Individual authentication of cocoa beans using Vis-NIR spectroscopy

Authors: Antoine Deryck[a], Anne Pinto[b], Giulia Victória Lima[b], Juan Antonio Fernández Pierna[a], Vincent Baeten[a], Hervé Rogez[b]

[a] Walloon Agricultural Research Center (CRA-W), Knowledge and Valorization of Agricultural Products Department
[b] Federal University of Pará (UFPA), Center for Valorization of Amazonian Bioactive Compounds (CVACBA)

Context

Theobroma cacao

Native to Central and South America

Grown in various tropical regions

Major consumers = Europe and the US

High economic importance

Major producers = West and Central Africa, South America, and Eastern Asia

Context

Theobroma cacao

Native to Central and South America

Grown in various tropical regions

Major consumers = Europe and the US

Major producers = West and Central Africa, South America, and Eastern Asia

High economic importance

- Brazil = 5th consumer and 6th producer globally
  Half from Amazonia region (mainly state of Pará)

Cocoa processing

1. Grow cocoa

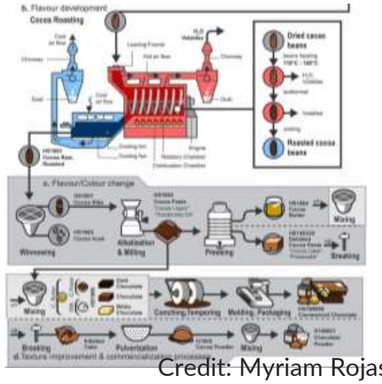
2. Harvest cocoa pods

3. Break cocoa pods to extract unfermented beans

4. Fermentation of beans and mucilage ⚠

5. Drying of beans ⚠

6. Collect fermented and dried beans



Credit: Myriam Rojas

7. Industrial steps

Importance of genotypes

- Composition (and later flavor/quality)

Genotype	°BRIX exter	°BRIX inter	pH externo	pH inter	L	a	b	h	Protein full	Lipids Full	Ash
PA169	3.98	13.0	5.83	4.92	26.73	18.21	12.19	33.8835	17.89	36.64	2.03
CCN51	3.43	13.0	5.87	5.08	23.52	15.88	15.85	44.97921	16.62	30.26	2.60
CAB324	3.01	11.0	6.02	5.15	20.34	16.34	14.39	41.46706	16.40	30.50	2.37
P7	2.85	10.0	5.82	5.11	22.04	13.82	14.80	32.68058	18.33	31.61	1.91
CAB499	2.96	11.0	5.98	5.11	26.17	16.66	10.70	32.81032	18.59	32.85	2.21
CA6	3.20	12.0	6.08	5.21	23.90	14.72	17.16	35.21722	18.57	29.74	2.03
PA121	3.66	14.0	5.76	4.94	22.90	15.43	15.28	44.64245	18.09	34.78	1.99
PA195	5.27	11.0	5.65	5.08	19.95	13.94	13.05	43.04929	18.43	30.64	2.01
BE10	3.72	13.5	5.79	4.85	23.41	15.37	11.74	37.5485	16.71	31.67	1.96
CAB270	2.78	15.0	6.13	5.16	25.46	15.31	15.80	41.52663	18.80	29.76	2.20
RB40	5.09	11.0	6.04	5.32	23.98	15.04	15.39	46.37243	20.29	30.37	2.30
RB36	4.35	12.0	5.91	4.87	22.59	15.56	12.78	39.89864	18.44	30.05	1.91
CAB208	4.45	12.5	6.33	5.17	26.01	15.86	17.56	48.11005	17.70	26.31	1.81
IMC67	3.37	14.0	6.15	4.91	22.18	13.81	14.67	47.18841	17.05	31.30	2.50
MO1	3.78	12.0	5.91	5.01	20.64	14.95	14.43	43.93385	16.74	28.34	2.26
CAB214	4.39	9.5	5.93	5.12	17.42	15.40	13.52	41.42286	18.94	29.51	1.77
MA11	4.96	11.5	5.75	4.82	24.44	16.44	13.09	38.86	19.15	35.40	1.98
MA15	3.08	13.0	5.81	4.99	24.41	15.88	12.58	38.76865	18.97	30.27	2.33

- Disease-resistant

- High yield

Importance of genotypes

- Composition (and later flavor/quality)

Genotype	°BRIX exter	°BRIX inter	pH externo	pH inter	L	a	b	h	Protein full	Lipids Full	Ash
PA169	3.98	13.0	5.83	4.92	26.73	18.21	12.19	33.8835	17.89	36.64	2.03
CCN51	3.43	13.0	5.87	5.08	23.52	15.88	15.85	44.97921	16.62	30.26	2.60
CAB324	3.01	11.0	6.02	5.15	20.34	16.34	14.39	41.46706	16.40	30.50	2.37
P7	2.85	10.0	5.82	5.11	22.04	13.82	14.80	32.68058	18.33	31.61	1.91
CAB499	2.96	11.0	5.98	5.11	26.17	16.66	10.70	32.81032	18.59	32.85	2.21
CA6	3.20	12.0	6.08	5.21	23.90	14.72	17.16	35.21722	18.57	29.74	2.03
PA121	3.66	14.0	5.76	4.94	22.90	15.43	15.28	44.64245	18.09	34.78	1.99
PA195	5.27	11.0	5.65	5.08	19.95	13.94	13.05	43.04929	18.43	30.64	2.01
BE10	3.72	13.5	5.79	4.85	23.41	15.37	11.74	37.5485	16.71	31.67	1.96
CAB270	2.78	15.0	6.13	5.16	25.46	15.31	15.80	41.52663	18.80	29.76	2.20
RB40	5.09	11.0	6.04	5.32	23.98	15.04	15.39	46.37243	20.29	30.37	2.30
RB36	4.35	12.0	5.91	4.87	22.59	15.56	12.78	39.89864	18.44	30.05	1.91
CAB208	4.45	12.5	6.33	5.17	26.01	15.86	17.56	48.11005	17.70	26.31	1.81
IMC67	3.37	14.0	6.15	4.91	22.18	13.81	14.67	47.18841	17.05	31.30	2.50
MO1	3.78	12.0	5.91	5.01	20.64	14.95	14.43	43.93385	16.74	28.34	2.26
CAB214	4.39	9.5	5.93	5.12	17.42	15.40	13.52	41.42286	18.94	29.51	1.77
MA11	4.96	11.5	5.75	4.82	24.44	16.44	13.09	38.86	19.15	35.40	1.98
MA15	3.08	13.0	5.81	4.99	24.41	15.88	12.58	38.76865	18.97	30.27	2.33

