

Artículo original**Comparación De Métodos Particionales Y De Consenso Para La Clasificación Del Germoplasma De *Musa* spp.****Comparison Of Partitioning And Consensus Methods For The Classification Of *Musa* spp. Germplasm**<https://orcid.org/0000-0003-4199-6652>

Osmany Molina Concepción*, Lianet Gonzáles Díaz, Yaselis Guillen López y Carmen C. Pons Pérez.

Instituto de
Investigaciones de
Viandas Tropicales
(INIVIT).
Apartado 6, Santo
Domingo, CP: 53 000,
Villa Clara, Cuba.

RESUMEN

Este estudio evaluó y comparó métodos de agrupamiento particional y técnicas de consenso para la clasificación de genotipos de plátano (*Musa* spp.) del banco de germoplasma del Instituto de Investigaciones de Viandas Tropicales (Inivit). A partir de un conjunto de datos balanceado (336 accesiones y cuatro grupos genómicos principales), se integraron 42 variables cualitativas y siete cuantitativas mediante análisis multivariante. Las variables categóricas se transformaron con análisis de correspondencia múltiple y las cuantitativas se normalizaron, integrando ambos conjuntos mediante distancia Euclidiana. Se aplicaron algoritmos de K-medias (Hartigan-Wong, Lloyd, Forgy, MacQueen), PAM, CLARA y FANNY, y se utilizaron siete métodos de consenso para combinar los resultados. Los agrupamientos se evaluaron mediante coeficiente de Silueta, matrices de confusión, área bajo la curva y diversas métricas externas reconocidas, como la precisión global, el *F1-score* y el índice de Rand ajustado. Los análisis se ejecutaron en R, facilitando la integración de datos y el procesamiento estadístico. Este enfoque permitió comparar objetivamente la correspondencia entre las soluciones obtenidas y la clasificación biológica de referencia, proporcionando una valoración integral de la calidad y robustez de los métodos aplicados. Los métodos de consenso, en especial *soft/symdiff* y GV3, mostraron mayor precisión, robustez y correspondencia con la clasificación biológica de referencia, superando a los algoritmos individuales. Estos resultados más que contribuir directamente a la conservación, proporcionan un conocimiento más detallado de la variabilidad conservada y favorecen su uso más eficiente en estrategias de mejora y conservación. El procedimiento es escalable y potencialmente aplicable a otras colecciones de germoplasma, apoyando la utilización sostenible de los recursos fitogenéticos.

* Autor para
correspondencia:
taxonumeric@inivit.cu

Palabras clave: agrupamiento particional, germoplasma, métodos no supervisados, validación.

ABSTRACT

This study evaluated and compared partitioning clustering methods and consensus techniques for the classification of plantain (*Musa* spp.) genotypes from the germplasm bank of the Institute of Tropical Root and Tuber

Research (Inivit). Based on a balanced dataset (336 accessions and four main genomic groups), 42 qualitative variables and seven quantitative variables were integrated through multivariate analysis. Categorical variables were transformed using multiple correspondence analysis,

and quantitative variables were normalized, integrating both sets through Euclidean distance. K-means algorithms (Hartigan–Wong, Lloyd, Forgy, MacQueen), PAM, CLARA, and FANNY were applied, and seven consensus methods were used to combine the results. Cluster solutions were evaluated using the coefficient of Silhouette, confusion matrices, area under the curve, and several recognized external metrics, including overall accuracy, the *F1-score*, and the adjusted Rand index. All analyses were conducted in R, facilitating data integration and statistical processing. This approach enabled an objective comparison between the obtained clustering solutions and the biological reference classification, providing a comprehensive assessment of the quality and robustness of the methods. Consensus approaches, particularly soft/symdiff and GV3, showed higher accuracy, robustness, and correspondence with the biological reference classification, outperforming individual algorithms. Rather than contributing directly to conservation, these results provide a more detailed understanding of the conserved variability and support its more efficient use in improvement and conservation strategies. The procedure is scalable and potentially applicable to other germplasm collections, supporting the sustainable utilization of plant genetic resources.

Keywords: partitional clustering, germplasm, unsupervised methods, validation.

INTRODUCCIÓN

El avance tecnológico de las últimas décadas ha revolucionado la capacidad de análisis de datos, permitiendo abordar investigaciones más complejas y multidisciplinarias. La interacción entre matemáticas, estadística y biología,

aunque no es un fenómeno reciente, ha ganado impulso gracias al desarrollo de herramientas informáticas avanzadas. Estas herramientas han facilitado la implementación de métodos sofisticados para el análisis e interpretación de grandes volúmenes de datos, respondiendo a las necesidades de disciplinas como Agronomía y Genética, así como del campo de la Conservación de Recursos Fitogenéticos.

El uso de técnicas estadísticas multivariadas ha tenido un papel relevante en diversas áreas del desarrollo agrícola. Ejemplos notables incluyen la clasificación morfoagronómica de genotipos (Morejón *et al.*, 2002), la conservación de germoplasma (Borges *et al.*, 2002) y la caracterización del germoplasma cubano de malanga (*Xanthosoma sagittifolium* (L.) Schott) (Milián *et al.*, 2018). Asimismo, líneas de investigación más recientes, como la Bioinformática, se han consolidado como campos prometedores para el análisis de datos biológicos.

La aplicación de métodos numéricos permite reducir la complejidad de los datos y cuantificar la variabilidad genética en colecciones de germoplasma (Ruiz De Galarreta, 1998).

Desde 1988, el Instituto de Investigaciones de Viandas Tropicales (Inivit) ha trabajado en la documentación y estudio de recursos genéticos de acuerdo con proyectos internacionales. Este esfuerzo ha generado una base de datos valiosa que recoge información resultante de evaluaciones morfológicas, agronómicas y moleculares de diversos genotipos.

La clasificación de genotipos de plátano (*Musa* spp.) representa un desafío debido a la alta variabilidad fenotípica y a la complejidad de integrar información proveniente de diferentes tipos de datos. La combinación de variables cualitativas y

cuantitativas mediante métodos multivariados avanzados permite explorar la estructura interna de las colecciones y evaluar la coherencia topológica de los agrupamientos obtenidos. Este tipo de análisis constituye una herramienta útil para apoyar la gestión y utilización de recursos fitogenéticos, y proporciona bases sólidas para el desarrollo de estudios más integrales y sistemáticos.

A pesar de los avances en la aplicación de técnicas de agrupamiento a colecciones genéticas de *Musa* spp., persisten desafíos metodológicos relevantes. La literatura muestra que la mayoría de los estudios utilizan algoritmos clásicos como k-medias, PAM o enfoques jerárquicos, pero enfrentan limitaciones debidas al desbalance de clases, la complejidad estructural de los datos y la integración de variables mixtas, lo que afecta la reproducibilidad y comparabilidad de los resultados (Jain, 2010; Vega-Pons y Ruiz-Shulcloper, 2011). Además, es escasa la evidencia sobre la efectividad de métodos de consenso en este contexto, y se discute poco el impacto de preprocesamientos como el balanceo por simulación o la combinación de variables cualitativas y cuantitativas sobre la calidad de las particiones (Strehl y Ghosh, 2002; Hornik *et al.*, 2025).

En este escenario, el presente estudio tiene como objetivo principal evaluar y comparar métodos de agrupamiento particional y técnicas de consenso para la clasificación de accesiones de *Musa* spp. A estos fines, se integran variables cualitativas y cuantitativas mediante procedimientos multivariados, y se emplea un conjunto de datos balanceado para garantizar la robustez de las evaluaciones. Asimismo, se utilizan métricas internas y externas con el propósito de valorar la coherencia topológica y biológica de los

agrupamientos generados. El alcance previsto de esta investigación incluye aportar herramientas replicables y evidencia metodológica que fortalezcan la gestión y utilización de recursos fitogenéticos en bancos de germoplasma.

MATERIALES Y MÉTODOS

Descripción del conjunto de datos

Los análisis ejecutados en esta investigación se realizaron con una colección de trabajo compuesta por 131 accesiones de plátano (*Musa* spp.), caracterizadas mediante 42 variables cualitativas (nominales y ordinales) y 7 cuantitativas, de acuerdo con los descriptores mínimos establecidos por IPGRI-INIBAP/CIRAD (1996) del banco de germoplasma, que se conserva en el Instituto de Investigaciones de Viandas Tropicales (Inivit).

