

ANÁLISIS DE CALIDAD DE CACAO UTILIZANDO ESPECTROSCOPIA DE INFRARROJO CERCANO PORTABLE: DESAFÍOS PARA DIFERENCIACIÓN GEOGRÁFICA

Paola Quesada-Arguedas¹, Carlos Hernández-Aguirre^{2/*},
Alejandro Vargas-Martínez³, Alejandra Mencia-Guevara⁴

Palabras clave: *Theobroma cacao* L.; NIR; análisis multivariado; origen geográfico; poscosecha.

Keywords: *Theobroma cacao* L.; NIR; multivariate analysis; geographic origin; post-harvest.

Recibido: 20/02/24


Aceptado: 28/05/24


RESUMEN

Introducción. La combinación de curvas de calibración generadas por espectroscopía de infrarrojo cercano (NIR) con aprendizaje profundo ofrece una oportunidad para desarrollar métodos de discriminación de la calidad y origen del cacao, apoyando estrategias de valorización territorial y trazabilidad de cacaos diferenciados. **Objetivo.** Validar la aplicabilidad de un método de discriminación de cacaos según su origen geográfico, mediante técnicas de espectroscopía NIR portátil y aprendizaje profundo. **Materiales y métodos.** Se recolectaron 193 muestras de granos de cacao fermentados y secos de diversas regiones de Costa Rica, utilizando 72 muestras para calibración y 121 para predicción. Las muestras se analizaron para determinar


su composición proximal, acidez titulable y compuestos fenólicos. Se empleó un espectrofotómetro NIR para recopilar datos espectrales (400-1700 nm). El preprocesamiento de los datos espectrales permitió desarrollar modelos de regresión para predecir características químicas. Para el modelo de clasificación geográfica, se eliminó el ruido de los espectros y se realizó un análisis de conglomerados, usando la distancia Gower y el método de agrupamiento Ward sobre los componentes obtenidos mediante Análisis de Componentes Principales (ACP). **Resultados.** La grasa fue el principal componente presente en el conjunto de muestras (>39,67%). El análisis espectral demostró que el NIR puede diferenciar el cacao según el grado de fermentación y el contenido de compuestos fenólicos. El modelo de regresión de mínimos cuadrados parciales

* Autor para correspondencia. Correo electrónico: carlos.hernandez.aguirre@una.cr

1 Estudiante de Licenciatura en Ingeniería en Agronomía, Universidad Nacional, Heredia, Costa Rica.
 0009-0009-2497-7717.

2 Laboratorio de Calidad e Innovación Agroalimentaria (LABCIAGRO), Universidad Nacional, Heredia, Costa Rica.
 0000-0002-0059-8028.

3 Escuela de Ciencias Agrarias, Universidad Nacional, Heredia, Costa Rica.
 0000-0001-8039-8984.

4 Laboratorio de Calidad e Innovación Agroalimentaria (LABCIAGRO), Universidad Nacional, Heredia, Costa Rica.
 0000-0003-2449-0510.

extendida (XLS) mostró la mejor capacidad predictiva para las propiedades químicas. El agrupamiento por origen geográfico identificó cuatro grupos influenciados principalmente por propiedades químicas relacionadas con prácticas poscosecha. **Conclusión.** El modelo de regresión lineal empleado demostró ser superior en la predicción de características químicas proximales. Se observó que la limitada diversidad genética y las prácticas poscosecha estandarizadas podrían reducir la variabilidad de calidad asociada al origen geográfico, limitando la utilidad del NIR en la identificación del origen y la trazabilidad. Se sugiere explorar espectros más amplios y equipos adicionales para análisis multivariados avanzados.

ABSTRACT

Quality analysis of cocoa using portable Near-Infrared Spectroscopy (NIR): Challenges towards geographical differentiation. Introduction. The combination of calibration curves generated by Near-Infrared (NIR) spectroscopy with deep learning offers an opportunity to develop methods for discriminating the quality and origin of cocoa, supporting strategies for territorial valorization and traceability of differentiated cocoas. **Objective.** To validate the applicability of a method for discriminating cocoa according to its geographical origin using portable NIR spectroscopy and deep learning techniques. **Materials and methods.** A total of 193 samples

of fermented and dried cocoa beans from different regions of Costa Rica were collected, using 72 samples for calibration and 121 for prediction. The samples were analyzed to determine their proximal composition, titratable acidity, and phenolic compounds. A NIR spectrophotometer was used to collect spectral data (400-1700 nm). Spectral data preprocessing allowed for the development of regression models to predict chemical characteristics. For the geographical classification model, noise was removed from the spectra, and cluster analysis was performed using the Gower distance and Ward's clustering method on components obtained through Principal Component Analysis (PCA). **Results.** Fat was the main component present in the sample set (>39.67%). Spectral analysis demonstrated that NIR can differentiate cocoa based on the degree of fermentation and phenolic compound content. The extended partial least squares (XLS) regression model showed the best predictive capacity for chemical properties. Clustering by geographical origin identified four groups mainly influenced by chemical properties related to post-harvest practices. **Conclusion.** The linear regression model used proved superior in predicting proximal chemical characteristics. It was observed that limited genetic diversity and standardized post-harvest practices could reduce the quality variability associated with geographical origin, limiting the utility of NIR in origin identification and traceability. It is suggested to explore broader spectra and additional equipment for advanced multivariate analyses.

INTRODUCCIÓN

El cacao (*Theobroma cacao* L.) es un cultivo de importancia económica y social en comunidades rurales del trópico y forma parte de un sistema agroalimentario fundamental en la evolución histórica, cultural y comercial de Centroamérica, vinculando alrededor de 15 000

familias que habitan zonas remotas y de difícil acceso (Orozco Aguilar *et al.* 2007). La región mesoamericana alberga una enorme riqueza de variabilidad genética y es fuente de cacaos con características especiales de sabores y aromas particulares (Zarrillo *et al.* 2018), los cuales se definen como “cacao fino y de aroma” de acuerdo con la Organización Internacional de Cacao.