- Disease-resistant

- High yield

→ High number of genotypes commonly planted in Pará

→ Co-planting of genotypes and blending of beans makes it hard to identify high-economic value genotypes

→ Beans size but not accurate

Identification by PCR but costly, destructive, and time-consuming

Importance of fermentation

1

$$C_6H_{12}O_6\ (glucose) \xrightarrow{\text{Yeast}} 2\ C_2H_5OH\ (ethanol) + 2\ CO_2\ (carbon\ dioxide)$$

$$C_6H_{12}O_6\ (glucose) \xrightarrow{\text{Bacteria}} 2C_3H_6O_3\ (lactic\ acid)$$

$$C_2H_5OH\ (ethanol) + O_2\ (oxygen) \xrightarrow{\text{Bacteria}} CH_3COOH\ (acetic\ acid) + H_2O\ (water)$$

Importance of fermentation

1

$$C_6H_{12}O_6\ (glucose) \xrightarrow{\text{Yeast}} 2\ C_2H_5OH\ (ethanol) + 2\ CO_2\ (carbon\ dioxide)$$

$$C_6H_{12}O_6\ (glucose) \xrightarrow{\text{Bacteria}} 2C_3H_6O_3\ (lactic\ acid)$$

$$C_2H_5OH\ (ethanol) + O_2\ (oxygen) \xrightarrow{\text{Bacteria}} CH_3COOH\ (acetic\ acid) + H_2O\ (water)$$

2

$$\text{Proteins} \xrightarrow{\text{Temperature + Enzymes}} \text{Peptides + Amino Acids}$$

$$\text{Polyphenols} + O_2\ (oxygen) \xrightarrow{\text{Temperature + Enzymes}} \text{Quinones} + H_2O\ (water)$$

Importance of fermentation

1

$$C_6H_{12}O_6\ (glucose) \xrightarrow{\text{Yeast}} 2\ C_2H_5OH\ (ethanol) + 2\ CO_2\ (carbon\ dioxide)$$

$$C_6H_{12}O_6\ (glucose) \xrightarrow{\text{Bacteria}} 2C_3H_6O_3\ (lactic\ acid)$$

$$C_2H_5OH\ (ethanol) + O_2\ (oxygen) \xrightarrow{\text{Bacteria}} CH_3COOH\ (acetic\ acid) + H_2O\ (water)$$

2

$$\text{Proteins} \xrightarrow{\text{Temperature + Enzymes}} \text{Peptides + Amino Acids}$$

$$\text{Polyphenols} + O_2\ (oxygen) \xrightarrow{\text{Temperature + Enzymes}} \text{Quinones} + H_2O\ (water)$$

- Color
- Flavor
- Texture
- Composition
- Quality

Importance of fermentation

1

$$C_6H_{12}O_6\ (glucose) \xrightarrow{\text{Yeast}} 2\ C_2H_5OH\ (ethanol) + 2\ CO_2\ (carbon\ dioxide)$$

$$C_6H_{12}O_6\ (glucose) \xrightarrow{\text{Bacteria}} 2C_3H_6O_3\ (lactic\ acid)$$

$$C_2H_5OH\ (ethanol) + O_2\ (oxygen) \xrightarrow{\text{Bacteria}} CH_3COOH\ (acetic\ acid) + H_2O\ (water)$$

2

$$\text{Proteins} \xrightarrow{\text{Temperature + Enzymes}} \text{Peptides + Amino Acids}$$

$$\text{Polyphenols} + O_2\ (oxygen) \xrightarrow{\text{Temperature + Enzymes}} \text{Quinones} + H_2O\ (water)$$

Color
Flavor
Texture
Composition
Quality

→ Blending of fermented and unfermented beans → Visual ID by cut-test but destructive and subjective

Objectives of the study

Efficacity of Vis-NIR spectroscopy for the non-destructive discrimination of

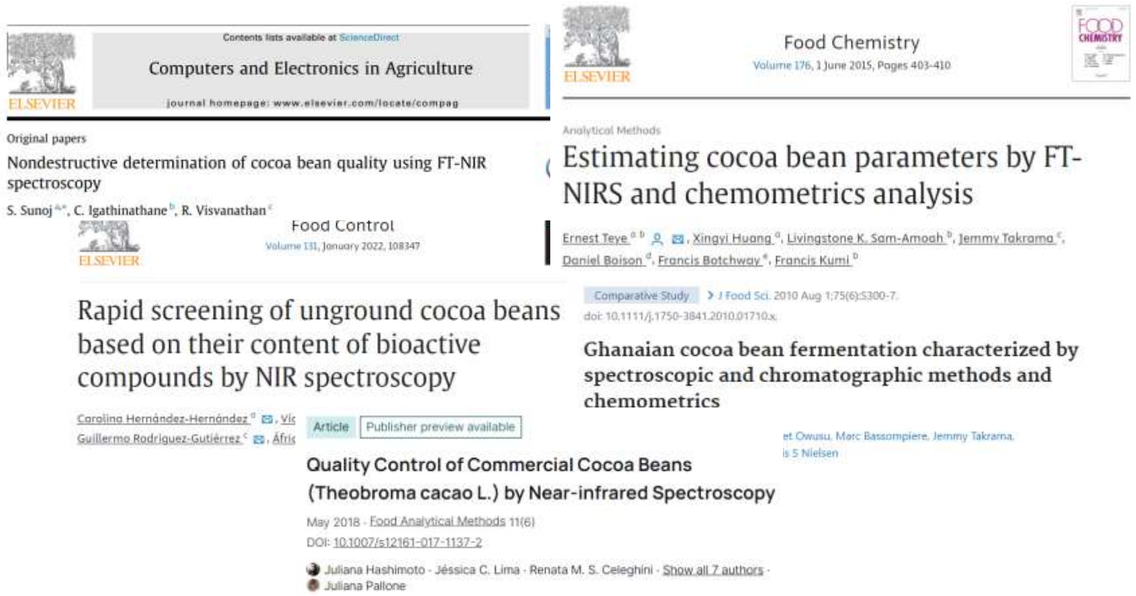
- Fermented from unfermented beans
- Forastero cocoa genotypes from the Eastern Amazonia