La investigación se realizó con una colección de trabajo de 131 accesiones procedentes del Banco de Germoplasma de *Musa* spp., conservado en el Instituto de Investigaciones de Viandas Tropicales (Inivit). Los análisis se ejecutaron utilizando la información generada por 42 variables cualitativas (nominales y ordinales) y siete variables cuantitativas, de acuerdo con los descriptores mínimos establecidos por IPGRI-INIBAP/CIRAD (1996).

Los métodos de agrupamiento y clasificación son sensibles a la desproporción entre clases, lo que puede distorsionar la formación de los conglomerados y favorecer la correcta identificación de grupos sobrerrepresentados en detrimento de las clases minoritarias. Para mitigar este efecto, se generó una versión balanceada del conjunto de datos mediante sobre muestreo sintético, siguiendo enfoques empleados en la literatura para tratar problemas de desbalance en el análisis multivariado (Chawla *et al.*, 2002;

Lunardon *et al.*, 2014). En este estudio, el sobre muestreo se aplicó exclusivamente a los grupos minoritarios, con el fin de reducir el sesgo inducido por la desigual representación de accesiones entre los distintos grupos genómicos.

En concreto, para cada grupo minoritario se preservaron los perfiles categóricos replicando las combinaciones de variables cualitativas observadas en accesiones originales seleccionadas al azar. Para las variables cuantitativas, se generaron valores adicionales mediante la función `rnorm()` de R, ajustando las simulaciones a la media y desviación estándar de cada clase. La validez de las muestras simuladas se verificó mediante pruebas de homogeneidad de varianzas (Levene) y comparación de distribuciones (Kolmogorov–Smirnov), confirmando que la dispersión y forma de las distribuciones generadas no difirieron significativamente de las observadas en los datos originales. Este procedimiento garantizó que el conjunto balanceado mantuviera un comportamiento estadístico y biológico consistente.

Este procedimiento de balanceo por simulación permitió mantener la coherencia y el realismo biológico del conjunto de datos, garantizando que la variabilidad intra-grupo de cada clase estuviera representada de forma adecuada. Así, se establecieron condiciones más equitativas para la aplicación y comparación de los métodos de agrupamiento, reforzando la validez y robustez de las conclusiones obtenidas.

Preprocesamiento y codificación

En un primer paso, las variables categóricas se transformaron mediante Análisis de Correspondencia Múltiple (ACM) utilizando la función `MCA()` del paquete *FactoMineR* (Tenenhaus, 1985), lo que permitió integrarlas con las variables cuantitativas. Estas variables cuantitativas se normalizaron

previamente mediante la función `data.Normalization(n1)` del paquete *clusterSim* (Walesiak, 2020).

La selección de dimensiones a retener en el ACM se basó en criterios combinados, considerando la regla de Kaiser–Guttman (*eigenvalues* > 1) (Guttman, 1953; Kaiser, 1960), el porcentaje de varianza explicada acumulada superior al 90 % (Greenacre, 1984; Husson, 2017) y la corrección de Benzécri–Greenacre (Benzécri, 1979; Greenacre, 2017).

Para cuantificar las diferencias y similitudes entre accesiones, se utilizó la distancia euclidiana implementada en la función `dist()` del paquete *stats*.

Si bien el estudio se centra en la clasificación no supervisada, la identificación inicial de los grupos genómicos se basó en información experta proporcionada por los curadores del banco de germoplasma.

Aplicación de métodos particionales

Se aplicaron métodos de aglomeración particional, incluyendo K-medias mediante la función `kmeans()` del paquete *stats*, empleando sus cuatro algoritmos clásicos: *Forgy* (Forgy, 1965), *Hartigan-Wong* (Hartigan y Wong, 1979), *Lloyd* (Lloyd, 1982) y *MacQueen* (MacQueen, 1967). En todos los casos, se fijó el número de clústeres en cuatro, correspondiente al número de grupos genómicos de referencia, y la inicialización de centroides se realizó de manera aleatoria.

El parámetro *iter.max* se fijó en 500 para asegurar que el algoritmo K-medias alcanzara la convergencia, y evitar la terminación prematura en particiones complejas (Jain, 2010). El parámetro *nstart* se configuró en 20 para asegurar la robustez de la solución, optimizando así la estabilidad y calidad de la agrupación (Hastie *et al.*, 2009).

Asimismo, se emplearon los métodos PAM, CLARA y FANNY implementados

en el paquete *cluster* (Maechler *et al.*, 2023), utilizando las funciones *pam()*, *clara()* y *fanny()* respectivamente, con cuatro clústeres y configuraciones por defecto recomendadas en la documentación. En el algoritmo FANNY se estableció el parámetro de exponenciación de membresía (*memb.exp*) en 1.3, tal como se recomienda en la literatura para lograr una adecuada discriminación entre clústeres, y evitar agrupaciones excesivamente blandas o rígidas (Kaufman y Rousseeuw, 1990; Maechler *et al.*, 2023).

La elección del número de conglomerados ($k=4$) se fundamentó en criterio biológico, la colección cubana de plátano y banano está estructurada en cuatro grupos genómicos de referencia (AAAB, AAB, AABB y ABB), por lo que resulta natural buscar una partición que refleje esta clasificación.

Análisis de consenso

El procedimiento de consenso entre los diferentes métodos particionales se implementó utilizando la función *cl_ensemble()* del paquete *clue* (Hornik *et al.*, 2023), combinando los resultados de los algoritmos individuales mediante siete enfoques de consenso disponibles en la función *cl_consenso()*: GV1, SM, SE, GV3, *soft/symdiff*, DWH y HE (Monti *et al.*, 2003; Vega-Pons y Ruiz-Shulcloper, 2011). La estabilidad y sensibilidad de las particiones consenso se analizaron evaluando la concordancia entre los diferentes consensos generados y su desempeño frente a la clasificación biológica de referencia. Para garantizar la robustez de los resultados, se aplicó validación cruzada interna a las soluciones consenso, siguiendo recomendaciones metodológicas de la literatura especializada.

Evaluación de los resultados de agrupamiento

La evaluación de la calidad de las agrupaciones, tanto para los métodos individuales como para las particiones consenso, se realizó aplicando un conjunto de métricas objetivas y complementarias. En todos los casos, la coherencia interna de los clústeres se cuantificó con el coeficiente de Silueta (Rousseeuw, 1987; Kaufman y Rousseeuw, 1990), calculado mediante la función *silhouette()* del paquete *cluster*, que permite interpretar la adecuación de la asignación de los elementos a los conglomerados formados, proporcionando una interpretación intuitiva de la calidad de la partición. De acuerdo con los criterios de interpretación propuestos por Rousseeuw (1987), valores del coeficiente de Silueta cercanos a 0.3 se corresponden con una estructura de clústeres débil a moderada, mientras que valores próximos a 1 indican agrupamientos bien definidos.

Además, se incluyeron dos métricas complementarias, ampliamente empleadas en la literatura de agrupamiento (Manning *et al.*, 2008; Powers, 2011):

- ✓ Pureza (*purity*). Para cada clúster C_j se determina la clase G_i más frecuente y se calcula:

$$purity(C_j) = \left(\frac{1}{N}\right) * \sum_{i=1}^k (max_i |C_j \cap G_i|)$$

La pureza global es la suma de estas fracciones ponderada por el tamaño de cada clúster. Una pureza elevada indica asignaciones mayoritariamente correctas con respecto a la clase dominante.

La pureza proporciona una valoración intuitiva de la coherencia interna de los agrupamientos y se ha consolidado como una métrica estándar para la evaluación externa en estudios de agrupamiento. El cálculo de la pureza implica identificar, para cada clúster, la clase con mayor

representación y sumar la cantidad de accesiones correctamente agrupadas, dividiendo posteriormente entre el total de observaciones.

- ✓ Entropía ponderada. Mide la diversidad de clases dentro de cada clúster:

$$H_w = \left(\frac{1}{N}\right) * \sum_{j=1}^k \left(|C_j| * \left(- \sum_i (p_{ij} * \log_2 p_{ij}) \right) \right)$$

donde P_{ij} es la proporción de accesiones de la clase G_i en C_j . La entropía global resulta de ponderar por el tamaño relativo de C_j (Shannon, 1948). Valores bajos reflejan homogeneidad y, por ende, mejor separación inter-clúster.

O sea, la entropía ponderada mide la diversidad de clases en cada clúster, proporcionando información sobre el grado de homogeneidad interna y la presencia de solapamientos entre grupos. La entropía de cada clúster se calcula considerando la distribución de las clases presentes, y la entropía total se obtiene como la suma ponderada según el tamaño relativo de cada clúster (Strehl y Ghosh, 2002; Meilä, 2003). Valores bajos de entropía indican una mayor pureza y menor dispersión de clases dentro de los conglomerados, mientras que valores elevados sugieren mayor heterogeneidad interna.

