

ÁRBOLES CONSENSO EN LA CLASIFICACIÓN DE BANCOS DE GERMOPLASMA VEGETAL

Osmany Molina Concepción*, Marilys Milián Jiménez¹, Carmen C. Pons Pérez¹ y Ricardo Grau Ábalo²

¹Instituto de Investigaciones de Viandas Tropicales (INIVIT), Cuba. INIVIT, Apartado 6, Santo Domingo, CP: 53 000, Villa Clara, Cuba.

²Universidad Central "Marta Abreu" de Las Villas (UCLV), Cuba.

*Autor para la correspondencia: taxonumeric@inivit.cu

RESUMEN

Partiendo de que no existe todavía un algoritmo de clasificación por excelencia que logre encontrar una mejor estructura para separar las accesiones y, que en la búsqueda de mejores algoritmos de clasificación aparece una tendencia a la combinación de algoritmos de agrupamiento en el mismo problema, el presente trabajo tiene como objetivo combinar varios algoritmos de agrupamiento en la conformación de conglomerados jerárquicos aglomerativos, en aras de lograr un mejor rendimiento y calidad de los resultados alcanzados por los algoritmos individuales y compensar posibles errores cometidos en el desempeño de cada uno sobre la clasificación del germoplasma en estudio. La investigación se realizó en la colección de malanga (*Xanthosoma* spp.) que se conserva en el germoplasma del INIVIT. Para combinar los resultados de los diferentes algoritmos de aglomeración se usó la función *cl_ensemble* y el árbol consenso se obtuvo con la función *cl_consensus* del paquete "clue", utilizando los tres métodos disponibles en esta función para calcular los conglomerados consenso en los algoritmos jerárquicos (*euclidean*, *manhattan* y *majority*). El árbol consenso con el método *manhattan* es el que mejor representa la estructura de la base de datos de malanga en estudio, sin embargo, no muestra estabilidad en diferentes corridas. En los análisis se utilizaron funciones implementadas sobre la base del lenguaje de programación R. Se demostró que la combinación de agrupamientos para lograr un dendrograma consenso es una técnica efectiva en la conformación de estructuras consistentes.

Palabras clave: bancos de germoplasma, conglomerados, consenso, taxonomía numérica.

CLUSTER ENSEMBLES IN THE CLASSIFICATION OF PLANT GERMOPLASMA BANKS

ABSTRACT

The present paper aims to combine several clustering algorithms in the conformation of agglomerative hierarchical clusters, based on the fact that there is not a classification algorithm *par excellence* that manages to find a better structure to separate the accessions and, that in the search for better algorithms of classification there is a tendency to combine clustering algorithms in the same problem. This study will contribute to attain a better performance and to improve quality of the results achieved by individual algorithms and to compensate for possible errors in the performance of each one on the classification of the germplasm under study. The research was carried out in the germplasm collection of cocoyam (*Xanthosoma* spp.) preserved in INIVIT's

germplasm. *CL_ensemble* function was used to combine the results of the different clustering algorithms, and the cluster ensemble was obtained with the *cl_consensus* function of the "clue" package, using the three methods available in this function to calculate the consensus clusters in the hierarchical algorithms (*euclidean*, *manhattan* y *majority*). The cluster ensemble with the *manhattan* method is the one that best represents the structure of the cocoyam database, however, it does not show stability in different runs. Implemented functions were used in the analyzes based on the R programming language. It was demonstrated that the combination of clusters to achieve a consensus dendrogram is an effective technique in the conformation of consistent structures.

Keywords: germplasm banks, clusters, consensus, numerical taxonomy.

INTRODUCCIÓN

Los métodos que se utilizan generalmente en el estudio de divergencias entre individuos siguen una aproximación fenética o numérica (Franco y Hidalgo, 2003).

La taxonomía numérica intenta construir clasificaciones "naturales", basadas en las semejanzas fenotípicas de los individuos, que se valoran partiendo de una adecuada elección de un coeficiente de similaridad (Cuadras, 1981). El problema que presenta es que no hay ningún método que permita asegurar que los grupos resultantes son naturales (descendencia de un ancestro común), ya que el resultado obtenido depende del orden de adición y no se pueden contrastar hipótesis.