Objectives of the study

Efficacity of Vis-NIR spectroscopy for the non-destructive discrimination of

Fermented from unfermented beans

Forastero cocoa genotypes from the Eastern Amazonia



Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Nondestructive determination of cocoa bean quality using FT-NIR spectroscopy

S. Sunoj [a,*], C. Igathinathane [b], R. Visvanathan [c]

Food Control

Volume 131, January 2022, 108347

Rapid screening of unground cocoa beans based on their content of bioactive compounds by NIR spectroscopy

Carolina Hernández-Hernández [a], Vís

Guillermo Rodríguez-Gutiérrez [c], Áfris

Food Chemistry

Volume 176, 1 June 2015, Pages 403-410

Analytical Methods

Estimating cocoa bean parameters by FT-NIRS and chemometrics analysis

Ernest Teye [a,b], Xingyi Huang [a], Livingstone K. Sam-Amoah [b], Jemmy Takrama [c], Daniel Boison [d], Francis Botchway [e], Francis Kumi [b]

Comparative Study > J Food Sci. 2010 Aug 1;75(6):S300-7.

doi: 10.1111/j.1750-3841.2010.01710.x.

Ghanaian cocoa bean fermentation characterized by spectroscopic and chromatographic methods and chemometrics

et Owusu, Marc Bassompiere, Jemmy Takrama, is S Nielsen

Article | Publisher preview available

Quality Control of Commercial Cocoa Beans (Theobroma cacao L.) by Near-infrared Spectroscopy

May 2018 · Food Analytical Methods 11(6)

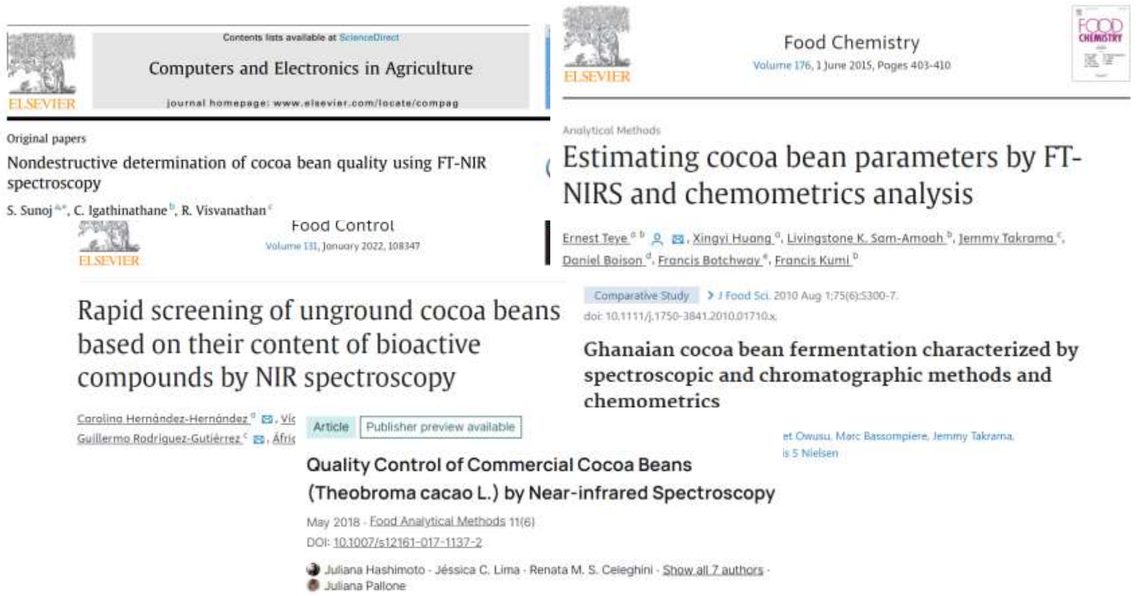
DOI: 10.1007/s12161-017-1137-2

Juliana Hashimoto · Jéssica C. Lima · Renata M. S. Celeghini · Show all 7 authors · Juliana Pallone

Objectives of the study

Efficacity of Vis-NIR spectroscopy for the non-destructive discrimination of

Fermented from unfermented beans



Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Nondestructive determination of cocoa bean quality using FT-NIR spectroscopy

S. Sunoj [a,*], C. Igathinathane [b], R. Visvanathan [c]

Food Control

Volume 131, January 2022, 108347

Food Chemistry

Volume 176, 1 June 2015, Pages 403-410

Analytical Methods

Estimating cocoa bean parameters by FT-NIRS and chemometrics analysis

Ernest Teye [a,b], Xingyi Huang [a], Livingstone K. Sam-Amoah [b], Jemmy Takrama [c], Daniel Boison [d], Francis Botchway [e], Francis Kumi [b]

Rapid screening of unground cocoa beans based on their content of bioactive compounds by NIR spectroscopy

Carolina Hernández-Hernández [a], Vis

Guillermo Rodríguez-Gutiérrez [c], Afris

Comparative Study · J Food Sci. 2010 Aug 1;75(6):S300-7.

doi: 10.1111/j.1750-3841.2010.01710.x.

Ghanaian cocoa bean fermentation characterized by spectroscopic and chromatographic methods and chemometrics

et Owusu, Marc Bassompiere, Jemmy Takrama, is S Nielsen

Article · Publisher preview available

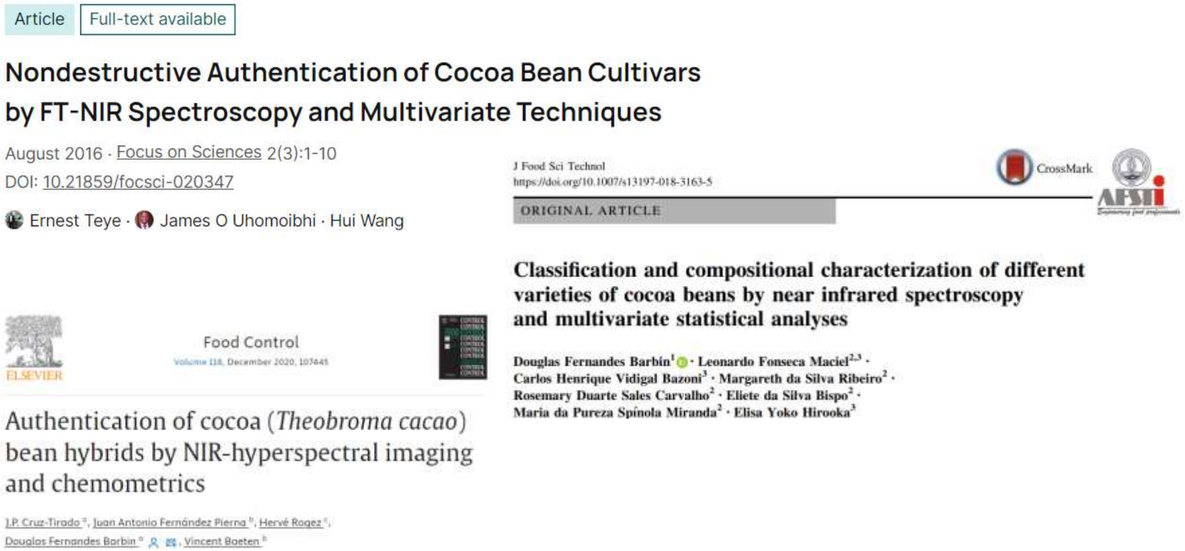
Quality Control of Commercial Cocoa Beans (Theobroma cacao L.) by Near-infrared Spectroscopy