Ambos indicadores se calcularon a partir de las matrices de contingencia grupo genómico \times clúster generadas para cada algoritmo mediante rutina desarrolladas en R (ver Anexo, Script 1).

Estas métricas complementan la evaluación tradicional mediante coeficiente de Silueta y métricas externas, proporcionando una visión más integral y objetiva sobre la calidad de las particiones obtenidas.

Para la validación externa y la comparación directa con la clasificación de referencia (grupos genómicos), se emplearon matrices de confusión (Kohavi y Provost, 1998; Provost y Fawcett,

2001). Las etiquetas predichas y reales se alinearon de forma óptima mediante el Algoritmo Húngaro (Kuhn, 1955) o, en su caso, utilizando el mapeo por mayoría.

El desempeño de las particiones obtenidas se evaluó mediante múltiples métricas de validación externa. En este estudio se reportan tanto la precisión global (*accuracy*) como la precisión (*precision*), ya que representan aspectos complementarios en la evaluación de los modelos de clasificación. La precisión global corresponde al porcentaje total de accesiones correctamente clasificadas en cualquiera de los grupos genómicos y proporciona una visión general del desempeño del modelo. Siguiendo criterios habituales en la evaluación de clasificadores (Hastie *et al.*, 2009; Kuhn y Johnson, 2013), se consideraron valores de precisión global y *F1-score* superiores a 0.85 como altos, entre 0.70 y 0.85 como moderados y por debajo de 0.70 como bajos.

Por su parte, la precisión se refiere al porcentaje de accesiones clasificadas dentro de una clase que realmente pertenecen a ella. Esta métrica permite evaluar la calidad de las asignaciones positivas y resulta especialmente útil en escenarios con desbalance de clases o cuando los falsos positivos tienen un costo elevado. Esta distinción metodológica está ampliamente documentada en la literatura científica (Powers, 2011; Saito y Rehmsmeier, 2015).

La precisión global y el coeficiente de acuerdo de Cohen (*Kappa*) se calcularon con el paquete *caret* (Kuhn, 2008), brindando una estimación inicial del nivel de concordancia entre la clasificación automática y los grupos de referencia. Adicionalmente, se emplearon métricas de promedio macro y ponderado de precisión, *recall* y *F1-score* utilizando el paquete *yardstick* (Kuhn *et al.*, 2024), lo

que permitió valorar de manera equilibrada el rendimiento en presencia de clases potencialmente desbalanceadas. Para evaluar la calidad de la asignación de los elementos a los grupos, se incluyeron el índice de Rand ajustado (ARI) (Hubert y Arabie, 1985) y el índice de Fowlkes-Mallows (Fowlkes y Mallows, 1983), ambos reconocidos en la literatura como indicadores robustos para medir el desempeño en métodos de agrupamiento.

Finalmente, se examinó la capacidad discriminante de cada partición utilizando el área bajo la curva ROC (AUC) multicategoría mediante la función `multiclass.roc()` del paquete *pROC* (Swets *et al.*, 2000a, Fawcett, 2003; Robin *et al.*, 2011).

Mejora relativa y reducción del error

Además de las métricas de validación externa, se cuantificó la mejora relativa de los métodos de consenso con respecto a un algoritmo particional tradicional de referencia para cada métrica M mediante la expresión:

$$\text{Mejora relativa (\%)} = \frac{M_{\text{método}} - M_{\text{ref}}}{M_{\text{ref}}} \times 100,$$

donde $M_{\text{método}}$ es el valor de la métrica para el método evaluado y M_{ref} el valor correspondiente al algoritmo de referencia. De forma análoga, se calculó la reducción relativa del error de clasificación utilizando:

$$\text{Reducción de errores (\%)} = \frac{\text{Error}_{\text{ref}} - \text{Error}_{\text{método}}}{\text{Error}_{\text{ref}}} \times 100,$$

Donde

$\text{Error} = 1 - \text{precisión global (accuracy)}$. Este conjunto de métricas se aplicó sistemáticamente tanto en el análisis de los métodos particionales individuales como en la evaluación de las particiones consenso, permitiendo una valoración exhaustiva de la coherencia interna, la

concordancia con la clasificación de referencia y la capacidad discriminante de las soluciones obtenidas.

Visualización y presentación de resultados

El entorno de trabajo empleado para todos los análisis fue el lenguaje R, versión 4.3.3 (*R Core Team*, 2024), ampliamente reconocido por su flexibilidad y potencia para el análisis estadístico, procesamiento de datos y visualización científica.

Para la generación de gráficos y visualizaciones de los resultados se utilizaron los paquetes *ggplot2* (para gráficos personalizados), *pheatmap* (mapas de calor) y funciones de resumen y manipulación de datos provenientes de los paquetes *dplyr* y *tidyr* (Wickham, 2023). Las tablas resumen de resultados y las métricas de evaluación se construyeron utilizando estas herramientas, garantizando una presentación clara y reproducible de los datos.

RESULTADOS Y DISCUSIÓN

Caracterización y preprocesamiento de los datos

Esta investigación se propuso resolver los desafíos asociados a la clasificación no supervisada en la colección cubana de plátanos y bananos. La colección original incluía 10 accesiones del grupo genómico AAAB, 84 del grupo AAB, dos del grupo AABB y 35 del grupo ABB. Para garantizar la comparabilidad y la robustez estadística en los análisis, se construyó una versión balanceada del conjunto de datos, compuesta por 336 accesiones distribuidas equitativamente entre los cuatro grupos genómicos de referencia (AAAB, AAB, AABB y ABB). Esta estrategia permitió igualar la representación de cada grupo y optimizar

las condiciones para la aplicación de los métodos de agrupamiento.

La determinación del número de dimensiones a retener se basó en el criterio de Kaiser–Guttman, seleccionándose únicamente aquellas cuyos valores propios superaron la media global. Este procedimiento condujo a la retención de 36 dimensiones, que explican el 91.13 % de la variabilidad total. Cabe señalar que, conforme a lo sugerido por Sorensen (1948) y por Gatna y Walesiak (2004), un umbral cercano al 70 % de inercia acumulada suele considerarse suficiente para aplicaciones prácticas. El mayor número de dimensiones retenidas en este estudio responde a la elevada cantidad de variables cualitativas incluidas en el conjunto de datos, lo cual incrementa de forma natural el número de ejes necesarios para alcanzar altos porcentajes de variabilidad explicada. Paralelamente, las siete variables cuantitativas ya normalizadas se integraron con las dimensiones obtenidas del ACM para conformar una matriz unificada de datos continuos. A partir de esta matriz se calcularon las distancias euclidianas mediante las funciones correspondientes del paquete estadístico *stats* de R.

Evaluación de métodos particionales

La aplicación de los métodos particionales al conjunto de datos balanceado permitió obtener cuatro conglomerados principales, en correspondencia con los grupos genómicos de referencia presentes en la colección analizada. Estas soluciones reflejan una estructura interna coherente con la variabilidad observada en las

accesiones y constituyen la base para la posterior comparación con los métodos de consenso.

La Tabla 1 muestra la correspondencia entre los grupos genómicos de referencia y los conglomerados generados por los distintos algoritmos particionales, revelando diferencias claras en la calidad de la clasificación no supervisada. En general, los métodos Lloyd y Hartigan–Wong lograron una asignación más precisa, con accesiones mayoritariamente concentradas en clústeres específicos, lo que indica una estructura genómica bien capturada y un mejor alineamiento con la clasificación biológica de referencia. Estas diferencias reflejan la sensibilidad variable de los algoritmos ante la estructura interna del conjunto de datos. Esta evaluación inicial constituye un punto de partida fundamental para los análisis cuantitativos posteriores mediante métricas de validación externa y visualizaciones complementarias.

La pureza y la entropía ponderada calculadas para cada método particional (Tabla 2) ofrecen una perspectiva complementaria sobre la calidad de los agrupamientos generados. Entre los métodos evaluados, Lloyd destaca por alcanzar la mayor pureza y la menor entropía, lo que indica una asignación más homogénea y coherente de accesiones dentro de los clústeres. Este desempeño superior puede atribuirse a la eficacia del algoritmo en la actualización iterativa de los centroides y, con ello, minimizar la variabilidad interna y maximizar la separación entre grupos.