La versatilidad del análisis de conglomerados jerárquicos radica en la posibilidad de utilizar distintos tipos de medidas para estimar la distancia existente entre los casos y la posibilidad de seleccionar uno entre una gran variedad de métodos. Pero no existe todavía un algoritmo de clasificación por excelencia, pues para un problema determinado es difícil seleccionar cuál será el método de aglomeración que logre encontrar una mejor estructura para separar las accesiones.

Para resolver en parte esta problemática se ha sugerido la idea de la combinación de agrupamientos (Clustering Ensemble) (Vega y Ruiz, 2010). Además, si no hay información *a priori* sobre las bases de datos en estudio, hacer una combinación de la diversidad de resultados responde a la idea intuitiva de que, si no se conoce la calidad de ciertos resultados individuales, la opción de combinarlos puede ser superior a seleccionar algún resultado simple, ya que mientras más diverso es el conjunto de resultados, mayor es la información disponible en la combinación.

En este trabajo se propone como objetivo combinar varios algoritmos de agrupamiento en el mismo problema sobre el germoplasma de malanga (*Xanthosoma* spp.).

MATERIALES Y MÉTODOS

El banco de germoplasma del Instituto de Investigaciones de Viandas Tropicales (INIVIT) se creó en 1967, lo conforman accesiones introducidas por diferentes vías: colectas nacionales e internacionales, introducciones desde otros países e instituciones, donaciones y material genético derivado de los programas de mejoramiento.

Para realizar la investigación, se usaron datos procedentes de un estudio de los diferentes genotipos de la colección de prueba de malanga (*Xanthosoma* spp.) (Milián, 2008) del Banco de Germoplasma, que se conserva en el INIVIT.

Dentro de la colección cubana de malanga se incluyen 57 accesiones de la especie *Xanthosoma sagittifolium* (L.) Schott (SAG) y 10 accesiones que conforman un grupo

que no tienen una especie definida, o cuya clasificación específica no está definida de *Xanthosoma* sp.(SP). Además, está presente la especie *Xanthosoma brasiliense* Engl. (BRA), cuyo único representante es el clon 'Belembe'; una de la especie *Xanthosoma atrovirens* Koch & Bouché (ATR), constituida por el clon 'Amarilla Trinidad'; mientras que la accesión 'Jardín' representa a la especie *X. nigrum* (Vell.) Mansf. (NIG).

En este estudio se evalúan variables continuas y categóricas en cada accesión o cultivar de los bancos de germoplasma, no existe una solución universal al problema de cómo combinarlas para construir una medida de distancia. Entre las alternativas metodológicas para abordar este problema está la métrica de Gower (1971). La cual está implementada en la función *gowdis* (), descrita en el paquete *FD*.

En esta investigación se usaron los métodos de aglomeración jerárquicos de Ward o varianza mínima (Ward, 1963), Promedio ó UPGMA (Sneath y Sokal, 1973), agrupación de enlace simple (Gower, 1967), agrupación de enlace completo (Sorensen, 1948) con la función *hclust*() en el paquete *stats* que forma parte de los paquetes de R que se instalan por defecto.

Una vez obtenido el resultado del método de aglomeración, en el análisis taxonómico es importante determinar si el conjunto de datos muestra una tendencia a formar grupos, lo cual se determina a través del coeficiente aglomerativo (CA) con la función *coef.hclust*() en el paquete *cluster*. Por otra parte, los métodos jerárquicos imponen cierta estructura sobre los datos y es necesario con frecuencia considerar si es aceptable o si se introducen distorsiones inaceptables en las relaciones originales. El método más usado para verificar este hecho, es el Coeficiente de Correlación Cofenético (CCC) (Sokal and Rohlf, 1962).

El paquete *clue* (Hornik, 2011) permite crear y analizar combinación de agrupamientos y obtener una estructura consenso. Para aglutinar los resultados de los diferentes algoritmos de aglomeración se usó la función *cl_ensemble* () de este paquete.

Para procesar la información se utilizó el lenguaje de programación, orientado a objetos, denominado R (R Development Core Team, 2016).