Se reconocen dos clases de calidades comerciales, cacao a granel “bulk” (calidad estándar) atribuido particularmente al tipo Forastero, y cacao fino (calidad alta) atribuido a los tipos Criollo y Trinitario. Sin embargo, diversos estudios muestran que independientemente del tipo de cacao, las prácticas postcosecha (fermentación y secado) tienen un papel central en la generación de la heterogeneidad en el perfil de calidad (Papalexandratou *et al.* 2019, Verce *et al.* 2021). Además, las condiciones geográficas y agronómicas así como el tipo de suelo, clima, y madurez de la mazorca también influyen en el perfil aromático. Así, los cacaos de distintos orígenes geográficos tienen diferentes características organolépticas que influyen en la calidad del chocolate.

La diferenciación geográfica registrada para granos de cacao es poco común, a causa de una asociación entre la calidad del chocolate y el lugar de origen de las fábricas y centros de transformación final, y no la geografía vinculada a los granos de cacao como materia prima. De forma que, los cacaos mesoamericanos se encuentran desprotegidos ante un uso inadecuado e informal de su procedencia, lo cual destaca la importancia de estudios técnicos en diferentes territorios con producción de cacao que permitan evidenciar una correlación entre calidad y origen geográfico (Hernandez y Granados 2021).

La espectroscopía de infrarrojo cercano (NIR) se posiciona como una técnica rápida y sencilla para evaluar alimentos y productos agrícolas. El principio de esta técnica se basa en la absorción de radiación electromagnética en longitudes de onda de 780 a 2500 nm en una matriz determinada. Sus principales ventajas: 1) la rapidez en la que realiza una medición (entre 15 y 90 segundos), 2) el procedimiento no destructivo de la muestra y 3) el nulo uso de reactivos químicos.

La espectroscopía NIR se ha aplicado en el análisis cualitativo y cuantitativo de granos de cacao: discriminación de granos por el grado de fermentación (Teye *et al.* 2014); análisis de los niveles de NH_3 en la determinación del origen geográfico (Hue *et al.* 2014); análisis de pH,

humedad, índice de fermentación y compuestos fenólicos totales (Anyidoho *et al.* 2021, Kutsanedzie *et al.* 2017, Sunoj *et al.* 2016, Teye 2022, Teye *et al.* 2015); estimación de compuestos químicos proximales como proteína, grasa, cenizas y carbohidratos totales (Barbin *et al.* 2018, Forte *et al.* 2023); incidencia de moho (Kutsanedzie *et al.* 2018); clasificación geográfica (Anyidoho *et al.* 2020), y diferenciación entre granos de orgánicos y convencionales (Anyidoho *et al.* 2021).

Todos los estudios anteriores se han basado en la implementación de enfoques de machine learning para la reducción del ruido espectral y el desarrollo de modelos de predicción. Para obtener la información de los datos espectrales se han aplicado técnicas de análisis quimiométrico, incluidos los mínimos cuadrados parciales (PLS), redes neuronales artificiales (ANN), y máquina de vectores de soporte (SVM).

La combinación de curvas de calibración generadas por espectroscopía NIR con aprendizaje profundo supone una oportunidad para crear métodos de discriminación de calidad y origen del cacao, apoyar estrategias de valorización territorial y gestionar sistemas de trazabilidad de productos diferenciados. Por lo tanto, el objetivo de este trabajo fue validar de la aplicabilidad de un método de discriminación de cacaos según su origen geográfico, mediante el uso de técnicas de espectroscopía NIR portátil y métodos de machine learning. Adicionalmente, se evaluó la aplicación del NIR en la predicción de características de calidad mediante la construcción de modelos generados por análisis proximal y cromatografía líquida de alta resolución (HPLC).

MATERIALES Y MÉTODOS

Muestras

Los granos de cacao fermentados y secos se obtuvieron de territorios de Costa Rica asociados a la producción de cacao fino y de aroma, los cuales, para motivos de este estudio, se agruparon de la siguiente manera: Región Huetar Norte (Upala, Guatuso, Los Chiles, Sarapiquí, Pococí,

Guácimo, Matina, Limón, Valle de la Estrella), Región Caribe (Talamanca, Puerto Viejo, Sixoala), Región Central (Turrialba), Región Brunca (Valle del General, Coto Brus, Ciudad Neily, Laurel, Paso Canoas), y Centroamérica (Nicaragua, Panamá y El Salvador) durante el periodo 2022 - 2023. Para cada muestra se registró: procedencia, tipología genética, tipo de fermentación. Dentro de los tipos de fermentación se encontró los siguientes: 1) sistemas artesanales con 5-6 días de fermentación en cajones de madera, y 2) sistemas industriales con volúmenes superiores a los 200 kg por lote y 6 días de fermentación en cajones de madera tipo escalera.

Se incluyeron 72 muestras en el conjunto de calibración y 121 muestras en el grupo de predicción. El conjunto de calibración se utilizó para desarrollar el modelo de predicción de las características químicas (Tabla 1), mientras que el conjunto de predicción se utilizó para evaluar el desempeño de los modelos desarrollados.

Tabla 1. Características químicas evaluadas en los granos de cacao utilizados para el desarrollo de la investigación.

| Variable | Unidad | Referencia |
|------------------|-----------------------|-----------------------------|
| Fibra cruda | % | ANKOM Technology (2006) |
| Proteína | mg mL ⁻¹ | AOAC (2005a) |
| Humedad | % | AOAC (2005b) |
| Cenizas | % | AOAC (2005c) |
| Grasa total | % | AOAC (2005d) |
| Acidez titulable | meq 100 ⁻¹ | Hinneh <i>et al.</i> (2018) |
| Epicatequina | mg g ⁻¹ | Rojas (2020) |
| Teobromina | mg g ⁻¹ | Rojas (2020) |
| Cafeína | mg g ⁻¹ | Rojas (2020) |

Instrumento e información espectral

El espectro de cada muestra se recopiló mediante el espectrofotómetro de infrarrojo cercano (BÜCHI, Suiza). Todas las muestras se procesaron en modo rotatorio durante 15 segundos y los escaneos abarcaron un rango de espectro de luz visible + NIR (400 - 1700 nm) con dos nanómetros de resolución (650 puntos). Se utilizaron dos lámparas distintas denominadas como vista superior o vista inferior del equipo, para los escaneos del conjunto de calibración. Para el conjunto de predicción se utilizó solamente la vista superior. El espectrómetro portátil proporcionó las medidas en partes de absorbancia proporcionales (log 1/R), donde R representa la reflectancia.