Tabla 1. Matriz de contingencia entre grupos genómicos de referencia (filas) y conglomerados (columnas) formados por los algoritmos de aglomeración particionales en estudio.

Métodos	Grupo Genómico	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Hartigan-Wong	AAAB	76	0	8	0
	AAB	11	2	0	71
	ABB	0	72	9	3
	AABB	0	0	42	42
Lloyd	AAB	1	0	73	10
	ABB	72	12	0	0
	AAAB	0	8	9	67
	AABB	0	84	0	0
Forgy	AAB	1	4	69	10
	ABB	71	0	0	13
	AAAB	0	9	0	75
	AABB	0	0	0	84
MacQueen	AAAB	8	59	17	0
	AABB	42	0	42	0
	ABB	9	0	75	0
	AAB	0	9	4	71
PAM	AABB	42	42	0	0
	ABB	39	45	0	0
	AAAB	0	17	7	60
	AAB	0	1	75	8
CLARA	AAAB	8	17	52	7
	AAB	36	1	3	44
	ABB	69	15	0	0
	AABB	0	42	0	42
FANNY	AAB	73	0	11	0
	ABB	38	26	20	0
	AAAB	0	8	76	0
	AABB	0	42	0	42

Tabla 2. Pureza y entropía ponderada de los agrupamientos obtenidos por los diferentes métodos particionales aplicados al conjunto de accesiones de *Musa spp.*

Métodos Particionales	Pureza	Entropía
Hartigan-Wong	0.777	0.764
Lloyd	0.881	0.549
Forgy	0.693	0.893
MacQueen	0.735	0.941
Pam	0.661	0.931
Clara	0.616	1.161
Fanny	0.693	0.973

Estos resultados subrayan la importancia de emplear métricas internas adicionales más allá de la exactitud externa para seleccionar el algoritmo particional más adecuado.

El objetivo central de las técnicas de agrupamiento es identificar grupos compactos y bien definidos dentro de los datos; sin embargo, resulta igualmente imprescindible evaluar la eficiencia de los métodos empleados. Para ello, la calidad de los agrupamientos obtenidos se valoró mediante el coeficiente de Silueta (Tabla 3).

Tabla 3. Coeficiente de Silueta para los diferentes algoritmos particionales.

Método	Coeficiente de Silueta
Hartigan-Wong	0.26
Lloyd	0.26
Forgy	0.22
MacQueen	0.28
PAM	0.26
CLARA	0.23
FANNY	0.22

Los valores observados para los métodos analizados se sitúan en un rango intermedio (0.22–0.28), lo cual, de acuerdo con los criterios de interpretación propuestos por Rousseeuw (1987), corresponde a una estructura de clústeres moderadamente definida. Estos valores se encuentran alejados de los umbrales típicos asociados a agrupamientos óptimos, generalmente superiores a 0.5. Si bien los valores de Silueta registrados no alcanzan niveles óptimos pues, según los criterios establecidos por Rousseeuw (1987), valores cercanos a 1 indican una estructura de clústeres bien definidos, resultan compatibles con lo esperable en conjuntos de datos biológicos con alta variabilidad fenotípica y estructuras parcialmente solapadas, especialmente cuando se integran variables mixtas (Kaufman y Rousseeuw, 1990; Everitt *et al.*, 2011). Así, estos resultados enfatizan la importancia de no depender de una única métrica para evaluar la calidad de los agrupamientos y subrayan la conveniencia de complementar el análisis con métricas externas y validación cruzada para obtener una valoración integral del desempeño de los algoritmos utilizados.

La Tabla 4 sintetiza los valores obtenidos para las principales métricas de validación externa empleadas para comparar el desempeño de los algoritmos de agrupamiento particional.

Los resultados evidencian que el algoritmo Lloyd alcanza el mejor desempeño global, con los valores más altos en todas las métricas evaluadas. Le siguen Hartigan–Wong y Forgy que muestran rendimientos intermedios, mientras que CLARA, PAM y FANNY presentan las puntuaciones más bajas, lo que indica una menor capacidad para reproducir la clasificación biológica de referencia. Bajo las condiciones de este estudio, estos resultados respaldan la

superioridad de Lloyd para la clasificación no supervisada de accesiones de *Musa* spp. en términos de correspondencia con

la clasificación de referencia y coherencia en la asignación de las accesiones.

Tabla 4. Comparación de métricas de validación externa para los diferentes métodos de agrupamiento particional aplicados al conjunto balanceado de accesiones de *Musa* spp.

Método	Precisión Global	Kappa	Precisión (por clase)	Recall	F1	Índice de Rand	Fowlkes-Mallows
Hartigan-Wong	0.776	0.702	0.793	0.777	0.774	0.566	0.677
Lloyd	0.88	0.841	0.889	0.881	0.881	0.712	0.784
Forgy	0.693	0.591	0.785	0.693	0.657	0.479	0.65
MacQueen	0.735	0.646	0.781	0.735	0.739	0.453	0.599
PAM	0.66	0.547	0.686	0.661	0.67	0.438	0.579
CLARA	0.616	0.488	0.647	0.616	0.619	0.333	0.504
FANNY	0.645	0.527	0.678	0.646	0.634	0.428	0.578

En conjunto, los resultados ponen de manifiesto que estas métricas deben interpretarse siempre a la luz de los objetivos biológicos y del contexto de aplicación, integrándolas con otros indicadores internos y externos para obtener una evaluación más completa del desempeño de los métodos.

Los resultados a partir de los distintos métodos de agrupamiento (Tabla 5) ponen de manifiesto notables diferencias en la capacidad discriminante de cada algoritmo según el grupo genómico considerado.

El algoritmo FANNY obtuvo el AUC más alto para el grupo AAAB y AAB. Mientras Lloyd fue el mejor clasificador de AABB y Forgy, Hartigan-Wong y Lloyd alcanzaron altos AUC para ABB. En contraste, CLARA y FANNY presentaron valores bajos en AABB, indicando dificultades para discriminar este grupo. En general, Lloyd y FANNY fueron los algoritmos más consistentes en desempeño por grupo, aunque con diferencias según la

estructura genética particular de cada uno.

Tabla 5. Área bajo la curva (AUC) para los diferentes algoritmos de agrupamiento particional aplicados a los grupos genómicos de *Musa* spp.

Método	AUC-AAAB	AUC-AAB	AUC-AABB	AUC-ABB
Hartigan-Wong	0.931	0.833	0.716	0.925
Lloyd	0.879	0.917	0.96	0.927
Forgy	0.546	0.911	0.806	0.921
MacQueen	0.833	0.923	0.716	0.821
PAM	0.841	0.933	0.673	0.649
CLARA	0.804	0.665	0.685	0.823
FANNY	0.986	0.982	0.501	0.808

La variabilidad observada en el rendimiento entre métodos y grupos subraya la importancia de comparar distintas estrategias de agrupamiento para optimizar la clasificación y el estudio

de la diversidad genética en colecciones complejas.

A pesar de los avances metodológicos en la clasificación no supervisada de la colección evaluada, los resultados globales evidencian la existencia de importantes diferencias en la capacidad de los algoritmos particionales para reproducir la estructura biológica subyacente.

El análisis conjunto de las métricas de validación interna y externa (Tablas 4 y 5) muestra que el método Lloyd presenta el desempeño más estable entre los algoritmos particionales evaluados. En particular, alcanza los valores más altos de AUC en los grupos AABB (0.960) y ABB (0.927), y se sitúa entre los de mejor rendimiento en AAAB (0.879) y AAB (0.917) (Tabla 5), además de obtener las mayores precisiones globales y valores de *F1-score* en la evaluación global (Tabla 4).

En contraste, otros algoritmos muestran una menor consistencia según el grupo genómico. Por ejemplo, FANNY alcanza AUC muy elevados en AAAB (0.986) y AAB (0.982), pero desciende hasta 0.501 en AABB; de forma similar, PAM y CLARA presentan valores bajos de AUC en AABB (0.673 y 0.685, respectivamente) y PAM también en ABB (0.649) (Tabla 5). Este patrón, en el que un mismo algoritmo ofrece rendimientos contrastantes según la estructura genética del grupo y la heterogeneidad de los datos, coincide con lo descrito en estudios previos sobre análisis de clústeres en datos biológicos complejos (Kaufman y Rousseeuw, 1990; Everitt *et al.*, 2011), donde se enfatiza la necesidad de comparar múltiples métodos y contextualizar su desempeño en función de la variabilidad del conjunto de datos.