May 2018 · Food Analytical Methods 11(6)

DOI: 10.1007/s12161-017-1137-2

Juliana Hashimoto · Jéssica C. Lima · Renata M. S. Celeghini · Show all 7 authors · Juliana Pallone

Forastero cocoa genotypes from the Eastern Amazonia



Article · Full-text available

Nondestructive Authentication of Cocoa Bean Cultivars by FT-NIR Spectroscopy and Multivariate Techniques

August 2016 · Focus on Sciences 2(3):1-10

DOI: 10.21859/focsci-020347

Ernest Teye · James O Uhomoibhi · Hui Wang

J Food Sci Technol

https://doi.org/10.1007/s13197-018-3163-5

ORIGINAL ARTICLE

Classification and compositional characterization of different varieties of cocoa beans by near infrared spectroscopy and multivariate statistical analyses

Douglas Fernandes Barbin[1] · Leonardo Fonseca Maciel[2,3] · Carlos Henrique Vidigal Bazoni[3] · Margareth da Silva Ribeiro[2] · Rosemary Duarte Sales Carvalho[2] · Eliete da Silva Bispo[2] · Maria da Pureza Spínola Miranda[2] · Elisa Yoko Hirooka[3]

Food Control

Volume 118, December 2020, 107445

Authentication of cocoa (*Theobroma cacao*) bean hybrids by NIR-hyperspectral imaging and chemometrics

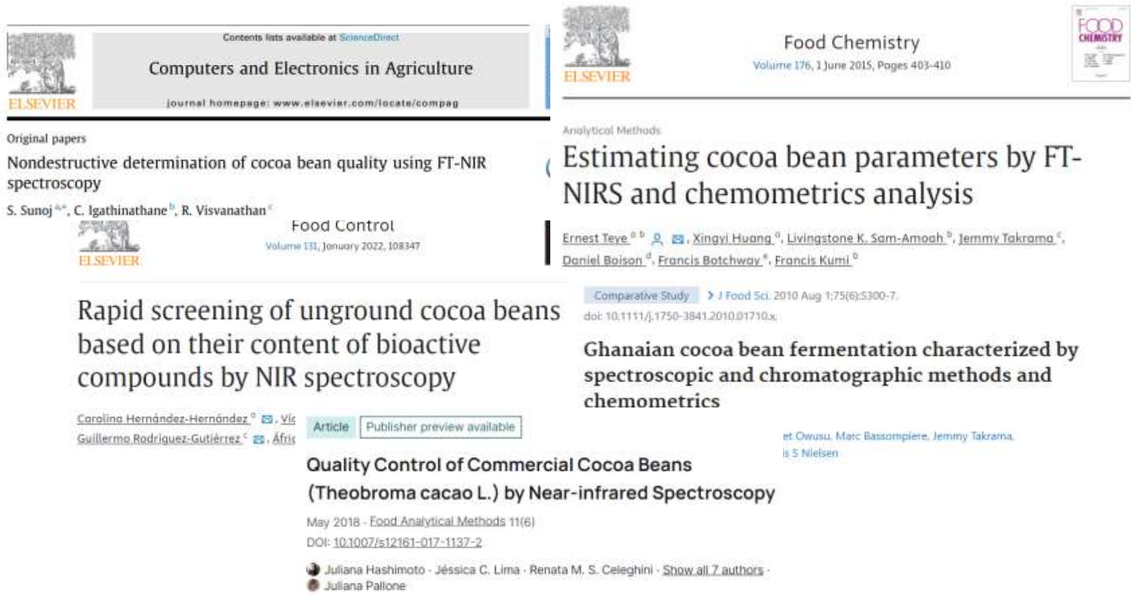
J.P. Cruz-Tirado [a], Juan Antonio Fernández Pierna [b], Hervé Rogez [c], Douglas Fernandes Barbin [a], Vincent Baeten [b]

Objectives of the study

Efficacity of Vis-NIR spectroscopy for the non-destructive discrimination of

Fermented from unfermented beans

Forastero cocoa genotypes from the Eastern Amazonia


Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Nondestructive determination of cocoa bean quality using FT-NIR spectroscopy

S. Sunoj [a,*], C. Igathinathane [b], R. Visvanathan [c]

Food Control

Volume 131, January 2022, 108347

Food Chemistry

Volume 176, 1 June 2015, Pages 403-410

Analytical Methods

Estimating cocoa bean parameters by FT-NIRS and chemometrics analysis

Ernest Teye [a,b], Xingyi Huang [a], Livingstone K. Sam-Amoah [b], Jemmy Takrama [c], Daniel Boison [d], Francis Botchway [e], Francis Kumi [b]

Rapid screening of unground cocoa beans based on their content of bioactive compounds by NIR spectroscopy

Carolina Hernández-Hernández [a], Vis

Guillermo Rodríguez-Gutiérrez [c], Áfris

Comparative Study · J Food Sci. 2010 Aug 1;75(6):S300-7.

doi: 10.1111/j.1750-3841.2010.01710.x.