RESULTADOS Y DISCUSIÓN

Al aplicar la métrica de Gower para variables mixtas a la matriz de datos de variables cualitativas y cuantitativas de la colección de malanga (*Xanthosoma* spp.), se obtuvo una matriz de distancia entre las accesiones, a la que se le aplicaron los cuatro métodos de aglomeración seleccionados para el estudio.

Como se observa en la Tabla 1, el método Ward es el que mayor CA tiene para las diferentes bases de datos, por lo cual hay estructuras bien definidas, y el de menor valor de CCC con relación a los demás métodos y UPGMA fue superior en el CCC, por lo cual es el que menos distorsiona la matriz de distancia original.

Tabla 1. Resumen de los coeficientes CA y CCC para los cuatros métodos de conglomerados en estudio.

Bases de datos	Métodos		Métrica de Gower		
			CA		CCC
malanga (<i>Xanthosoma</i> spp.)	Ward	1	0,885	4	0,578
	Promedio	3	0,558	1	0,762
	Simple	4	0,466	2	0,701
	Completo	2	0,632	3	0,628

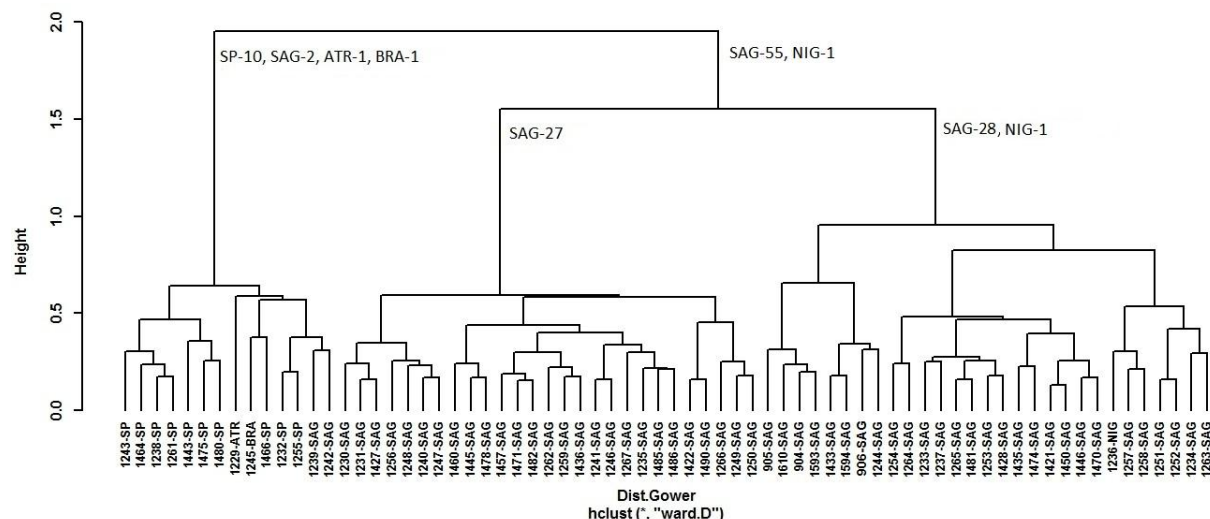
Los dendrogramas (figura 1) obtenidos para los cuatro métodos de aglomeración analizados con la base de datos en estudio, tienen una relativa diferencia de estructura entre ellos, dentro de la base de datos; en el método de Ward se definen claramente a la altura de 1,7 dos grupos bien separados: el grupo uno está formado por una mezcla de 10 accesiones de SP, dos de la especie SAG, una de la especie ATR y una de la especie BRA, el grupo dos está formado por 55 accesiones de la especie SAG y una de la especie NIG, la cual no debe de estar ubicada en este grupo donde la mayoría es de la especie *Xanthosoma sagittifolium* (L.) Schott.

En el dendrograma obtenido por el método promedio haciendo un corte a la altura 0,4 se forman cuatro grupos: el grupo uno tiene una accesión de la especie ATR, el dos una de la especie BRA, el grupo tres, mezcla dos de la especie SAG y 10 de la SP, el grupo cuatro mezclas 55 accesiones de la especie SAG con una de la especie NIG.