Esta variabilidad, observable tanto a nivel global como en la discriminación de grupos particulares, resalta no solo la

complejidad intrínseca del conjunto de datos caracterizado por una alta heterogeneidad y posibles solapamientos entre clases, sino también las limitaciones inherentes a los enfoques de agrupamiento clásicos cuando se aplican de manera aislada. La falta de robustez y reproducibilidad detectada, sumada a la sensibilidad de cada algoritmo a la distribución de las observaciones y la influencia de las variables, puede conducir a interpretaciones erróneas o parciales si se basa únicamente en el resultado de un método individual.

En este contexto, la adopción de estrategias de consenso representa una alternativa metodológica respaldada por la literatura científica, que subraya la importancia de combinar los resultados de múltiples métodos para mejorar la robustez y precisión del agrupamiento. Strehl y Ghosh (2002) argumentan que la combinación de distintos algoritmos puede mitigar las debilidades inherentes a cada uno y ofrecer soluciones más completas, mientras que Hastie *et al.* (2009) señalan que el uso de múltiples particiones permite capturar diversos aspectos de la estructura de los datos, lo que conduce a agrupaciones más informativas y significativas. Así, la integración de soluciones de diferentes algoritmos no solo aprovecha las fortalezas particulares de cada enfoque, sino que también compensa sus limitaciones, generando particiones más estables, reproducibles y representativas de la diversidad genética de la colección. Diversos estudios han confirmado la eficacia de los métodos de consenso en contextos de alta complejidad y variabilidad, como el abordado en este trabajo. Por tanto, la aplicación de técnicas de consenso responde tanto a una exigencia de robustez estadística como a la necesidad de garantizar la objetividad y reproducibilidad de las

conclusiones derivadas del estudio de la diversidad genética y la clasificación en bancos de germoplasma.

En vista de estos resultados, y con el objetivo de superar las limitaciones detectadas, la siguiente sección examina en profundidad la aplicación de estrategias consenso, evaluando su capacidad para mejorar la robustez y precisión de la clasificación no supervisada en la colección de *Musa* spp.

Evaluación del agrupamiento consenso

La tabla 6 muestra que los métodos de consenso GV3 y *soft/symdiff* lograron la mejor correspondencia con la clasificación biológica de referencia, asignando

correctamente casi la totalidad de las accesiones de AABB y ABB a clústeres únicos. Mientras, GV1, SM, DWH y HE también ofrecieron agrupamientos consistentes, especialmente para AAAB y AAB, aunque con mayor dispersión en AABB y ABB. En contraste, SE mostró la mayor dispersión entre clústeres, con una clasificación menos precisa de todos los grupos. Los algoritmos GV3 y *soft/symdiff* sobresalen por su capacidad para generar particiones más definidas y alineadas con la estructura genómica esperada.

Tabla 6. Matriz de contingencia entre los grupos genómicos de referencia (filas) y los conglomerados (columnas) generados por los diferentes algoritmos de consenso evaluados.

Método	Grupo Genómico	Cluster 1	Cluster 2	Cluster 3	Cluster 4
GV1	AAAB	67	17	0	0
	AAB	10	0	2	72
	AABB	0	42	42	0
	ABB	0	10	74	0
SM	AAB	72	10	2	0
	AAAB	0	67	0	17
	AABB	0	0	42	42
	ABB	0	0	74	10
SE	AAAB	7	60	0	17
	AAB	37	9	38	0
	ABB	73	0	1	10
	AABB	0	0	42	42
GV3	AAB	2	72	10	0
	ABB	74	0	0	10
	AAAB	0	0	67	17
	AABB	0	0	0	84
<i>soft/symdiff</i>	AAAB	67	0	17	0
	AAB	10	72	0	2
	AABB	0	0	84	0
	ABB	0	0	9	75

DWH	AAAB	67	0	17	0
	AAB	10	2	0	72
	AABB	0	42	42	0
	ABB	0	74	10	0
HE	AAAB	67	0	0	17
	AAB	10	72	2	0
	AABB	0	0	42	42
	ABB	0	0	74	10

En síntesis, la matriz de contingencia evidencia que los métodos consensos GV3 y *soft/symdiff*, son capaces de reproducir en buena medida la estructura genética esperada, logrando una mejor correspondencia que algunos métodos individuales y disminuyendo la dispersión de accesiones entre clústeres, cómo se observa en la tabla 7 alcanzan las purezas más altas y las entropías más bajas, lo que confirma su capacidad para generar particiones consistentes y bien definidas. Sin embargo, la calidad de la correspondencia varía según el algoritmo de consenso y el grupo genómico, lo que subraya la importancia de una evaluación sistemática y el uso de métricas adicionales para determinar la solución óptima en estudios de diversidad genética.

Tabla 7. Pureza y entropía ponderada de los agrupamientos obtenidos por los diferentes algoritmos de consenso.

Métodos Consensos	Pureza	Entropía
GV1	0.759	0.772
SM	0.759	0.772
SE	0.646	1.066
GV3	0.884	0.508
<i>soft/symdiff</i>	0.887	0.498
DWH	0.759	0.772
HE	0.759	0.772

Desde la perspectiva biológica, estos hallazgos son altamente relevantes: una

mayor pureza indica que los clústeres agrupan accesiones con alta similitud genética y fenotípica, facilitando estrategias más precisas de conservación, caracterización y utilización del material vegetal. Al mismo tiempo, una baja entropía asegura una clara separación entre grupos genómicos y minimiza la mezcla de accesiones de distintas clases, reduciendo el riesgo de errores de clasificación y mejorando la eficiencia en la gestión y selección de materiales para programas de mejora genética o conservación.

En suma, la implementación de métodos consenso que maximizan la pureza y minimizan la entropía contribuye a una gestión más eficaz y estratégica de los bancos de germoplasma, consolidando su papel en la sostenibilidad y la innovación agrícola.

La Tabla 8 indica que los métodos GV1, SM, DWH y HE alcanzaron los coeficientes de Silueta más altos (0.30), ligeramente superiores a los obtenidos por GV3 y *soft/symdiff* (0.26) y por encima del valor registrado para SE (0.24), que fue el más bajo. De acuerdo con los criterios de interpretación propuestos por Rousseeuw (1987), estos valores se sitúan en un rango compatible con una estructura de clústeres débil a moderadamente definida, lejos de los valores próximos a 1 que caracterizan agrupamientos muy bien separados. En este contexto, los métodos de consenso

que muestran mejor desempeño en términos de precisión externa no necesariamente presentan la mayor cohesión interna, aunque mantienen una partición razonablemente consistente de la colección.

Tabla 8. Coeficiente de Silueta de los algoritmos de consenso evaluados en la clasificación no supervisada de accesiones de *Musa spp.*

Método Consenso	Coeficiente de Silueta
GV1	0.30
SM	0.30
SE	0.24
GV3	0.26
<i>soft/symdiff</i>	0.26
DWH	0.30
HE	0.30

La Tabla 9 muestra que los métodos *soft/symdiff* y GV3 alcanzan los mejores resultados globales entre los algoritmos

Tabla 9. Métricas de validación externa para los diferentes algoritmos de consenso aplicados al agrupamiento de accesiones de *Musa spp.*

Método	Precisión Global	Kappa	Precisión (por clase)	Recall	F1	Índice de Rand	Fowlkes-Mallows
GV1	0.758	0.678	0.776	0.759	0.759	0.539	0.657
SM	0.758	0.678	0.776	0.759	0.759	0.539	0.657
SE	0.633	0.511	0.643	0.634	0.63	0.381	0.539
GV3	0.883	0.845	0.9	0.884	0.885	0.716	0.788
<i>soft/symdiff</i>	0.886	0.849	0.902	0.887	0.888	0.723	0.793
DWH	0.758	0.678	0.776	0.759	0.759	0.539	0.657
HE	0.758	0.678	0.776	0.759	0.759	0.539	0.657

En conjunto, *soft/symdiff* se posiciona como el método de consenso más robusto y preciso para la clasificación de accesiones de *Musa spp.* Su enfoque, basado en la consideración simultánea de coincidencias y discrepancias entre

de consenso, con precisiones globales de 0.886 y 0.883, valores de *F1-score* de 0.888 y 0.885, e índices de Rand de 0.793 y 0.716, respectivamente. En el extremo opuesto, SE presenta los valores más bajos en todas las métricas (precisión global = 0.633, *F1-score* = 0.630, índice de Rand = 0.381), lo que confirma su menor capacidad clasificatoria. Los métodos GV1, SM, DWH y HE conforman un grupo intermedio: muestran resultados prácticamente idénticos entre sí (precisión global = 0.758, *F1-score* = 0.759, índice de Rand = 0.539), claramente superiores a los de SE, pero inferiores a los registrados por *soft/symdiff* y GV3. Este patrón indica que, dentro de las condiciones evaluadas, *soft/symdiff* y GV3 constituyen las alternativas de consenso más eficaces, mientras que GV1, SM, DWH y HE ofrecen un desempeño aceptable y SE se comporta de manera claramente menos favorable.

particiones individuales, le otorga una mayor resiliencia frente a la variabilidad del conjunto de datos y al solapamiento entre clases, especialmente en contextos con grupos parcialmente superpuestos o accesiones atípicas.