Ghanaian cocoa bean fermentation characterized by spectroscopic and chromatographic methods and chemometrics

et Owusu, Marc Bassompiere, Jemmy Takrama, is S Nielsen

Article · Publisher preview available

Quality Control of Commercial Cocoa Beans (Theobroma cacao L.) by Near-infrared Spectroscopy

May 2018 · Food Analytical Methods 11(6)

DOI: 10.1007/s12161-017-1137-2

Juliana Hashimoto · Jéssica C. Lima · Renata M. S. Celeghini · Show all 7 authors · Juliana Pallone

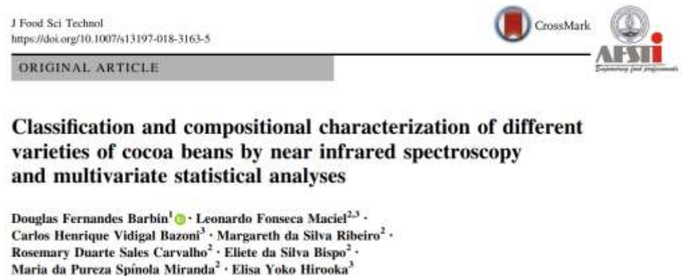

Article · Full-text available

Nondestructive Authentication of Cocoa Bean Cultivars by FT-NIR Spectroscopy and Multivariate Techniques

August 2016 · Focus on Sciences 2(3):1-10

DOI: 10.21859/focsci-020347

Ernest Teye · James O Uhomoibhi · Hui Wang


J Food Sci Technol

https://doi.org/10.1007/s13197-018-3163-5

ORIGINAL ARTICLE

Classification and compositional characterization of different varieties of cocoa beans by near infrared spectroscopy and multivariate statistical analyses

Douglas Fernandes Barbin[1] · Leonardo Fonseca Maciel[2,3] · Carlos Henrique Vidigal Bazoni[3] · Margareth da Silva Ribeiro[2] · Rosemary Duarte Sales Carvalho[2] · Eliete da Silva Bispo[2] · Maria da Pureza Spínola Miranda[2] · Elisa Yoko Hirooka[3]


Food Control

Volume 118, December 2020, 107445

Authentication of cocoa (*Theobroma cacao*) bean hybrids by NIR-hyperspectral imaging and chemometrics

J.P. Cruz-Tirado [a], Juan Antonio Fernández Pierna [b], Hervé Rogez [c], Douglas Fernandes Barbin [a], Vincent Baeten [b]

→ BUT focus on maximum 5 genotypes

Data collection

Sample collection	Sample preparation	Analyses

- 19 Forastero genotypes from Eastern Amazonia (high importance)

Genotype	Origin	Q1	Genetic group	Q2	Genetic group
CA6	Medicilândia	0.3239	Iquitos	0.2041	Nanay
PA169	Tucumã	0.4665	Marañón	0.2535	Amenolado
PA121	Medicilândia	0.9240	Marañón	0.924	Marañón
PA195	Tucumã	0.6744	Marañón	0.6744	Marañón
BE10	Medicilândia	0.3022	Nanay	0.2576	Manañón
CAB499	Tucumã	0.5731	Purús	0.5731	Purús
CCN51	Medicilândia	0.4608	Criollo	0.2930	Iquitos
IMC67	Medicilândia	0.6602	Iquitos	0.6602	Iquitos
CAB324	Tucumã	0.4620	Purús	0.4598	Nanay
CAB214	Medicilândia	0.5386	Purús	0.4155	Contamana
MA11	Tucumã	0.4192	Purús	0.2835	Amelonado
P7	Medicilândia	0.5295	Nanay	0.4257	Contamana
RB36	Tucumã	0.9511	Purús	0.9511	Purús
RB40	Medicilândia	0.8646	Purús	0.8646	Purús
CAB270	Medicilândia	0.3238	Purús	0.2792	Guiana
MO1	Medicilândia	0.3729	Amelonado	0.2650	Purús
CAB208	Medicilândia	0.7415	Purús	0.7415	Purús
MA15	Medicilândia	0.8389	Purús	0.8389	Purús
CAB314	Tucumã	0.4860	Purús	0.3489	Nanay

- 19x70 fruits collected by CEPLAC in July 2020

Data collection

Sample collection

- 19 Forastero genotypes from Eastern Amazonia (high importance)

Genotype	Origin	Q1	Genetic group	Q2	Genetic group
CA6	Medicilândia	0.3239	Iquitos	0.2041	Nanay
PA169	Tucumã	0.4665	Marañón	0.2535	Amenolado
PA121	Medicilândia	0.9240	Marañón	0.924	Marañón
PA195	Tucumã	0.6744	Marañón	0.6744	Marañón
BE10	Medicilândia	0.3022	Nanay	0.2576	Manañón
CAB499	Tucumã	0.5731	Purús	0.5731	Purús
CCN51	Medicilândia	0.4608	Criollo	0.2930	Iquitos
IMC67	Medicilândia	0.6602	Iquitos	0.6602	Iquitos
CAB324	Tucumã	0.4620	Purús	0.4598	Nanay
CAB214	Medicilândia	0.5386	Purús	0.4155	Contamana
MA11	Tucumã	0.4192	Purús	0.2835	Amelonado
P7	Medicilândia	0.5295	Nanay	0.4257	Contamana
RB36	Tucumã	0.9511	Purús	0.9511	Purús
RB40	Medicilândia	0.8646	Purús	0.8646	Purús
CAB270	Medicilândia	0.3238	Purús	0.2792	Guiana
MO1	Medicilândia	0.3729	Amelonado	0.2650	Purús
CAB208	Medicilândia	0.7415	Purús	0.7415	Purús
MA15	Medicilândia	0.8389	Purús	0.8389	Purús
CAB314	Tucumã	0.4860	Purús	0.3489	Nanay

- 19x70 fruits collected by CEPLAC in July 2020

Sample preparation

- +/- 1 kg beans per genotype Unfermented

- +/- 1 kg beans per genotype Fermented (6 days)

- Sun-dried (5 days – moisture < 8%) and stored in refrigerator

Analyses

Data collection

Sample collection

- 19 Forastero genotypes from Eastern Amazonia (high importance)

Genotype	Origin	Q1	Genetic group	Q2	Genetic group
CA6	Medicilândia	0.3239	Iquitos	0.2041	Nanay
PA169	Tucumã	0.4665	Marañón	0.2535	Amenolado
PA121	Medicilândia	0.9240	Marañón	0.924	Marañón
PA195	Tucumã	0.6744	Marañón	0.6744	Marañón
BE10	Medicilândia	0.3022	Nanay	0.2576	Manañón
CAB499	Tucumã	0.5731	Purús	0.5731	Purús
CCN51	Medicilândia	0.4608	Criollo	0.2930	Iquitos
IMC67	Medicilândia	0.6602	Iquitos	0.6602	Iquitos
CAB324	Tucumã	0.4620	Purús	0.4598	Nanay
CAB214	Medicilândia	0.5386	Purús	0.4155	Contamana
MA11	Tucumã	0.4192	Purús	0.2835	Amelonado
P7	Medicilândia	0.5295	Nanay	0.4257	Contamana
RB36	Tucumã	0.9511	Purús	0.9511	Purús
RB40	Medicilândia	0.8646	Purús	0.8646	Purús
CAB270	Medicilândia	0.3238	Purús	0.2792	Guiana
MO1	Medicilândia	0.3729	Amelonado	0.2650	Purús
CAB208	Medicilândia	0.7415	Purús	0.7415	Purús
MA15	Medicilândia	0.8389	Purús	0.8389	Purús
CAB314	Tucumã	0.4860	Purús	0.3489	Nanay