Conjunto de calibración

Se utilizaron 72 muestras para la creación del modelo de análisis proximal. Se procesaron 300 g de granos enteros en una licuadora durante 45 segundos (tres intervalos de licuado de cinco segundos con pausas de 10 segundos) con la mínima intensidad. No se realizó descascarillado de los granos ni tamizado del polvo resultante del licuado. Se colocaron 250 g de muestra procesada para la medición con la vista inferior y 10 g de la misma muestra para realizar la medición con la vista superior.

Conjunto de predicción

Se procesaron 50 g de granos enteros en una licuadora durante 45 segundos (tres intervalos de licuado de cinco segundos con pausas de 10 segundos) con la mínima intensidad. No se realizó descascarillado de los granos ni tamizado del polvo resultante del licuado. Se colocaron 10 g de cacao para obtener los espectros según la vista superior del equipo.

Análisis proximal

Se tomaron los 250 g de 72 muestras de granos de cacao molidos utilizados en los escaneos previos para realizar análisis proximales de

fibra cruda (ANKOM 2006), proteína (AOAC 2005a), humedad (AOAC 2005b), cenizas (AOAC 2005c) y grasa total (AOAC 2005d).

Compuestos fenólicos

El análisis de cromatografía líquida (HPLC) se elaboró siguiendo la metodología descrita por Rojas (2020), donde los estándares químicos utilizados para comparar los tiempos de retención y las longitudes de onda con las muestras seleccionadas fueron: cafeína, (-) epicatequina y teobromina (Sigma Aldrich, Estados Unidos). Los solventes metanol (MeOH), acetonitrilo y ácido acético (CH₃-COOH) fueron grado HPLC (Sigma Aldrich, Estados Unidos). El agua desionizada se obtuvo mediante el sistema Millipore Alpha Q.

Los granos de cacao destinados para el análisis del epicatequina y metilxantinas (teobromina y cafeína) fueron liofilizados (OPERON modelo FDU-7006) durante 48 horas, bajo las condiciones de 0 atm y -83°C. Seguidamente, se les removió la cáscara y fueron molidos con un mortero, posteriormente se almacenaron en un congelador a -20°C.

Para la extracción de grasas, se tomaron 2 g de cacao molido en un tubo plástico cónico, se agregaron 10 ml de hexano y se colocaron en baño ultrasónico a 40°C por 20 minutos y se sometieron a centrifugación a 25°C y 6607 RCF rpm durante 10 minutos y se descartó el sobrenadante. La muestra fue extraída con 10 ml de metanol-agua a una proporción de 15:6. Se colocó en baño ultrasónico a 40°C por 20 minutos, se centrifugó a 4807 RCF por 10 minutos, y el sobrenadante fue vertido en un falcon de 50 ml.

El procedimiento se ejecutó tres veces consecutivas, se combinaron las tres extracciones y se aforaron (con el mismo solvente de extracción) a un volumen de 40 ml. El sobrenadante, se volvió a centrifugar por 10 min a 25°C y 6607 RCF, y se filtró a través de filtros de 0,45 µm (Sartorius stedim biotech). Finalmente, se incorporaron en viales para HPLC.

El análisis cromatográfico se efectuó mediante el sistema HPLC Agilent serie 1200

(Agilent Technologies, Estados Unidos), el cual consta de un muestreador automático para viales de 2 mL, una bomba cuaternaria, un detector de diodos (DAD) y un detector ultravioleta (UV). Los componentes fueron separados utilizando una columna Zorbax Eclipse Plus C18 de 4,6 x 100 mm y 3,5 µm (Agilent Technologies, Estados Unidos).

El sistema solvente aplicado consistió en: eluyente A (agua desionizada con 1% ácido acético) y eluyente B (acetonitrilo). El gradiente de elución se programó con un flujo de 1 mL min⁻¹, de la siguiente manera: 100% A por 10 minutos, 75% A por cinco minutos, 50% A por dos minutos, 0% A por un minuto; con un post-acondicionamiento de un minuto. La detección se realizó con el DAD a longitudes de onda de λ 275,4 nm y λ 325,5 nm.

La identificación de los compuestos se realizó con patrones de cafeína, (-) epicatequina, y teobromina utilizando las mismas condiciones descritas del método de HPLC. Los resultados fueron expresados en mili unidades de absorbancia por segundo (mUA s⁻¹), cuyos valores medios (±) de las desviaciones estándar representan un valor aproximado del contenido de tales compuestos. Sin embargo, esto no alude con exactitud la concentración por gramo de cacao desgrasado.

Acidez titulable

Siguiendo la metodología realizada por Hinneh *et al.* (2018), se molieron 10 g de granos enteros usando una licuadora, los cuales fueron añadidos a un erlenmeyer con 90 mL de agua en punto de ebullición. El homogeneizado se filtró y se enfrió a 20–25°C. A continuación, se tituló una alícuota de 50 mL hasta un punto final de pH de 8,1 con NaOH 0,1 N (Sigma Aldrich, Estados Unidos). En el titulador, la acidez se expresó en meq 100⁻¹ (%). Es importante señalar que este procedimiento no fue para cuantificar el pH real del grano de cacao en sí, sino para medir la acidez derivada de los ácidos del grano que se difunden en el agua.

Preprocesamiento de la información espectral

El análisis quimiométrico se realizó con el programa NIRWise Plus para Windows 10. Se realizó un preprocesamiento de la información espectral para mejorar el conjunto de datos sin procesar y reducir el ruido en los espectros con el método Escalado y Centrado (SNV), el cual resta el promedio y divide por la desviación estándar del vector. El pretratamiento estándar normal variable elimina las interferencias multiplicativas de la dispersión, el tamaño de las partículas y el cambio de la distancia de la luz, por lo que es altamente recomendable para muestras sólidas (Teye *et al.* 2020). En este caso, se calculó la primera derivada mediante cinco puntos de datos y una ventana de suavizado de cero puntos de datos (BÜCHI s.f.).