Asimismo, el proceso de balanceo y la integración de variables categóricas y cuantitativas pueden amplificar la heterogeneidad de los resultados entre métodos consenso, generando estructuras internas que favorecen a unos algoritmos más que a otros. Es relevante señalar que, a pesar de estas diferencias, la mayoría de los métodos consenso igualan o superan el rendimiento de los algoritmos individuales menos robustos, lo que evidencia el valor añadido del enfoque de consenso para incrementar la fiabilidad de la clasificación en contextos de diversidad genética.

Por último, estas variaciones resaltan la importancia de seleccionar el método de consenso más adecuado según el tipo de datos y el objetivo biológico del estudio, así como la necesidad de complementar la interpretación cuantitativa con un análisis biológico cualitativo de los grupos resultantes.

La Tabla 10 resume los valores de área bajo la curva ROC (*AUC*) obtenidos por los distintos algoritmos de consenso para cada grupo genómico de *Musa* spp.

Tabla 10. Área bajo la curva ROC (*AUC*) para los algoritmos de consenso según grupo genómico en *Musa* spp.

Algoritmo	AUC-AAAB	AUC-AAB	AUC-AABB	AUC-ABB
GV1	0.985	0.966	0.865	0.948
SM	0.862	0.924	0.917	0.929
SE	0.985	0.739	0.500	0.922
GV3	0.973	0.952	0.962	0.918
<i>soft/symdiff</i>	0.902	0.941	0.95	0.959
DWH	0.985	0.966	0.865	0.948
HE	0.879	0.929	0.696	0.853

En conjunto, los resultados de *AUC* indican que los métodos de consenso *soft/symdiff* y GV3 ofrecen la mayor capacidad discriminante de manera

consistente entre los grupos genómicos, mientras que SE, se sitúa sistemáticamente como la alternativa menos eficaz, especialmente en el grupo AABB. Los algoritmos GV1, DWH, SM y HE conforman un grupo intermedio, con un desempeño aceptable pero inferior al de *soft/symdiff* y GV3. Este patrón refuerza la utilidad de los enfoques de consenso para mejorar la clasificación no supervisada de accesiones de *Musa* spp., al potenciar la separación entre grupos genómicos y reducir la ambigüedad en la identificación de accesiones, lo cual resulta particularmente relevante para la gestión de la diversidad genética y la definición de colecciones núcleo en bancos de germoplasma.

El análisis de la Tabla 11 demuestra que el método de consenso GV3 mejora sobre Lloyd en todas las métricas consideradas, reduciendo los errores de clasificación en un 2.5 %. Por su parte, *soft/symdiff* amplía esa ventaja, con una reducción de errores del 5 % y mejoras próximas o superiores al 1 % en Kappa, precisión y F1-score. Biológicamente, estas ganancias implican una asignación más fiable de accesiones a los grupos genómicos correctos, lo que se traduce en una selección de material vegetal más segura para programas de conservación y mejoramiento.

Desde el punto de vista biológico, estos resultados implican que el uso de métodos de consenso, en particular *soft/symdiff*, permite identificar grupos genéticos más puros y bien definidos. Esto facilita la caracterización, conservación y manejo eficiente de los recursos genéticos en bancos de germoplasma, reduciendo el riesgo de mezclas y mejorando la toma de decisiones para la conservación y el mejoramiento vegetal.

Tabla 11. Métricas globales y mejora relativa (%) de los métodos de consenso GV3 y *soft/symdiff* frente al mejor algoritmo particional tradicional (Lloyd) en *Musa* spp.

Métrica	Lloyd (Mejor Tradicional)	GV3	<i>soft/symdiff</i>	Mejora GV3 vs Lloyd (%)	Mejora <i>soft/symdiff</i> vs Lloyd (%)
Precisión global	0.880 (88.0%)	0.883 (88.3%)	0.886 (88.6%)	0.34	0.68
Kappa	0.841	0.845	0.849	0.48	0.95
Precisión (por clase)	0.889	0.9	0.902	1.24	1.46
Recall	0.881	0.884	0.887	0.34	0.68
F1-score	0.881	0.885	0.888	0.45	0.79
Índice de Rand ajustado	0.712	0.716	0.723	0.56	1.54
Fowlkes-Mallows	0.784	0.788	0.793	0.51	1.15
Pureza	0.881	0.884	0.887	0.34	0.68
Entropía	0.549	0.508	0.498	-7.47	-9.29
Reducción de errores (%)	—	2.5	5	2.5	5

Nota: La mejora relativa (%) de cada métrica se calculó tomando a Lloyd como referencia, y la reducción de errores (%) se basó en el error de clasificación definido como 1– precisión global.

El método consenso *Soft/symdiff* también logra mejoras notables en grupos específicos:

- **AAAB:** +2.6% en AUC (18.6% de reducción de error)
- **AAB:** +2.6% en AUC (29.3% de reducción de error)
- **AAB:** +3.5% en AUC (43.8% de reducción de error).

Solo en el grupo AAB se observa una ligera disminución en AUC (-1.0%), atribuible a su alta heterogeneidad interna.

Asimismo, al comparar ambos métodos de consenso, *soft/symdiff* supera de forma consistente a GV3 en todas las métricas globales, lo que la identifica

como la opción más eficiente y robusta para la clasificación no supervisada en *Musa* spp. En este contexto, se recomienda priorizar su uso en bancos de germoplasma, reservando GV3 como alternativa complementaria para validaciones metodológicas. La implementación de esta función no solo garantiza clasificaciones reproducibles, sino que también facilita el manejo estratégico de las colecciones, con implicaciones directas para la conservación y el uso sostenible del germoplasma.

Aunque la reducción del error de clasificación pueda parecer modesta en términos absolutos, debe interpretarse en

el contexto de modelos altamente precisos y escenarios de clasificación multiclase, donde las mejoras marginales son difíciles de alcanzar y pueden tener un impacto práctico considerable. En estudios previos sobre evaluación de clasificadores se ha señalado que incrementos del orden de 0.5–1 % en la precisión pueden considerarse relevantes en contextos similares (Hastie *et al.*, 2009; Fawcett, 2006; Kuhn y Jonhson, 2013). En este marco, una reducción del 5 % en los errores de clasificación representa un avance sustancial en robustez y exactitud para aplicaciones reales en bancos de germoplasma.

Para visualizar el rendimiento obtenido, a continuación, se presentan la matriz de confusión, el mapa de calor (*heatmap*) correspondiente, el gráfico de Silueta y las curvas ROC multiclase para el método de consenso *soft/symdiff*.

La tabla 12 resume el número de accesiones clasificadas correctamente (diagonal principal) y los errores de

clasificación (valores fuera de la diagonal) para cada grupo genómico. Una alta concentración de valores en la diagonal indica buena correspondencia entre la clasificación obtenida y la biológica de referencia.

Tabla 12. Matriz de confusión para la clasificación consenso (*soft/symdiff*).

Predicho/ Actual	AAAB	AAB	AABB	ABB
AAAB	67	10	0	0
AAB	0	72	0	0
AABB	17	0	84	9
ABB	0	2	0	75

En el mapa de calor (Figura 1), la intensidad del color refleja la proporción de accesiones clasificadas en cada combinación de clase real y predicha, permitiendo visualizar de forma intuitiva las coincidencias y discrepancias más relevantes entre grupos.