- 19x70 fruits collected by CEPLAC in July 2020

Sample preparation

- +/- 1 kg beans per genotype Unfermented

- +/- 1 kg beans per genotype Fermented (6 days)

- Sun-dried (5 days – moisture < 8%) and stored in refrigerator

Analyses

- Spectral: 10 random beans per genotype (20 for replicates) with Foss XDS 400-2500 nm (each 2 nm)

- Parameters: Standard analytical methods

Data treatment - Fermentation

Exploratory analyses	Datasets preparation	Models
• Mahalanobis distance and Z-score method for outliers – none discarded • 4 preprocessing combinations: SNV, SG (w = 21, d = 1, p = 2), SNV + SG and SG + SNV • PCA of raw and preprocessed spectra		

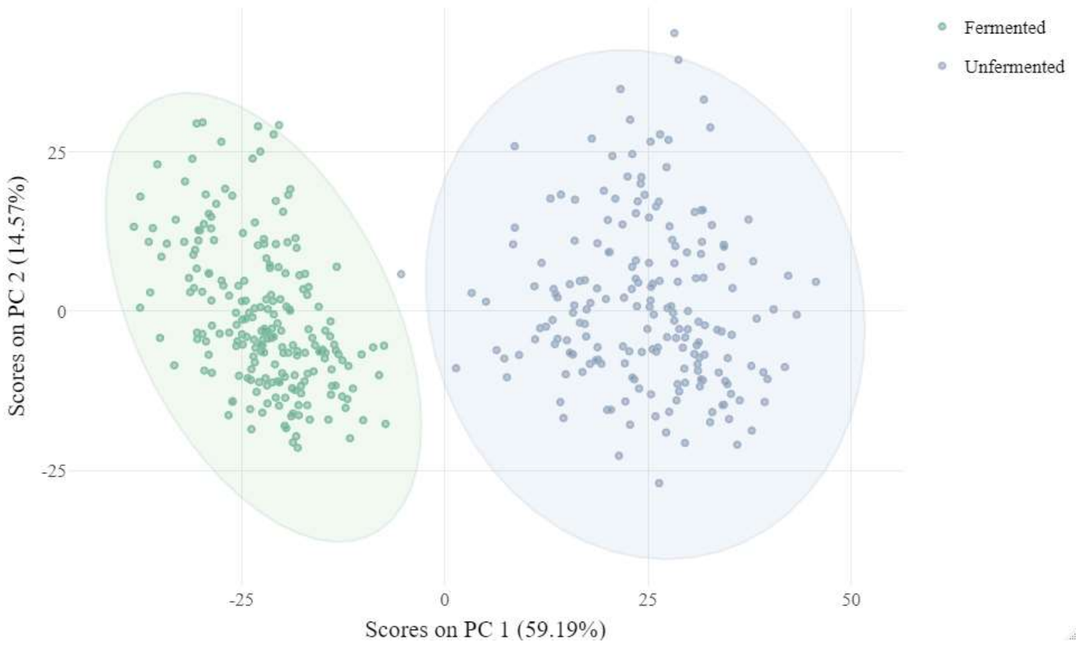
Data treatment - Fermentation

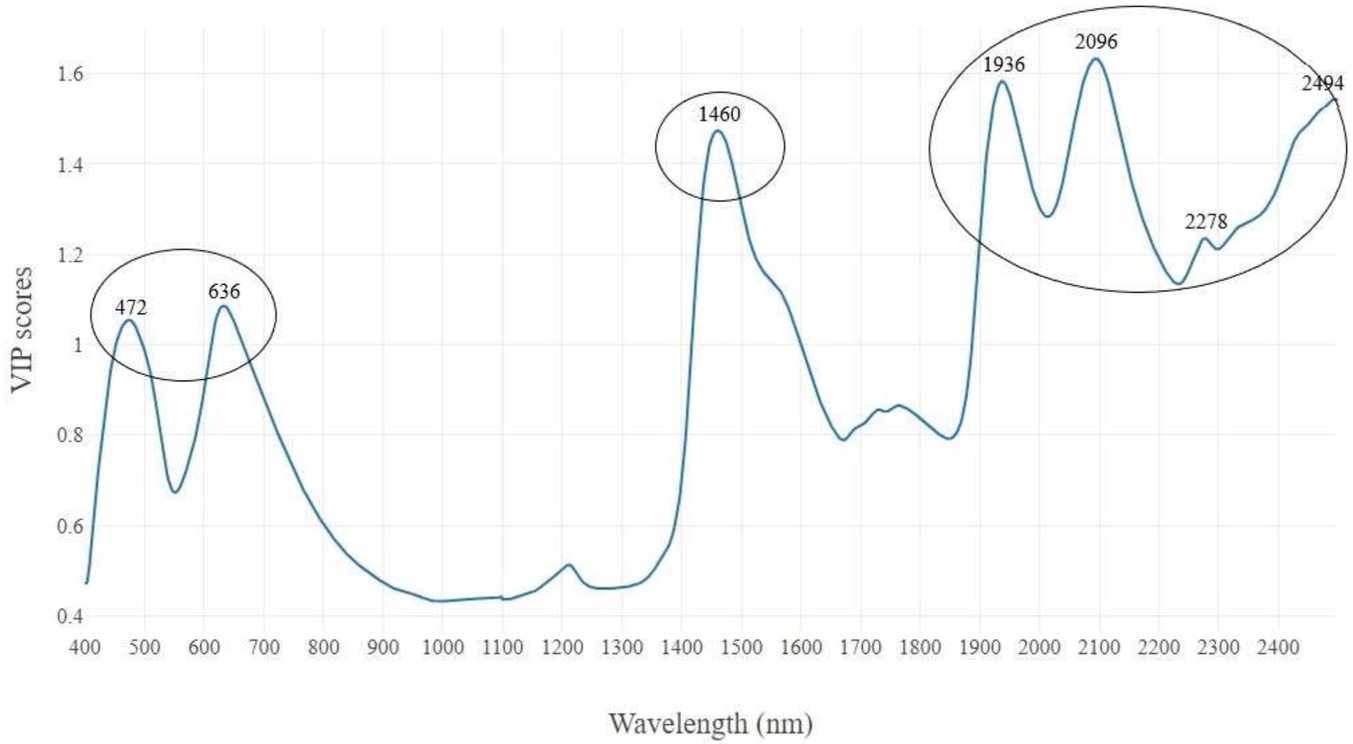
Exploratory analyses	Datasets preparation	Models
• Mahalanobis distance and Z-score method for outliers – none discarded • 4 preprocessing combinations: SNV, SG (w = 21, d = 1, p = 2), SNV + SG and SG + SNV • PCA of raw and preprocessed spectra	• CreateDataPartition for hold-out validation (proportion of fermentation or not in cal/val) • Calibration dataset = 280 beans (147 F, 133 U) • Validation dataset = 120 beans (63 F, 57 U)	