Modelos de regresión lineal de predicción para características químicas

Los modelos de calibración para las características químicas estimadas fueron desarrollados con el programa NIRWise Plus para Windows 10 utilizando tres tipos de algoritmos de regresión: regresión de mínimos cuadrados parciales extendida (XLS), regresión de mínimos cuadrados parciales (PLS) y regresión de componentes principales (PCR).

Se realizó una validación cruzada utilizando las 72 muestras de calibración, en donde el algoritmo seleccionó un 80% de dichas muestras como conjunto de calibración y el 20% restante como conjunto de predicción. Para la evaluación de los resultados de la validación individual, se tomaron en cuenta los siguientes estadísticos: R^2 de calibración, error estándar de la calibración (SEC), R^2 de predicción y error estándar de la validación cruzada (SECV). Esto permitió la comparación de diferentes modelos de predicción.

Modelo de clasificación geográfica

Se utilizó el programa Rstudio para eliminar el ruido de las muestras y corregir debilidades de señal. Se aplicaron técnicas de Escalado y

Centrado (SNV), el método de la primer y segunda derivada (D1 y D2), corrección de señal multiplicativa extendida (EMSC) y corrección de dispersión multiplicativa (MSC) y sus combinaciones bajo distinto orden de aplicación. Las derivadas se calcularon a partir del algoritmo de Savitzky-Golay.

Se aplicó un análisis de componentes principales (ACP) para reducir la dimensionalidad de 499 longitudes de onda a la cantidad de componentes donde el porcentaje de varianza acumulada explicase alrededor del 90%.

Se utilizaron las dimensiones obtenidas en ACP y se procedió a realizar un análisis de conglomerados utilizando la distancia Gower, método de agrupamiento Ward y se tomó como criterio para la determinación del número de grupos una línea de corte al 33% del rango de distancia observadas. Además, con los grupos formados, se hicieron tablas de contingencia para determinar la asociación ($p < 0,05$) con las variables “Tipo de producción”, “Región” y “Genética”. Con las variables que fueron significativas se hizo un análisis de correspondencias múltiples para lograr describir los grupos.

RESULTADOS Y DISCUSIÓN

Propiedades químicas proximales

Los parámetros descriptivos de las propiedades químicas proximales de cacao seco se muestran en la Tabla 2. La grasa fue el mayor constituyente, en términos porcentuales, del conjunto de muestras, seguido por la proteína. Indicadores tales como epicatequina, la relación teobromina/cafeína, acidez y humedad, mostraron mayor peso en cuanto al coeficiente de variación. Lo anterior apunta a la relevancia del proceso poscosecha en la generación de variabilidad, y resalta la importancia de dichas variables en los criterios de discriminación de muestras según su origen. Estos resultados coincidieron con los publicados por Forte *et al.* (2023) donde de igual forma se encuentra mayor peso en el coeficiente de variación asociado a indicadores de la fermentación y el manejo poscosecha.

Tabla 2. Resultados de los parámetros químicos estimados en granos de cacao secos durante el desarrollo de la investigación.

| | Mínimo | Máximo | Media | Desviación estándar | Coefficiente de variación (%) |
|------------------------------------|--------|--------|--------|---------------------|-------------------------------|
| Acidez (meq 100 ⁻¹) | 0,78 | 14,35 | 7,354 | 2,573 | 34,980 |
| Humedad (m m ⁻¹ ; %) | 4,34 | 12,70 | 7,761 | 1,595 | 20,552 |
| Grasa (m m ⁻¹ ; %) | 39,67 | 56,50 | 47,991 | 3,428 | 7,142 |
| Proteína (m m ⁻¹ ; %) | 8,83 | 21,43 | 13,114 | 1,633 | 12,450 |
| Cenizas (m m ⁻¹ ; %) | 2,24 | 4,62 | 3,539 | 0,371 | 10,489 |
| Relación teobromina/cafeína | 1,14 | 18,38 | 6,037 | 3,015 | 49,939 |
| Epicatequina (mg g ⁻¹) | -1,20 | 9,62 | 3,143 | 1,750 | 55,684 |

Nota: meq = miliequivalentes, % m/m = porcentaje masa/masa.

Características espectrales de los granos de cacao secos

El conjunto de espectros crudos de 193 muestras de cacao seco se presenta en la Figura 1A. El comportamiento espectral observado en las longitudes de onda de 400 a 1700 nm es similar al reportado por otros autores (Forte *et al.* 2023, Anyidoho *et al.* 2021). Se aplicó el preprocesamiento quimiométrico de Escalado y Centrado (SNV) a dicho espectro para obtener solo las características útiles de las muestras y construir un modelo confiable. El espectro que resultó de dicho análisis se presenta en la Figura 1B. En ambos gráficos es notable que el espectro visible de menores longitudes de onda (400 – 500 nm) es la región donde se generó mayor ruido.

Los espectros claves fueron en los rangos de 1150-1250, y 1400-1500 nm, correspondientes a grupos funcionales O-H, C-H, C-H₂, C-H₃, NH₂, los cuales están a su vez asociados con carbohidratos complejos, agua, alcoholes alifáticos, grasas, polifenoles, ácidos orgánicos y anillos bencénicos. Específicamente, el contenido de grasa está relacionado con la banda de absorción alrededor de 1200 nm (Hayati *et al.* 2021), mientras que el contenido de humedad ha sido reportado en las regiones de 1460-1490 nm (Caporaso *et al.* 2018).

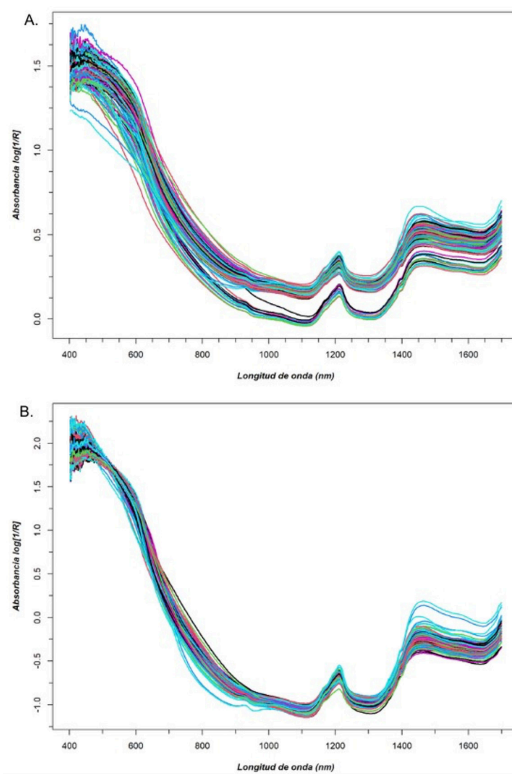


Figura 1. Espectro crudo (A) y espectro posterior a la aplicación del pretratamiento SNV (B) de las muestras de cacao seco analizadas durante el desarrollo de la investigación.