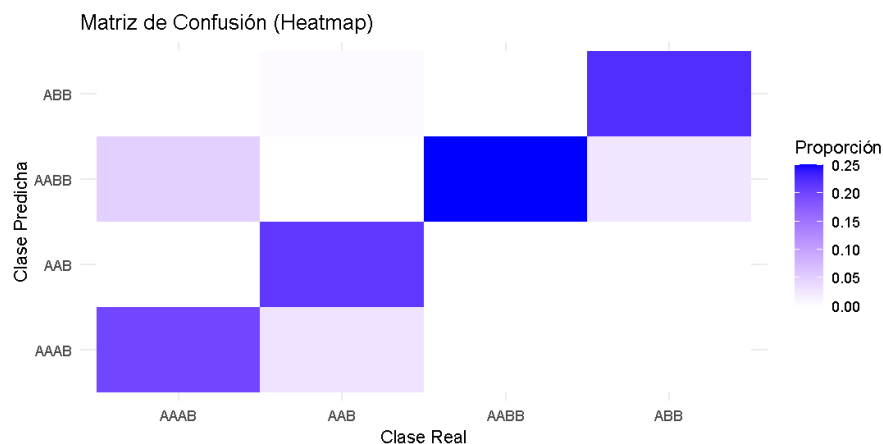


Figura 1. Mapa de calor de la matriz de confusión para la clasificación consenso por el algoritmo *soft/symdiff*.

El gráfico de Silueta (Figura 2) representa la cohesión interna y separación entre los clústeres formados. Un valor promedio de Silueta moderado (0.26) indica que,

aunque existe una estructura de agrupamiento razonable, persiste cierto grado de solapamiento entre algunos grupos.

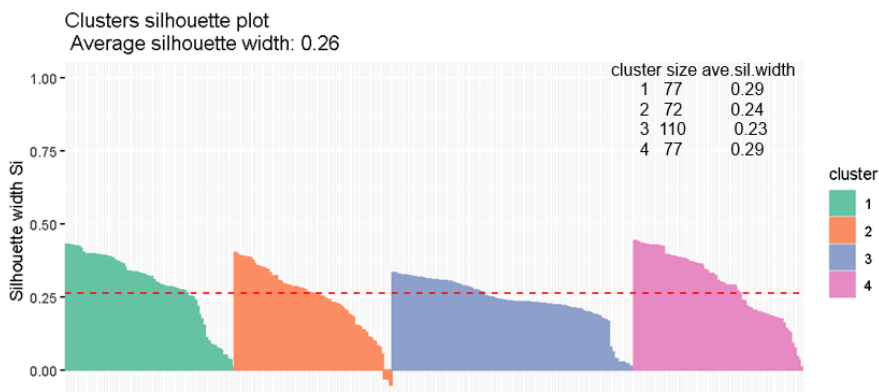


Figura 2. Gráfico de Silueta para los clústeres generados por el algoritmo consenso *soft/symdiff*.

Las curvas ROC (Figura 3) muestran la capacidad discriminante del modelo para cada grupo genómico. Los valores elevados de AUC en todas las clases

sugieren un buen desempeño global del método, especialmente en la identificación de los grupos ABB y AAB.

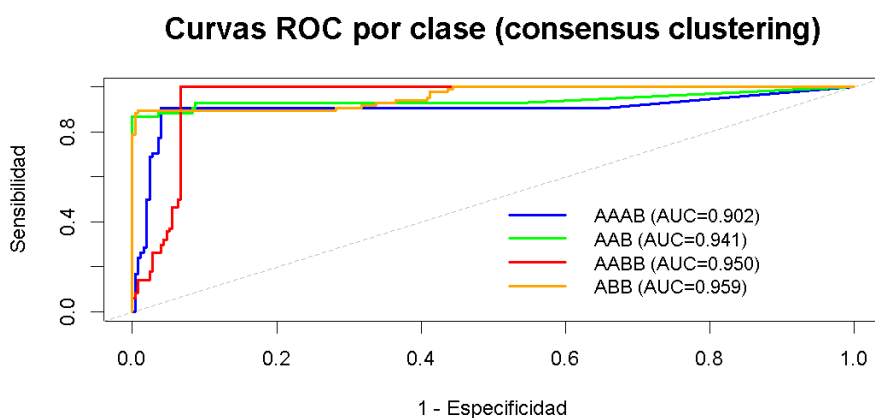


Figura 3. Curvas ROC por clase (AUC) para el algoritmo consenso *soft/symdiff*.

La variabilidad en el rendimiento de los métodos de agrupamiento, así como las diferencias observadas en la discriminación de los grupos genómicos, evidencian la complejidad inherente a las colecciones de *Musa* spp. y subrayan la importancia de emplear enfoques metodológicos robustos.

Si bien el balanceo por simulación permitió evaluar los métodos de agrupamiento en condiciones controladas y equitativas, es importante reconocer que este enfoque no necesariamente refleja la estructura real de las colecciones de germoplasma, las cuales

suelen estar desbalanceadas en la práctica. En estudios anteriores, se realizaron análisis complementarios empleando la base de datos original no balanceada, lo que permitió validar la robustez y aplicabilidad de los métodos seleccionados en escenarios más cercanos a la realidad. Aunque el balanceo optimiza la detección de patrones en todas las clases, los resultados obtenidos con los datos originales evidencian la tendencia natural a una mayor dificultad en la discriminación de los grupos minoritarios. Esto refuerza la necesidad de interpretar

las conclusiones considerando las particularidades y el contexto específico de cada colección analizada.

En síntesis, este estudio amplía el conocimiento sobre la integración de técnicas de agrupamiento particional y consenso en la clasificación de bancos de germoplasma vegetal. La adopción de enfoques más robustos y precisos permite avanzar significativamente en el mejoramiento genético y la conservación del germoplasma, aportando así al fortalecimiento de la sostenibilidad y la diversidad agrícola.

CONCLUSIONES

1. El análisis comparativo de los métodos de agrupamiento particional mostró que el algoritmo de Lloyd fue el que alcanzó el mejor desempeño global, aunque su capacidad para reproducir la clasificación biológica de referencia varió entre los diferentes grupos genómicos evaluados.
2. Los métodos de consenso mejoraron de forma consistente la calidad de la clasificación no supervisada respecto al mejor método particional, con incrementos en precisión global y *F1-score* y una reducción del error de clasificación de hasta un 5 %, manteniendo al mismo tiempo niveles aceptables de cohesión interna de los clústeres.
3. Entre los enfoques de consenso evaluados, *soft/symdiff* se identificó como la alternativa más eficiente y robusta, al combinar la mayor correspondencia con la clasificación genómica de referencia con la mejor capacidad discriminante por grupo. GV3 se confirmó como una opción complementaria, con resultados muy próximos a *soft/symdiff*.
4. El flujo metodológico implementado integración de variables cualitativas y cuantitativas mediante ACM y

normalización, balanceo sintético de clases y evaluación con métricas internas y externas constituye una herramienta replicable para la clasificación de accesiones de *Musa* spp., y ofrece una base sólida para apoyar la gestión y utilización de recursos fitogenéticos en bancos de germoplasma.

BIBLIOGRAFÍA

- Borges, M.; L.M. González; N. Aguilera; B. Malaurie; J. Vázquez; Z. Infante y M. Fonseca. 2002. Aplicación de la estadística multivariada al estudio de la conservación in vitro de germoplasma de *Dioscorea alata* L. En Congreso Científico del INCA (13:2002, nov 12-15, La Habana) *Memorias*. CD-ROM. Instituto Nacional de Ciencias Agrícolas. ISBN 959-7023-22-9.
- Benzécri, J.P. 1979. Sur l'analyse d'un tableau dont l'une des colonnes à un poids prédominant. *Les cahiers de l'analyse des données*, Volume 4 no.4, pp. 413-416.
- Chawla, N.V.; K.W. Bowyer; I.O. Hall and W.P. Kegelmeyer. 2002. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16: 321-357.
- Everitt, B.S.; S. Landau; M. Leese and D. Stahl. 2011. *Cluster Analysis* (5th ed.). Wiley.
- Fawcett, T. 2003. ROC Graphs: Notes and Practical Considerations for Data Mining Researchers. Technical Report HPL-2003-4. Hewlett-Packard Laboratories, Palo Alto, California, Estados Unidos.
- Forgy, E.W. 1965. Cluster analysis of multivariate data: efficiency vs interpretability of classifications. *Biometrics*, 21:768-769.
- Fowlkes, E.B. and C.L. Mallows. 1983. A method for comparing two