Data treatment – Fermentation

Exploratory analyses	Datasets preparation	Models
• Mahalanobis distance and Z-score method for outliers – none discarded • 4 preprocessing combinations: SNV, SG (w = 21, d = 1, p = 2), SNV + SG and SG + SNV • PCA of raw and preprocessed spectra	• CreateDataPartition for hold-out validation (proportion of fermentation or not in cal/val) • Calibration dataset = 280 beans (147 F, 133 U) • Validation dataset = 120 beans (63 F, 57 U)	• PLS-DA on 400-2500 nm • PLS-DA on specific ranges based on VIP scores (< 1) • LDA on two pairs of wavelengths based on VIP scores • Optimization by CV - AUC • Accuracy, sensitivity, and specificity to evaluate models

Results - Fermentation





	PLS-DA						LDA	
	400-2500 nm	400-700 nm	1400-1600 nm	1900-2500 nm	2000-2250 nm	2250-2350 nm	472 + 636 nm	2096 + 2278 nm
NLV	3	3	2	3	2	2	/	/
Sensitivity	1	1	1	1	1	1	0.937	1
Specificity	1	1	0.984	1	0.984	0.984	0.982	1
Accuracy	1	1	0.992	1	0.992	0.992	0.958	1

Data treatment - Genotypes

Exploratory analyses	Datasets preparation	Models
• Mahalanobis distance and Z-score method for outliers – none discarded • 4 preprocessing combinations: SNV, SG (w = 21, d = 1, p = 2), SNV + SG and SG + SNV • PCA + t-SNE of raw and preprocessed spectra		


Journal of CHEMOMETRICS

RESEARCH ARTICLE

Use of *t*-distributed stochastic neighbour embedding in vibrational spectroscopy

François Stevens ✉ Beatriz Carrasco, Vincent Baeten, Juan A. Fernández Pierna

First published: 23 March 2024 | https://doi.org/10.1002/cem.3544

Data treatment - Genotypes

Exploratory analyses	Datasets preparation	Models
• Mahalanobis distance and Z-score method for outliers – none discarded • 4 preprocessing combinations: SNV, SG (w = 21, d = 1, p = 2), SNV + SG and SG + SNV • PCA + t-SNE of raw and preprocessed spectra	• Split between fermented and unfermented spectra • CreateDataPartition for hold-out validation (proportion of genotypes in cal/val) • Fermented dataset : 210 beans (147 cal – 7/genotypes, 63 val – 3/genotypes) • Unfermented dataset = 190 beans (133 cal – 7/genotypes, 57 val – 3/genotypes)	


Journal of CHEMOMETRICS

RESEARCH ARTICLE

Use of *t*-distributed stochastic neighbour embedding in vibrational spectroscopy

François Stevens ✉ Beatriz Carrasco, Vincent Baeten, Juan A. Fernández Pierna

First published: 23 March 2024 | https://doi.org/10.1002/cem.3544

Data treatment - Genotypes

Exploratory analyses	Datasets preparation	Models
• Mahalanobis distance and Z-score method for outliers – none discarded • 4 preprocessing combinations: SNV, SG (w = 21, d = 1, p = 2), SNV + SG and SG + SNV • PCA + t-SNE of raw and preprocessed spectra	• Split between fermented and unfermented spectra • CreateDataPartition for hold-out validation (proportion of genotypes in cal/val) • Fermented dataset : 210 beans (147 cal – 7/genotypes, 63 val – 3/genotypes) • Unfermented dataset = 190 beans (133 cal – 7/genotypes, 57 val – 3/genotypes)	• PLS-DA on raw + preprocessed spectra • Optimization by CV – Accuracy • Accuracy


Journal of CHEMOMETRICS

RESEARCH ARTICLE

Use of *t*-distributed stochastic neighbour embedding in vibrational spectroscopy

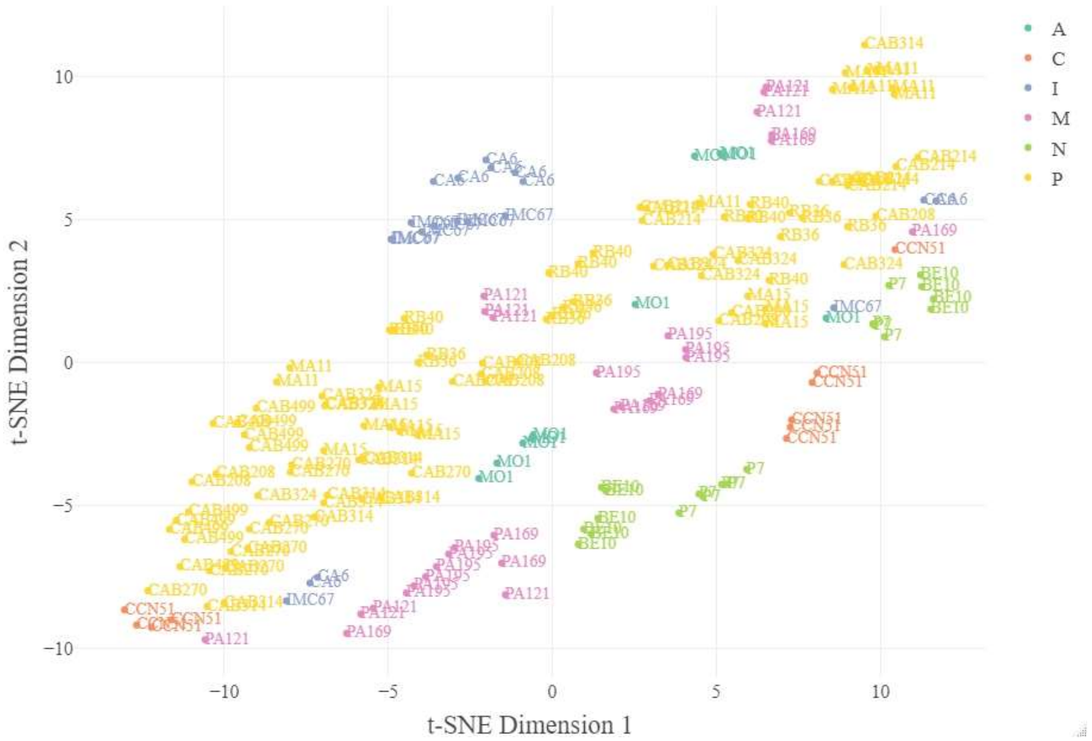
François Stevens ✉ Beatriz Carrasco, Vincent Baeten, Juan A. Fernández Pierna