El rango 1200 – 1460 nm está relacionado con C-H, C-H₂, C-H₃, lo cual se asocia a compuestos con anillos bencénicos y fenólicos (tales como epicatequina). De esta manera, este estudio demuestra una alta capacidad del NIR para diferenciar cacao según su grado de fermentación, en respuesta a la variación de epicatequina y polifenoles, los cuales a su vez son indicadores de acidez, amargor y astringencia observado por Anyidoho *et al.* (2021).

Modelos de regresión lineal para la predicción de parámetros químicos proximales

Los modelos de predicción fueron establecidos con la variable independiente (x) dada por

el espectro crudo y con la variable dependiente (y) dada por los contenidos respectivos de cada propiedad. Los estadísticos de cada modelo de predicción para las propiedades químicas se muestran en la Tabla 3. El modelo que obtuvo los valores máximos de coeficiente de correlación para la totalidad de las propiedades químicas fue XLS, en donde la propiedad de humedad presentó el valor máximo ($R^2 = 0,922$) y grasa el valor mínimo ($R^2 = 0,811$), bajo este modelo de regresión lineal. Asimismo, la Figura 2 ilustra los gráficos de regresión lineal para cada una de las propiedades proximales.

Tabla 3. Rendimiento de tres modelos de regresión lineal para siete propiedades medidas utilizando la vista inferior del NIR en cacao seco durante el desarrollo de la investigación.

| Método de regresión | Estadísticos | Acidez | Humedad | Proteína | Grasa | Cenizas | Epicatequina | Teobromina/ cafeína |
|--|--------------|--------|---------|----------|-------|---------|--------------|------------------------|
| Regresión de mínimos cuadrados parciales (PLS) | SEC | 0,825 | 0,447 | 0,752 | 1,311 | 0,087 | 0,635 | 0,495 |
| | R2 | 0,845 | 0,863 | 0,622 | 0,872 | 0,845 | 0,857 | 0,847 |
| | SECV | 1,505 | 0,721 | 1,536 | 2,315 | 0,215 | 0,756 | 1,241 |
| Regresión de componentes principales (PCR) | R2 | 0,543 | 0,676 | 0,031 | 0,635 | 0,297 | 0,793 | 0,297 |
| | SEC | 1,251 | 0,566 | 0,648 | 2,465 | 0,186 | 0,672 | 1,164 |
| | R2 | 0,628 | 0,806 | 0,759 | 0,576 | 0,293 | 0,840 | 0,405 |
| Regresión de mínimos cuadrados parciales extendida (XLS) | SECV | 1,675 | 0,654 | 1,235 | 3,177 | 0,222 | 0,785 | 1,533 |
| | R2 | 0,400 | 0,747 | 0,475 | 0,353 | 0,089 | 0,786 | 0,095 |
| | SEC | 0,636 | 0,358 | 0,500 | 1,643 | 0,075 | 0,521 | 0,460 |
| | R2 | 0,902 | 0,922 | 0,822 | 0,811 | 0,885 | 0,904 | 0,883 |
| Regresión de mínimos cuadrados parciales extendida (XLS) | SECV | 0,749 | 0,438 | 0,583 | 2,009 | 0,087 | 0,588 | 0,519 |
| | R2 | 0,865 | 0,888 | 0,759 | 0,722 | 0,848 | 0,878 | 0,853 |

Nota: SEC = error estándar de calibración, R^2 = coeficiente de correlación de la regresión, SECV = error estándar de la validación cruzada.

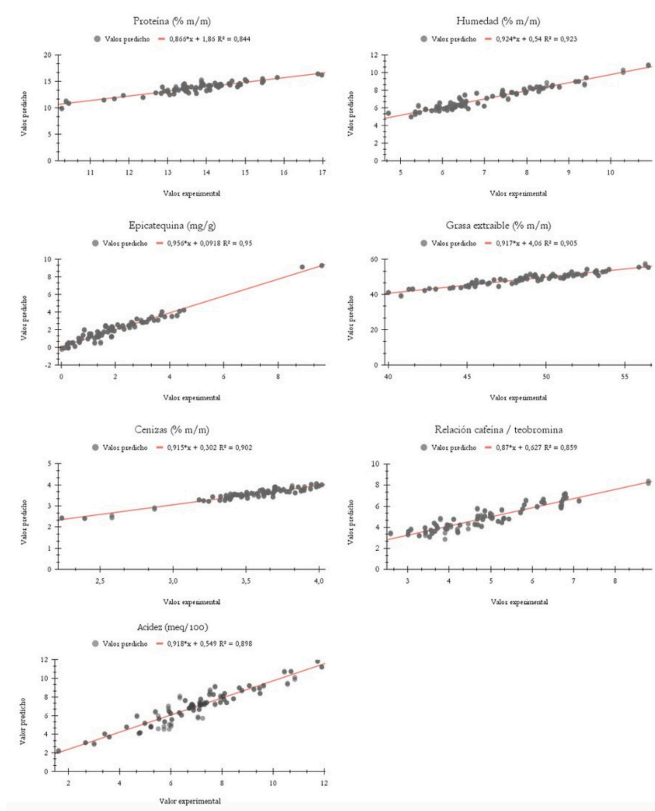


Figura 2. Curvas de calibración del modelo XLS para los parámetros químicos analizados.

Aplicación del NIR como herramienta de predicción de origen

El análisis de Componentes Principales (ACP) utilizó 499 lecturas de longitudes a lo largo del espectro implementado (400-1700 nm), reduciendo la dimensionalidad a solamente tres componentes que lograron explicar el 93% de la variabilidad de los datos. Los resultados del agrupamiento basado en un modelo de Clasificación No Supervisada (CNS), utilizando las dimensiones seleccionadas, se muestran mediante un dendrograma en la Figura 3, donde se evidenciaron al menos cuatro grupos a una altura de 15 en el eje Y. Se graficaron dichos grupos mediante la asignación de color en la Figura 4.