El método de aglomeración de enlace simple no forma un dendrograma con témpanos bien identificados.

El método de aglomeración de enlace completo establece un dendrograma con dos grupos bien definidos a la altura de 0,6: El grupo uno está formado por las 55 accesiones de la especie SAG y una de la especie NIG, el grupo dos tiene una gran mezcla formada por 10 accesiones de SP, seis de la especie SAG, una de la especie ATR y una de la especie BRA.

Como se constata en los dendrogramas de la figura 1 la agrupación exacta no es una tarea sencilla, y es difícil de hacer recomendaciones generales. Teniendo en cuenta los resultados obtenidos, es conveniente aglutinar el de los diferentes algoritmos de aglomeración en aras de lograr una mejor calidad del resultado final en comparación con el alcanzado por los algoritmos individuales, y compensar posibles errores en el desempeño de cada uno.



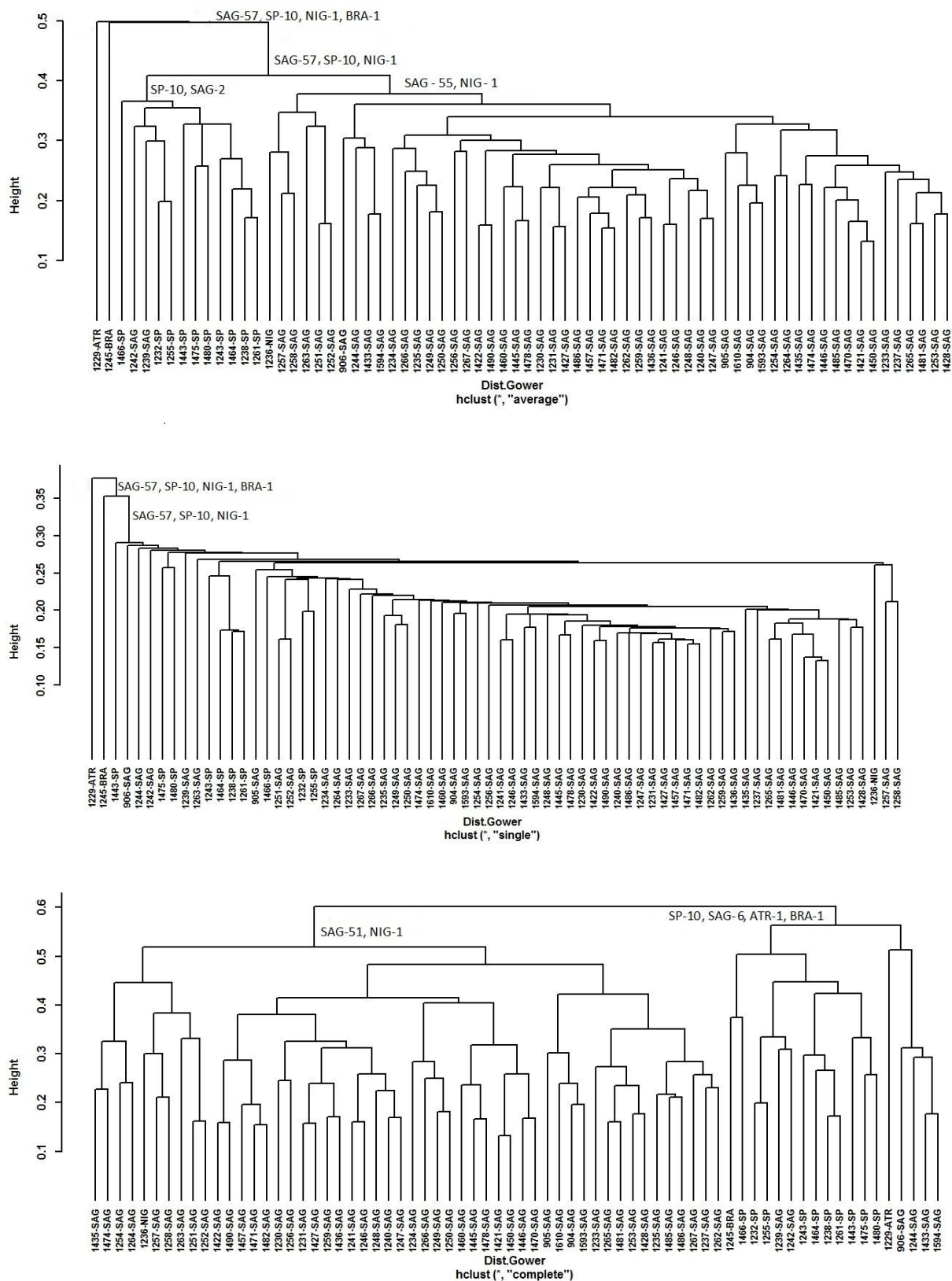
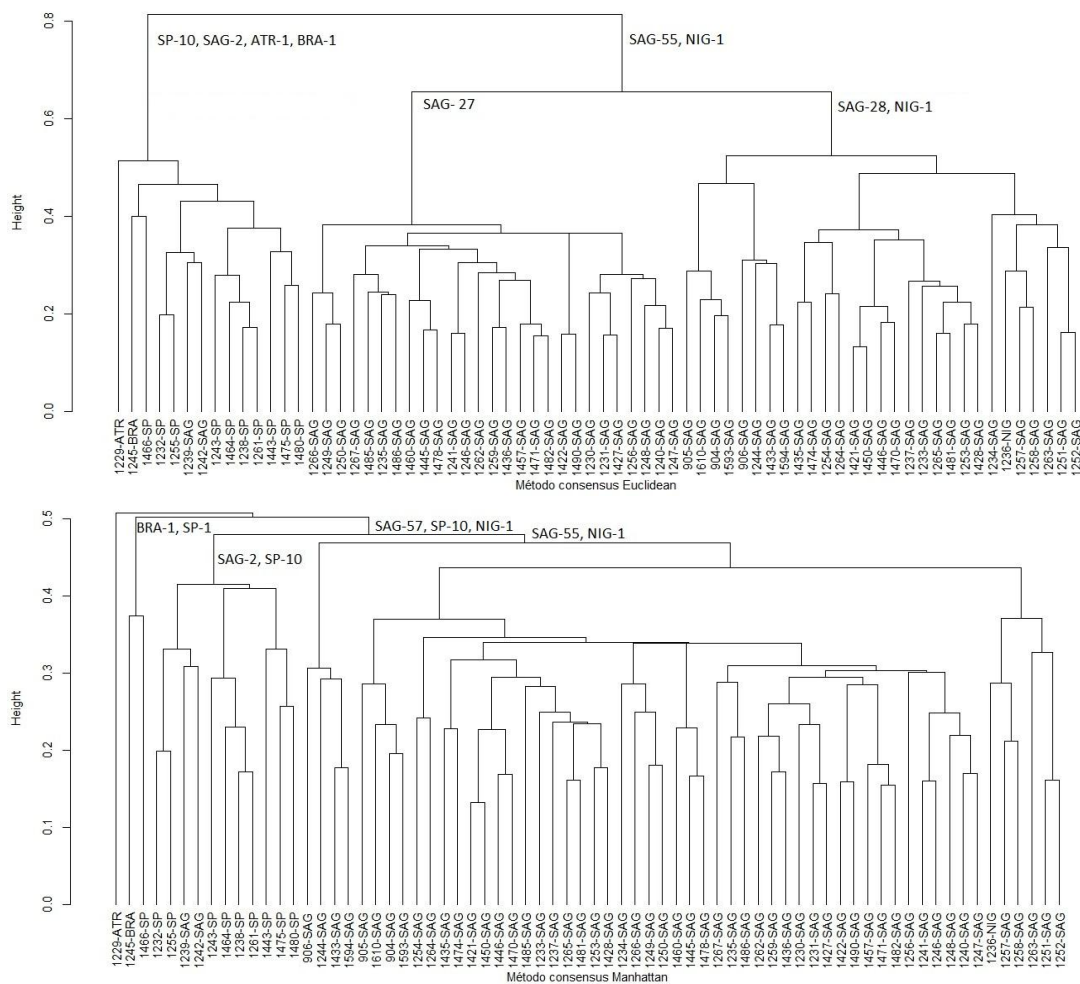


Figura 1. Dendrogramas de los cuatro métodos de conglomerados en estudio. Base de datos de la colección del germoplasma de malanga (*Xanthosoma* spp.).