- hierarchical agrupamientos. *Journal of the American Statistical Association*, 78(383), 553-569. <https://doi.org/10.1080/01621459.1983.10478008>.
- Gatnar, E. and M. Walesiak. 2004. *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*. Wrocław: Wydawnictwo Akademii Ekonomicznej we Wrocławiu, 393 p.
- Guttman, L. 1953. Image theory for the structure of quantitative variates. *Psychometrika*, 18, 277-296.
- Greenacre, M.J. 1984. *Theory and Applications of Correspondence Analysis*. Pp.334. Academic Press, London.
- Greenacre, M. 2017. *Correspondence analysis in practice* (3rd Ed.), pp. 74-76.
- Hartigan, J.A. and M.A. Wong. 1979. Algorithm AS 136: A K-means agrupamiento algorithm. *Applied Statistics*, 28:100-108.
- Hastie, T.; R. Tibshirani and J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer. Disponible en: <https://web.stanford.edu/~hastie/ElemStatLearn/>.
- Husson, F.; S. Le and J. Pagès. 2017. Exploratory multivariate analysis by example using R. *CRC press*, 15:1-60.
- Hornik, K.; C. Buchta and N. X. Vinh. 2023. clue: Cluster Ensembles (Version 0.3-65) [R package]. <https://CRAN.R-project.org/package=clue>.
- Hubert, L. and P. Arabie. 1985. Comparing partitions. *Journal of Classification*, 2(1), 193-218. <https://doi.org/10.1007/BF01908075>.
- IPGRI-INIBAP/CIRAD. 1996. *Descriptores para el banano (Musa spp.)*. Instituto Internacional de Recursos Fitogenéticos, Roma, Italia;
- Red internacional para el mejoramiento del Banano y el Plátano, Montpellier, Francia; Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Montpellier, Francia. ISBN 92-9043-307-8.
- Jain, A.K. 2010. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651-666. <https://doi.org/10.1016/j.patrec.2009.09.011>.
- Kaufman, L. and P. J. Rousseeuw. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, New York, 342 p.
- Kaiser, H.F. 1960. The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.
- Kohavi, R. and F. Provost. 1998. Glossary of terms. *Machine Learning*, 30(2-3), 271-274. <https://doi.org/10.1023/A:1017181826899>
- Kuhn, H.W. 1955. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2), 83-97. <https://doi.org/10.1002/nav.3800020109>
- Kuhn, M. 2008. Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(5), 1-26. <https://www.jstatsoft.org/article/view/v028i05>.
- Kuhn, M. and K. Johnson. 2013. *Applied Predictive Modeling*. Springer, New York.
- Kuhn, M.; D. Vaughan and J. Silge. 2024. *Yardstick: Tidy Characterizations of Model Performance* (Version 1.2.0) [R package]. <https://yardstick.tidymodels.org/>.
- Lunardon, N.; G. Menardi y N. Torelli. 2014. ROSE: A Package for Binary Imbalanced Learning. *The R Journal*, 6(1): 79-89. DOI: 10.32614/RJ-2014-008.

- Lloyd, S.P. 1982. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2): 129-137.
- Morejón, R.; H.D. Sandra y DE J.P. Noraida. 2002. Técnicas multivariadas en la clasificación morfoagronómica de genotipos de arroz (*Oryza sativa* L.) obtenidos en la estación experimental "Los Palacios". En Congreso Científico del INCA (13:2002, nov 12-15, La Habana) Memorias. CD-ROM. Instituto Nacional de Ciencias Agrícolas, 2002. ISBN 959-7023-22-9.
- Milián, M.; Molina, O. and Y. Figueroa. 2018. Integrated Characterization of Cuban Germplasm of Cocoyam (*Xanthosoma Sagittifolium* (L.) Schott). *Journal of Plant Genetics and Crop Research*, 1(1):1-18.
- Macqueen, J. 1967. Some methods for classification and analysis of multivariate observations. In: L. M. Le Cam y J. Neyman (eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, University of California Press, Berkeley, pp. 281-297.
- Manning, C.D.; P. Raghavan and H. Schütze. 2008. Introduction to Information Retrieval. Cambridge University Press.
- Meilă, M. 2003. Comparing clusterings by the variation of information. En: B. Schölkopf y M. K. Warmuth (eds.), *Learning Theory and Kernel Machines* (COLT/Kernel 2003, Washington, DC). Lecture Notes in Computer Science, vol. 2777. Springer, pp. 173–187.
- Maechler, M.; P. Rousseeuw; A. Struyf; M. Hubert and K. Hornik. 2023. cluster: Cluster Analysis Basics and Extensions (Version 2.1.6) [R package]. <https://CRAN.R-project.org/package=cluster>.
- Monti, S.; P. Tamayo; J. Mesirov and T. Golub. (2003). Consenso agrupamiento: a resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learning*, 52(1-2), 91-118. <https://link.springer.com/article/10.1023/A:1023949509487>.
- Provost, F. and T. Fawcett. 2001. Robust classification for imprecise environments. *Machine Learning*, 42(3), 203-231. <https://doi.org/10.1023/A:1007601015854>.
- Powers, D.M.W. 2011. Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies*, 2, 37–63.
- R Core Team. 2024. R: A Language and Environment for Statistical Computing (version 4.3.3). R Foundation for Statistical Computing. Vienna, Austria.
- Ruiz De Galarreta, J.I. 1998: Agrupación de poblaciones locales de maíz (*Zea mays* L.) mediante caracteres morfológicos y parámetros ambientales, Servicio de Publicaciones de la Universidad de Lleida, p.161.
- Robin, X.; N. Turck; A. Hainard; N. Tiberti; F. Lisacek; J.C. Sánchez and M. Müller. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12, 77. <https://doi.org/10.1186/1471-2105-12-77>
- Rousseeuw, P.J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).

- Shannon, C.E. 1948. A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379-423.
- Strehl, A. and J. Ghosh. 2002. Cluster Ensembles - A Knowledge Reuse Framework for Combining Multiple Partitions. *Journal of Machine Learning Research*, 3, 583-617.
- Saito, T. and M. Rehmsmeier. 2015. *The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets*. *PLoS ONE*, 10(3): e0118432. <https://doi.org/10.1371/journal.pone.0118432>
- Swets, J.A.; R.M. Dawes and J. Monahan. 2000. Better decisions through science. *Scientific American*, 283(4), 82-87. <https://www.jstor.org/stable/26058990>.
- Vega-Pons, S. and J. Ruiz-Shulcloper. 2011. A survey of agrupamiento ensemble algorithms. *International Journal of Pattern Recognition and Artificial Intelligence*, 25(3), 337-372. <https://arxiv.org/abs/1105.1421>.
- Walesiak, M. and A. Dudek. 2020. clusterSim: Searching for Optimal

- Agrupamiento Procedure for a Data Set (Version 0.49-2) [R package]. <https://CRAN.R-project.org/package=clusterSim>.
- Wickham, H. and J. Bryan. 2023. *R packages*. "O'Reilly Media, Inc."

ANEXOS

Script 1

```
## ---- Pureza y entropía ponderada -----  
## cont: matriz de contingencia (filas =  
      clases, columnas = clústeres)  
pureza_entropia <- function(cont) {  
  N <- sum(cont) # total de accesiones  
  purity <- sum(apply(cont, 2, max)) / N #  
    pureza global  
  # entropía de un clúster  
  entropy_cluster <- function(v) {  
    p <- v[v > 0] / sum(v)  
    -sum(p * log2(p))  
  }  
  ent_vec <- apply(cont, 2, entropy_cluster)  
  # entropía por clúster  
  ent_total <- sum(colSums(cont) * ent_vec) /  
    N # entropía ponderada  
  c(pureza = purity, entropia = ent_total)}
```

Recibido: 28 de julio de 2025; Aceptado: 3 de diciembre de 2025

Conflicto de intereses: Los autores declaran no tener conflictos de intereses.

Contribución de los autores:

Conceptualización y curación de datos: Osmany Molina Concepción, Lianet Gonzáles Díaz

Investigación: Osmany Molina Concepción, Lianet Gonzáles Díaz

Software: Osmany Molina Concepción, Yaselis Guillen López y Carmen C. Pons Pérez

Validación: Osmany Molina Concepción, Lianet Gonzáles Díaz

Escritura-borrador original: Osmany Molina Concepción

Redacción-revisión y edición: Osmany Molina Concepción, Yaselis Guillen López y Carmen C. Pons Pérez

Administración de proyectos: Osmany Molina Concepción

Ética: El autor para la correspondencia confirma que todos los demás autores han leído y aprobado el manuscrito y que no existen cuestiones éticas involucradas.

La referencia a marcas comerciales de equipos, instrumentos o materiales específicos se realiza únicamente con fines de identificación, sin que ello implique ningún compromiso promocional por parte de los autores ni del editor.