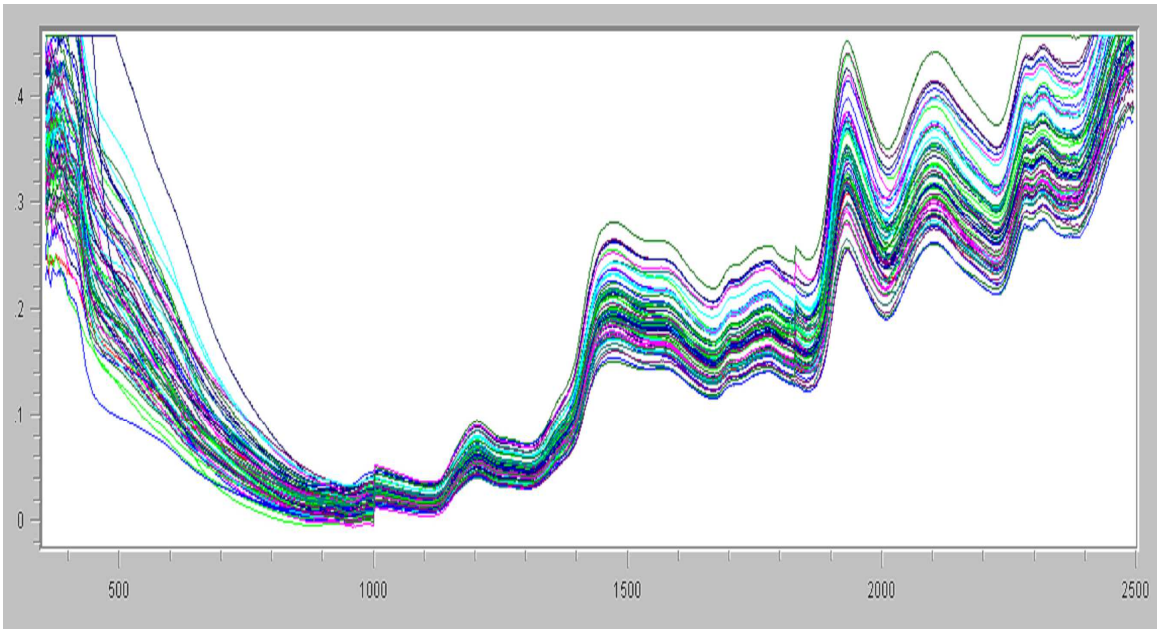


Figure 2

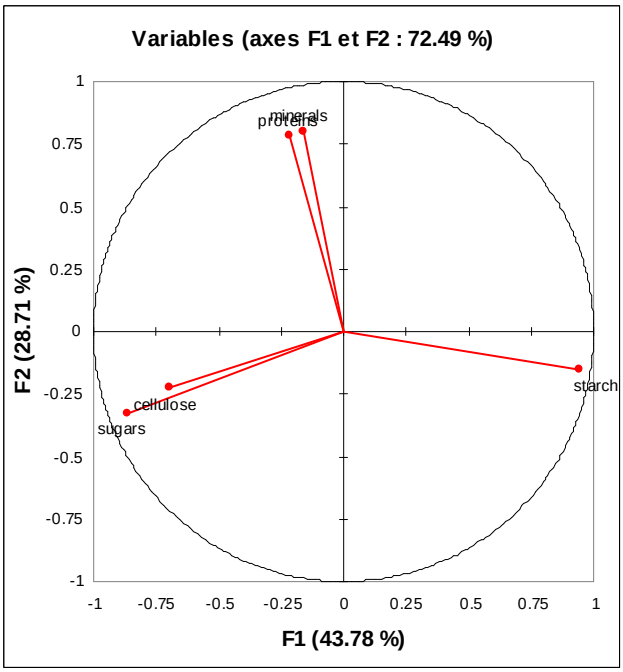


Figure 3

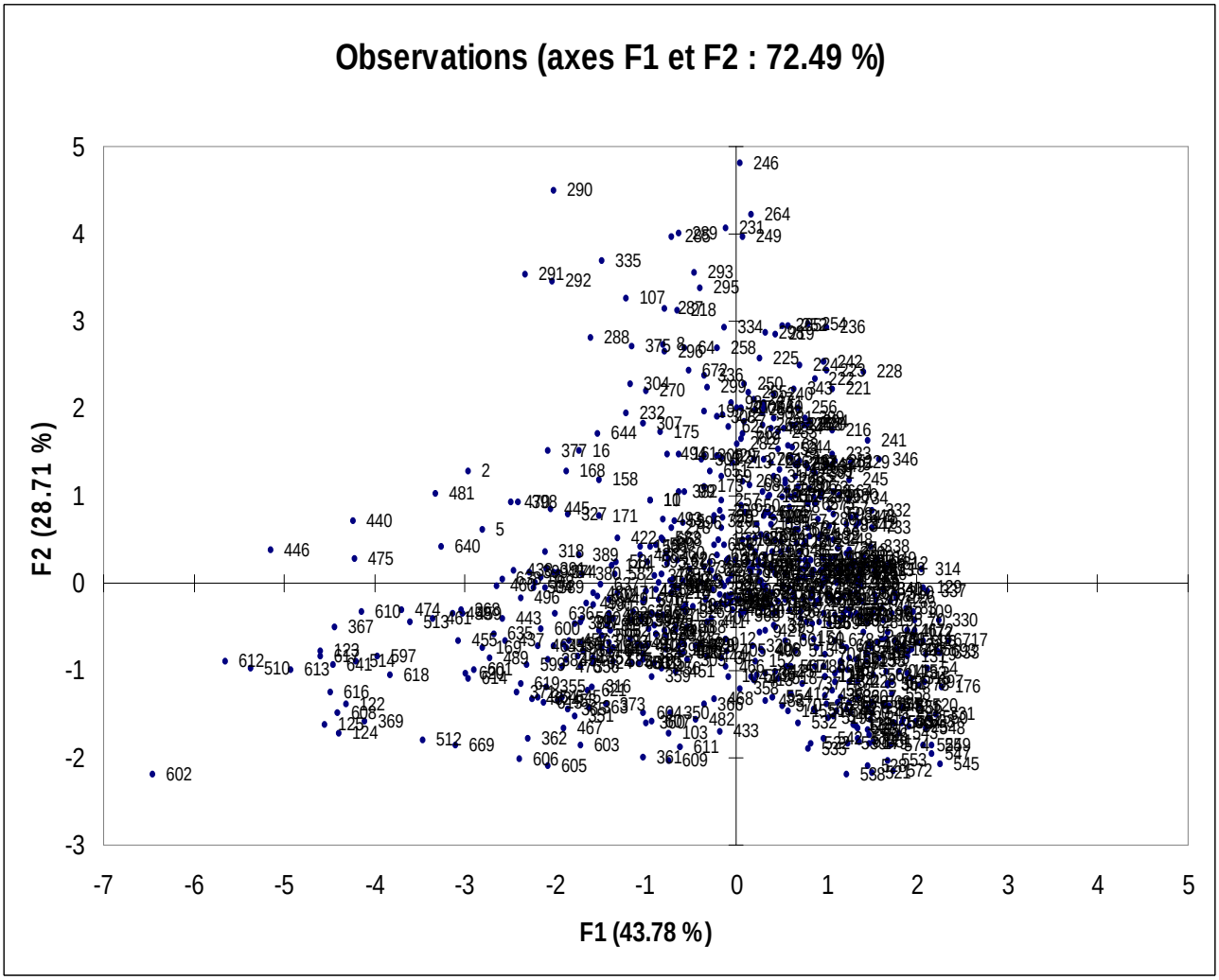
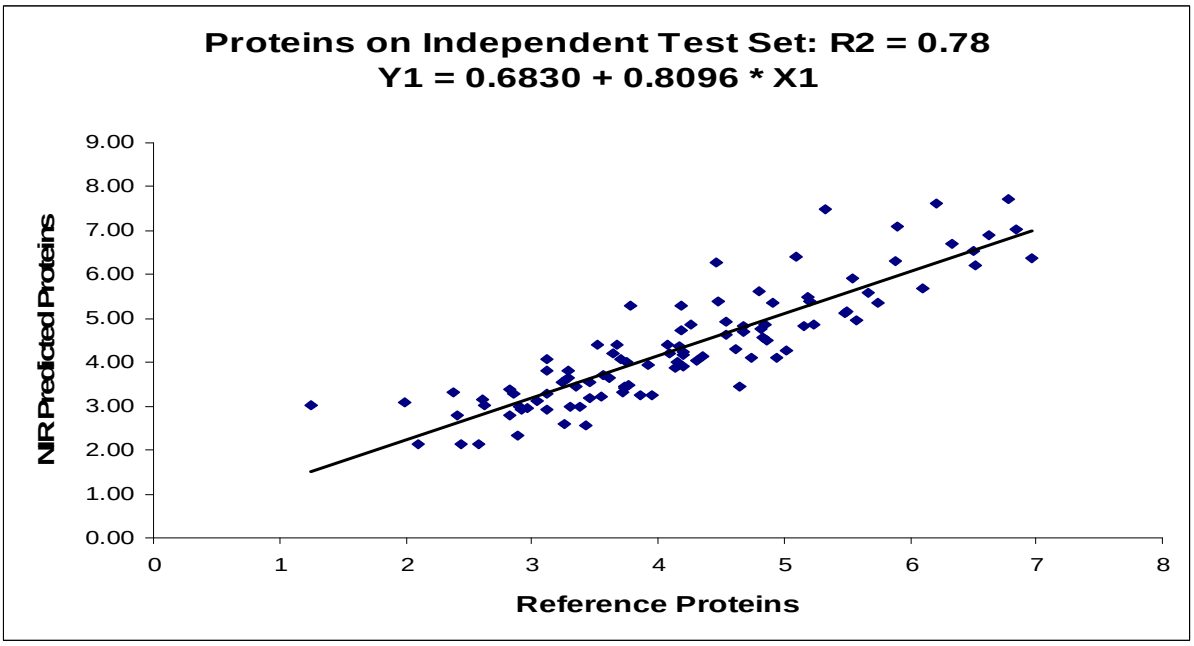