En la Tabla 4 se observa la distribución de muestras por origen, según el grupo establecido previamente. El grupo 1 y 2 presentan en su mayoría muestras del Caribe y Región Huetar Norte; en el grupo 3 se agrupan en su mayoría las muestras de la Región Central, Caribe y Región Huetar Norte; y en el grupo 4 se agrupan menor cantidad de todos los orígenes. En un esfuerzo por comprender el peso de propiedades químicas en el agrupamiento por origen, se observa que el grupo 1 incluye muestras con un alto potencial de calidad atribuido a altos contenidos de grasa, así como valores bajos en la relación teobromina/cafeína y en el contenido de epicatequina. En este grupo se observan muestras predominantemente procesadas de manera artesanal, con genéticas basadas en híbridos trinitarios y clones CATIE.

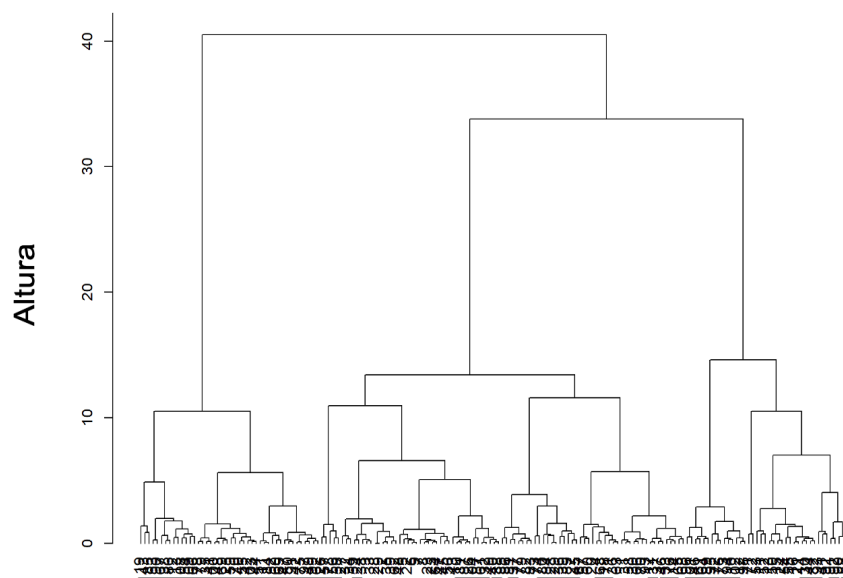


Figura 3. Dendrograma de agrupación a partir de los componentes principales obtenidos de las muestras de cacao seco analizadas durante el desarrollo de la investigación.

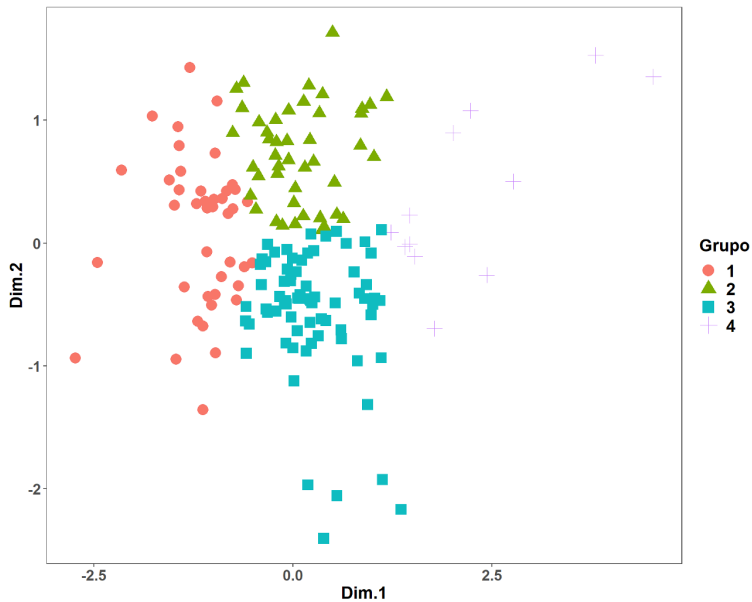


Figura 4. Agrupación de las muestras de cacao seco analizadas durante el desarrollo de la investigación según los resultados del dendrograma.

Tabla 4. Asociación de los grupos formados y el origen geográfico de las muestras de cacao seco analizadas durante el desarrollo de la investigación.

| Región | Grupo | | | |
|---------------|-------|----|----|---|
| | 1 | 2 | 3 | 4 |
| Brunca | 3 | 7 | 14 | 2 |
| Caribe | 22 | 22 | 26 | 3 |
| Huetar Norte | 9 | 13 | 10 | 3 |
| Centroamérica | 4 | 0 | 6 | 0 |
| Central | 4 | 3 | 18 | 4 |

En el grupo 2 se incluye un amplio número de muestras con altos valores de acidez y humedad, evidenciando algunas carencias en el manejo poscosecha. En el grupo 3 las muestras presentan menor humedad y acidez, así como bajos valores de epicatequina, con genéticas diversas incluyendo cacaos acriollados e híbridos. El grupo 4 incluye muestras de calidad inferior, reflejada en altos valores de acidez, humedad y epicatequina.

La débil agrupación que se obtuvo en este estudio demostró que las prácticas poscosecha contribuyen mayormente a la diferenciación de calidad en comparación con el origen geográfico de las muestras, un aspecto que ha sido revisado recientemente por autores (D’Souza *et al.* 2017, Fanning *et al.* 2023).

De manera similar Anyidoho *et al.* (2020) realizaron un estudio para la discriminación por origen en consideración de cuatro países en África, atribuyendo los criterios de segregación a propiedades fisicoquímicas, aunque no demuestran con claridad un agrupamiento por origen. Se demostró el potencial de dicha herramienta en la predicción de propiedades fisicoquímicas derivadas del manejo poscosecha.