First published: 23 March 2024 | https://doi.org/10.1002/cem.3544

Results - Genotypes

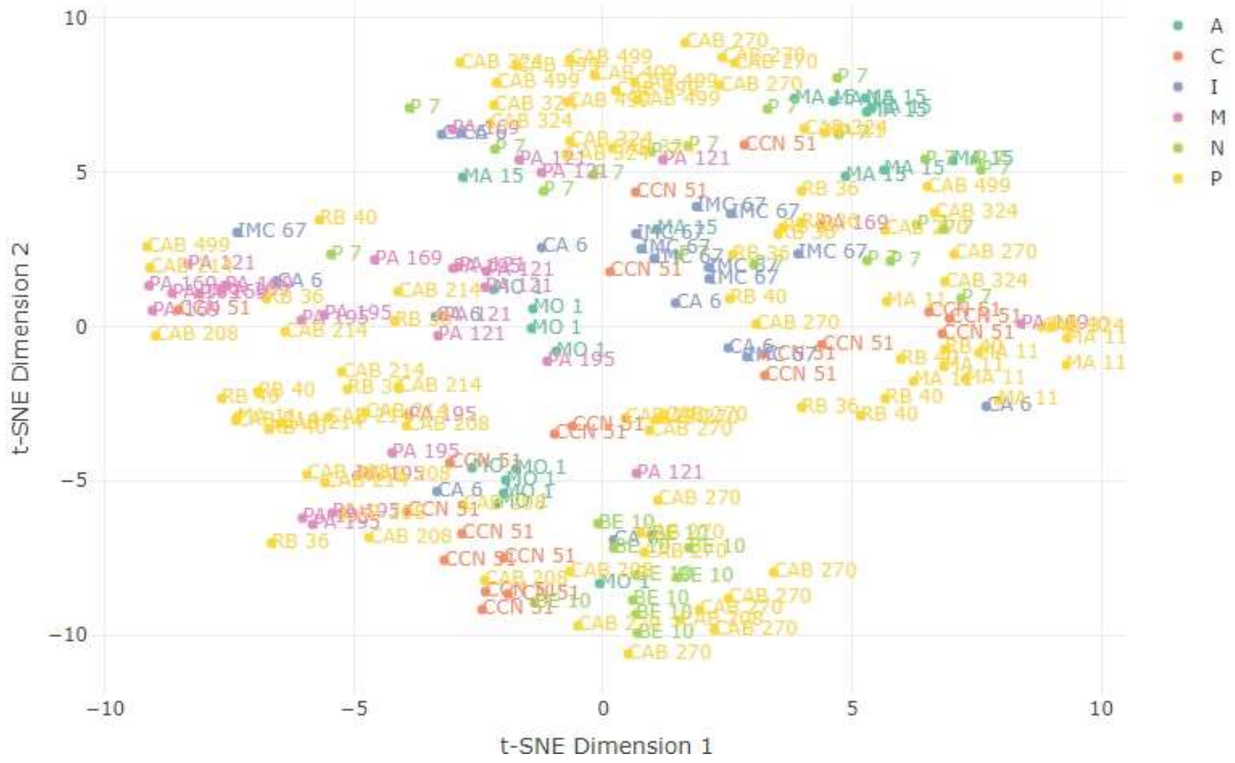
- Unfermented model: Accuracy = 0,84 (Raw)

9 misclassified ⟶ 6 sharing genetic ancestry



- Fermented model: Accuracy = 0,86 (SG)

9 misclassified ⟶ 1 sharing genetic ancestry



Conclusion

Fermented/Unfermented	Genotypes
• Vis and NIR spectral regions for discrimination • High performances with two pairs of wavelengths (472-636 nm and 2096-2278 nm) • Potential for tailored spectral sensors (smaller, cheaper, handheld) 	• Great potential of Vis-NIR for discriminating main cocoa genotypes of Pará • Remain cautious : Additional analyses with more samples • Analyses of unfermented beans seem more reliable

→ Vis-NIR = Practical and efficient solution to address challenges related to quality control and authentication

Acknowledgements

Food Control 164 (2024) 110559

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

Advances in the individual authentication of cocoa beans: Vis/NIR spectroscopy as a tool to distinguish fermented from unfermented beans and classify genotypes in the eastern Amazonia

Anne Pinto [a], Antoine Deryck [b], Giulia Victória Lima [a], Ana Caroline de Oliveira [b], Fabio Gomes Moura [a], Douglas Fernandes Barbin [c], Juan Antonio Fernández Pierna [b], Vincent Baeten [b], Hervé Rogez [a,*]

[a] Center for Valorization of Amazonian Bioactive Compounds (CVACBA), Federal University of Pará (UFPA), Belém, PA, Brazil
[b] Walloon Agricultural Research Center (CRA-W), Knowledge and Valorization of Agricultural Products Department, Quality and Authentication of Products Unit, Gembloux, Belgium
[c] Department of Food Engineering, University of Campinas (UNICAMP), Campinas, SP, Brazil

- Financement: Cooperation project between the Wallonie-Bruxelles International and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES-WBI/2017–2019/361394)
- Donation of the cocoa fruit samples: CEPLAC (Pará, Brazil)
- Grant and financial support: CNPQ (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and PROPESP/UFPA