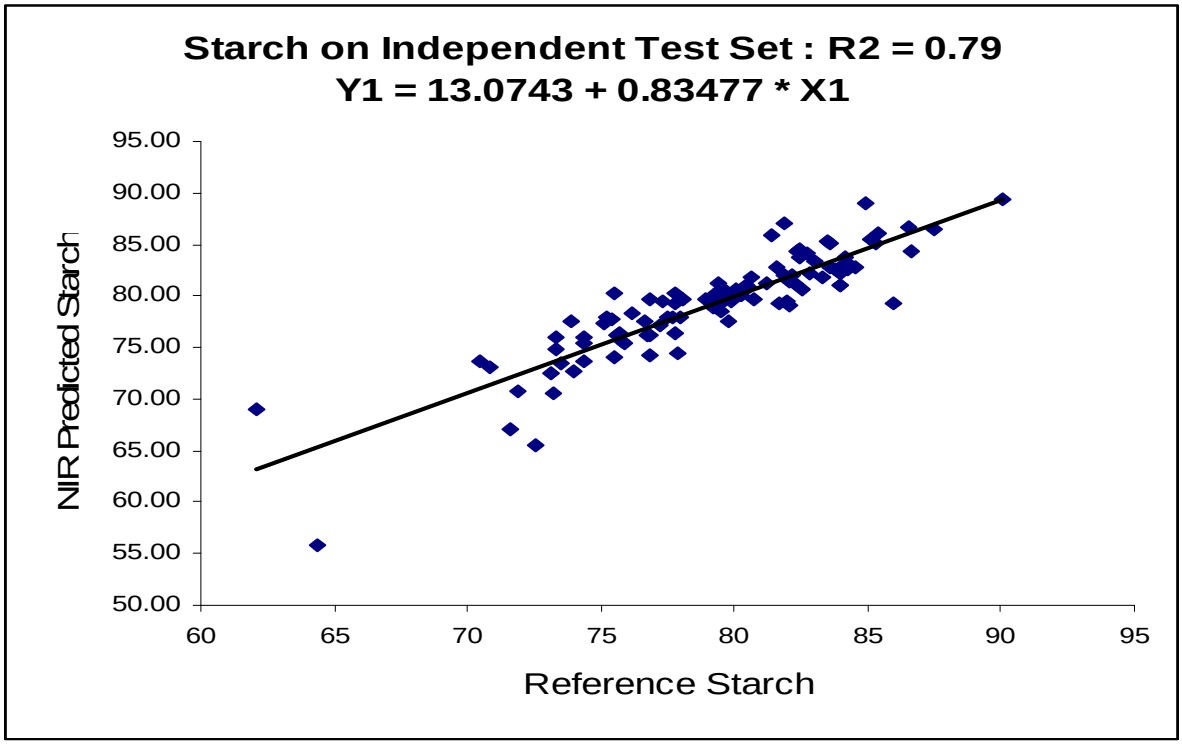
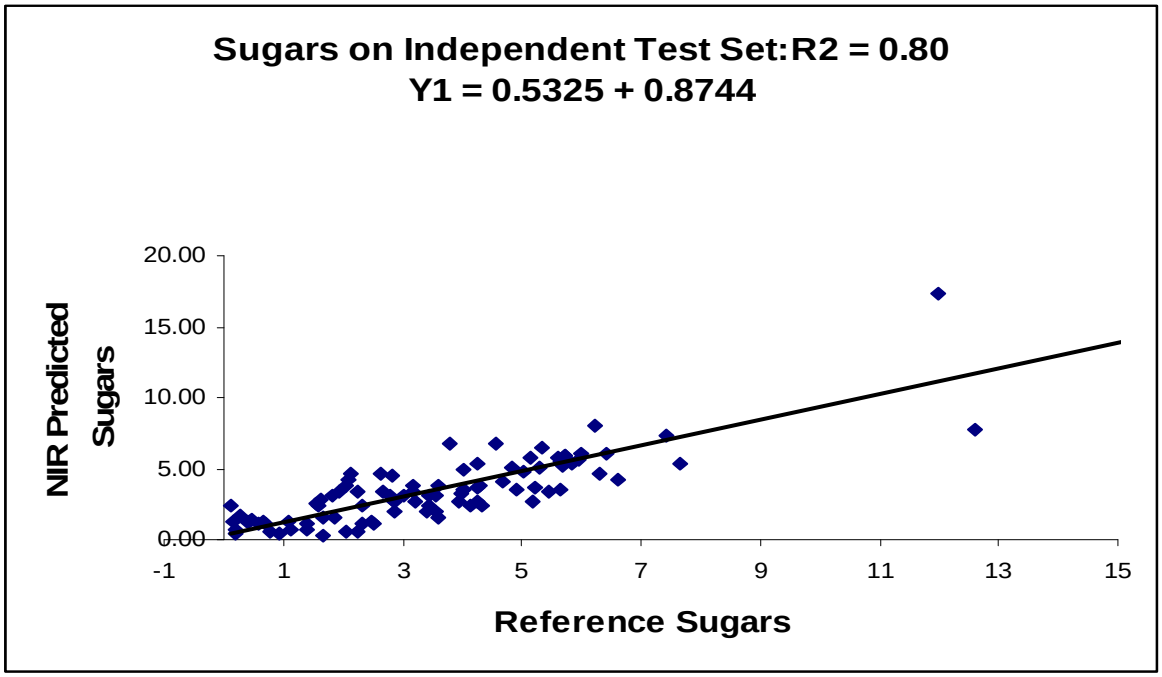
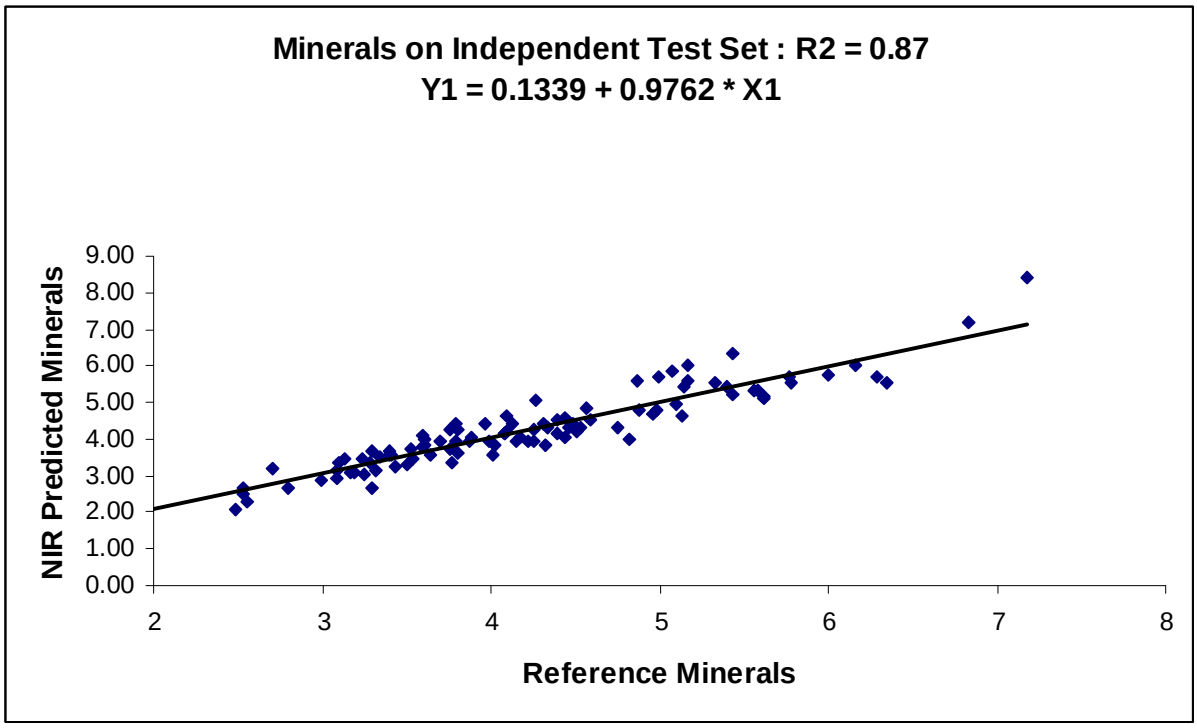


Figure 4


Proteins on Independent Test Set: R2 = 0.78
Y1 = 0.6830 + 0.8096 * X1
NIR Predicted Proteins
Reference Proteins
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
0
1
2
3
4
5
6
7
8


Starch on Independent Test Set : R2 = 0.79
Y1 = 13.0743 + 0.83477 * X1
NIR Predicted Starch
Reference Starch
95.00
90.00
85.00
80.00
75.00
70.00
65.00
60.00
55.00
50.00
60
65
70
75
80
85
90
95


Sugars on Independent Test Set:R2 = 0.80
Y1 = 0.5325 + 0.8744
NIR Predicted Sugars
20.00
15.00
10.00
5.00
0.00
-1
1
3
5
7
9
11
13
15
Reference Sugars


Minerals on Independent Test Set : R2 = 0.87
Y1 = 0.1339 + 0.9762 * X1
NIR Predicted Minerals
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
2
3
4
5
6
7
8
Reference Minerals