La fermentación y el secado del cacao son procesos espontáneos y realizados en finca, con múltiples factores que influyentes que han sido descritos por varios autores (Kongor *et al.* 2016, Febrianto *et al.* 2022) provocando una calidad variable del producto terminado. Por su parte, la genética aporta características de composición

molecular homogéneas según el cultivar, específicamente contenido de proteína, polisacáridos y polifenoles. Por tanto, es altamente probable producir granos de cacao con un perfil de calidad pobre, según el proceso de fermentación y el secado realizado (De Vuyst y Weckx 2016).

Es una práctica común de los productores de cacao, utilizar mezclas de granos fermentados y secos de diferentes orígenes y prácticas para complementar o contrastar las inconsistencias en calidad, dadas por el proceso poscosecha. Se resalta así la utilidad del NIR primordialmente en el análisis de compuestos químicos macromoleculares, así como de parámetros relacionados con la fermentación, los cuales reducen el peso del efecto de atributos asociados al origen geográfico.

CONCLUSIONES

Este estudio demuestra la utilidad del NIR mediante un abordaje estadístico multivariado para la discriminación rápida, no destructiva y eficiente del cacao por calidad, en consideración de parámetros proximales e indicadores relevantes. Los resultados muestran que un modelo de regresión lineal combinado a pretratamiento de Clasificación No Supervisada (CNS), permite la identificación de calidad de los granos de cacao al mostrar superioridad en la predicción de características químicas proximales. Las técnicas propuestas podrían ser implementadas en el estudio y control de calidad de granos de cacao secos hacia una estimación de parámetros relevantes de fácil operación.

Se debe anotar que el establecimiento de plantaciones de limitada diversidad genética (clones internacionales y comerciales), junto a prácticas poscosecha estandarizadas, contribuye con una minimización del grado de diversidad de calidad asociada al origen geográfico. De manera que, la utilidad del NIR se ve limitada en la identificación de origen y trazabilidad por procedencia. En perspectiva de un espectro limitado al rango de 400-1700 nm, podrían recomendarse estudios con capacidad superior (>1700 nm) y el

uso de equipos portátiles complementarios tales como GC MS, NMR, FTIR, entre otros, mediante enfoques de análisis multivariado.

AGRADECIMIENTOS

A las Agencias Regionales del Ministerio de Agricultura y Ganadería por su apoyo logístico en la recolección de muestras, al igual que productores de cacao del país por la facilitación de estas. Así como al Laboratorio de Análisis y Servicios Químicos, por el apoyo brindado durante esta investigación.

LITERATURA CITADA

- ANKOM Technology. 2006. Method 10. Crude Fiber Analysis in Feeds By Filter Bag Technique (For A2000, A20001). s. p.
- Anyidoho, EK; Teye, E; Agbemafle, R. 2020. Non-destructive authentication of the regional and geographical origin of cocoa beans by using a handheld NIR spectrometer and multivariate algorithm. *Analytical Methods* 12(33):4150-4158. DOI: <https://doi.org/10.1039/d0ay00901f>
- Anyidoho, EK; Teye, E; Agbemafle, R. 2021. Differentiation of Organic Cocoa Beans and Conventional Ones by Using Handheld NIR Spectroscopy and Multivariate Classification Techniques. *International Journal of Food Science*. DOI: <https://doi.org/10.1155/2021/1844675>
- Anyidoho, EK; Teye, E; Agbemafle, R; Amuah, CLY; Boadu, VG. 2021. Application of portable near infrared spectroscopy for classifying and quantifying cocoa bean quality parameters. *Journal of Food Processing and Preservation* 45(5):1-13. DOI: <https://doi.org/10.1111/jfpp.15445>
- AOAC. 2005a. Meat and Meat Products: Official Method 981.10. *In* Official Methods of Analysis 39.1.19. p. 7.
- AOAC. 2005b. Meat and Meat Products: Official Method 950.46. *In* Official Methods of Analysis 39.1.02. p. 1.
- AOAC. 2005c. Animal Feed: Official Method 942.05. *In* Official Methods of Analysis 4.1.10. p. 8.
- AOAC. 2005d. Meat and Meat Products: Official Method 991.36. *In* Official Methods of Analysis 39.1.08. p. 3.
- Barbin, DF; Maciel, LF; Bazoni, CHV; Ribeiro, MS; Carvalho, RDS; Bispo, ES; Miranda, MPS; Hirooka, EY. 2018. Classification and compositional characterization of different varieties of cocoa beans by near infrared spectroscopy and multivariate statistical analyses. *Journal of Food Science and Technology* 55(7):2457-2466. DOI: <https://doi.org/10.1007/s13197-018-3163-5>
- Caporaso, N; Whitworth, MB; Fowler, MS; Fisk, ID. 2018. Hyperspectral imaging for non-destructive prediction of fermentation index, polyphenol content and antioxidant activity in single cocoa beans. *Food Chemistry* 258:343-351. DOI: <https://doi.org/10.1016/j.foodchem.2018.03.039>
- D'Souza, RN; Grimbs, S; Behrends, B; Bernaert, H; Ullrich, MS; Kuhnert, N. 2017. Origin-based polyphenolic fingerprinting of Theobroma cacao in unfermented and fermented beans. *Food Research International* 99:550-559. DOI: <https://doi.org/10.1016/j.foodres.2017.06.007>
- De Vuyst, L; Weckx, S. 2016. The cocoa bean fermentation process: from ecosystem analysis to starter culture development. *Journal of Applied Microbiology* 121(1): 5-17. DOI: <https://doi.org/10.1111/jam.13045>
- Fanning, E; Eyres, G; Frew, R; Kebede, B. 2023. Linking cocoa quality attributes to its origin using geographical indications. *Food Control* 151:109825. DOI: <https://doi.org/10.1016/j.foodcont.2023.109825>
- Febrianto, NA; Wang, S; Zhu, F. 2022. Chemical and biological properties of cocoa beans affected by processing: a review. *Critical Reviews in Food Science and Nutrition* 62(30):8403-8434. DOI: <https://doi.org/10.1080/10408398.2021.1928597>
- Forte, M; Currò, S; Van de Walle, D; Dewettinck, K; Mirisola, M; Fasolato, L; Carletti, P. 2023. Quality Evaluation of Fair-Trade Cocoa Beans from Different Origins Using Portable Near-Infrared Spectroscopy (NIRS). *Foods* 12(1):4. DOI: <https://doi.org/10.3390/foods12010004>
- Hayati, R; Zulfahrizal, Z; Munawar, AA. 2021. Robust prediction performance of inner quality attributes in intact cocoa beans using Near-Infrared spectroscopy and multivariate analysis. *Heliyon* 7(2):e06286. DOI: <https://doi.org/10.1016/j.heliyon.2021.e06286>
- Hernández, CE; Granados, L. 2021. Quality differentiation of cocoa beans: implications for geographical indications. *Journal of the Science of Food and Agriculture* 101(10):3993-4002. DOI: <https://doi.org/10.1002/jsfa.11077>
- Hinne, M; Semanhyia, E; Van de Walle, D; De Winne, A; Tzompa-Sosa, DA; Scalone, GLL; De Meulenaer, B; Messens, K; Van Durme, J; Afoakwa, EO; De Cooman, L; Dewettinck, K. 2018. Assessing the influence of pod storage on sugar and free amino acid profiles and the implications on some Maillard reaction related flavour volatiles in Forastero cocoa beans. *Food Research International* 111:607-620. DOI: <https://doi.org/10.1016/j.foodres.2018.05.064>
- Hue, C; Gunata, Z; Bergounhou, A; Assemat, S; Boulanger, R; Sauvage, FX; Davrieux, F. 2014. Near infrared spectroscopy as a new tool to determine cocoa fermentation levels through ammonia nitrogen quantification. *Food Chemistry* 148:240-245. DOI: <https://doi.org/10.1016/j.foodchem.2013.10.005>