A partir de la combinación de agrupamientos, se obtienen tres árboles consensos por los métodos *euclidean*, *manhattan* y *majority* (Figura 2).

Con el método consenso de la distancia *Euclidiana* se obtiene un dendrograma con una formación clara de dos grupos a la altura de 0,75 con igual distribución de las accesiones por grupo si se estableciera un corte para dos grupos en el dendrograma de la figura 1 por el método de Ward.



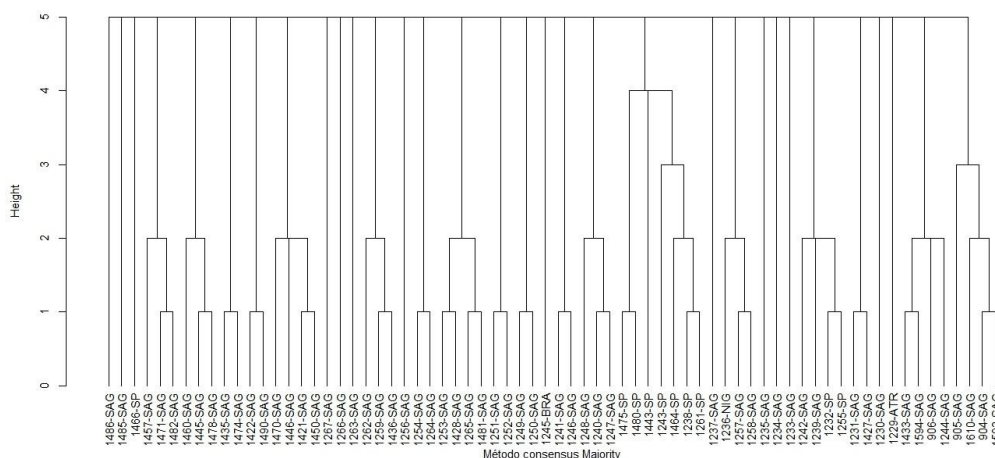


Figura 2. Dendrogramas consensos con los métodos *euclidean*, *manhattan* y *majority*. Base de datos de la colección del germoplasma de malanga (*Xanthosoma* spp.).

En el dendrograma consenso obtenido por el método *Manhattan* de la figura 2, para un corte que conforma cuatro, se representa con claridad la estructura de la base de datos en estudio; en el primer grupo está la única accesión de la especie ATR, un segundo grupo está formando por la accesión “1245” de la especie BRA y la accesión “1446” de la SP, lo cual sugiere hacer un estudio molecular de esta accesión para definir con claridad su ubicación, ya que, excepto el dendrograma resultante del método promedio (Figura 1), todos los demás ubican estas dos accesiones del grupo dos muy cercanas. El tercer grupo lo forman 10 de la SP y dos de la especie SAG, este resultado también sugiere el estudio molecular de estas dos accesiones para precisar con mayor detalle su ubicación biológica real. El cuarto grupo está formando por 55 de la especie SAG y la accesión “1236” de la especie NIG, la cual también debe ser evaluada por análisis molecular para definir con mayor precisión su ubicación biológica.

Por su parte el método *majority* no logra una clara descripción biológica de la base de datos. En múltiples iteraciones con este método, se obtienen dendrogramas iguales sin variaciones en la estructura de los grupos, en altura y en el orden de las accesiones.

Sin embargo, en múltiples iteraciones para el método *Euclidean*, se obtienen dendrogramas con pequeñas variaciones en la estructura de los grupos inferiores (Figura 3), en cuanto a altura y orden de las accesiones dentro de los dos grupos principales, pero no entre ellos.