- Kongor, JE; Hinnah, M; de Walle, D Van; Afoakwa, EO; Boeckx, P; Dewettinck, K. 2016. Factors influencing quality variation in cocoa (*Theobroma cacao*) bean flavour profile - A review. *Food Research International* 82:44-52. DOI: <https://doi.org/10.1016/j.foodres.2016.01.012>
- Kutsanedzie, FYH; Chen, Q; Hassan, MM; Yang, M; Sun, H; Rahman, MH. 2018. Near-Infrared system coupled chemometric algorithms for enumeration of total fungi count in cocoa beans neat solution. *Food Chemistry* 240:231-238. DOI: <https://doi.org/10.1016/j.foodchem.2017.07.117>
- Kutsanedzie, FYH; Chen, Q; Sun, H; Cheng, W. 2017. In situ cocoa beans quality grading by Near-Infrared-chemodyes systems. *Analytical Methods* 9(37):5455-5463. DOI: <https://doi.org/10.1039/c7ay01751k>
- Orozco Aguilar, L; Deheuvels, O; Villalobos, M; Somarriba Chávez, RE. 2007. El sector cacao en Centroamérica (en línea). s.l., s.e. Consultado 24 de noviembre de 2023 . Disponible en www.catie.ac.cr
- Papalexandratou, Z; Kaasik, K; Kauffmann, LV; Skorstengaard, A; Bouillon, G; Espensen, JL; Hansen, LH; Jakobsen, RR; Blennow, A; Krych, L; Castro-Mejía, JL; Nielsen, DS. 2019. Linking cocoa varieties and microbial diversity of Nicaraguan fine cocoa bean fermentations and their impact on final cocoa quality appreciation. *International Journal of Food Microbiology* 304(June):106-118. DOI: <https://doi.org/10.1016/j.ijfoodmicro.2019.05.012>
- Rojas, K. 2020. Transformaciones bioquímicas asociadas a la calidad de algunos grupos genéticos de cacao (*Theobroma cacao* L.) mediante un proceso de fermentación controlada. Tesis Lic. Heredia, Costa Rica, Universidad Nacional. 96 p.
- Sunoj, S; Igathinathane, C; Visvanathan, R. 2016. Nondestructive determination of cocoa bean quality using FT-NIR spectroscopy. *Computers and Electronics in Agriculture* 124:234-242. DOI: <https://doi.org/10.1016/J.COMPAG.2016.04.012>
- Teye, E. 2022. Mini shortwave spectroscopic techniques and multivariate statistical analysis as a tool for testing intact cocoa beans at farmgate for quality control in Ghana. *Infrared Physics and Technology* 122. DOI: <https://doi.org/10.1016/J.INFRARED.2022.104092>
- Teye, E; Anyidoho, E; Agbemafle, R; Sam-Amoah, LK; Elliott, C. 2020. Cocoa bean and cocoa bean products quality evaluation by NIR spectroscopy and chemometrics: A review. *Infrared Physics and Technology* 104. DOI: <https://doi.org/10.1016/j.infrared.2019.103127>
- Teye, E; Huang, X yi; Lei, W; Dai, H. 2014. Feasibility study on the use of Fourier transform Near-Infrared spectroscopy together with chemometrics to discriminate and quantify adulteration in cocoa beans (en línea). *Food Research International* 55:288-293. DOI: <https://doi.org/10.1016/j.foodres.2013.11.021>
- Teye, E; Huang, X; Sam-Amoah, LK; Takrama, J; Boison, D; Botchway, F; Kumi, F. 2015. Estimating cocoa bean parameters by FT-NIRS and chemometrics analysis. *Food Chemistry* 176:403-410. DOI: <https://doi.org/10.1016/J.FOODCHEM.2014.12.042>
- Verce, M; Schoonejans, J; Hernandez Aguirre, C; Molina-Bravo, R; De Vuyst, L; Weckx, S. 2021. A combined metagenomics and metatranscriptomics approach to unravel Costa Rican cocoa box fermentation processes reveals yet unreported microbial species and functionalities. *Frontiers in microbiology* 12:641185.
- Zarrillo, S; Gaikwad, N; Lanaud, C; Powis, T; Viot, C; Lesur, I; Fouet, O; Argout, X; Guichoux, E; Salin, F; Solorzano, RL; Bouchez, O; Vignes, H; Severt, P; Hurtado, J; Yepez, A; Grivetti, L; Blake, M; Valdez, F. 2018. The use and domestication of *Theobroma cacao* during the mid-Holocene in the upper Amazon. *Nat. Ecol. Evol.* 2:1879-1888.