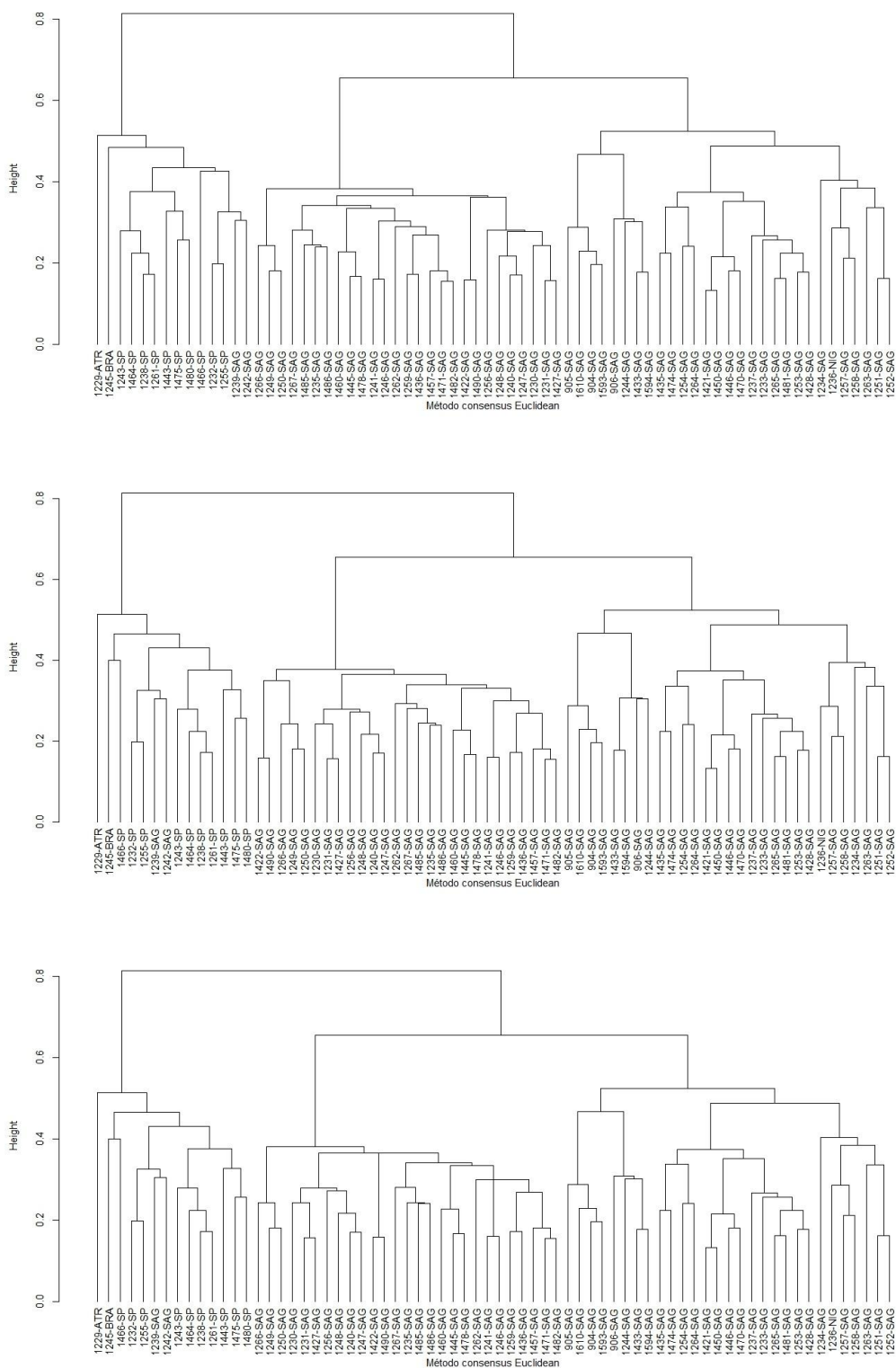


Figura 3. Dendrogramas consensos con el método *Euclidean*, en una segunda, tercera y cuarta iteración. Base de datos de la colección del germoplasma de malanga (*Xanthosoma* spp.).

A continuación, se puede observar los pequeños cambios en cuanto a la altura y el orden para las iteraciones dos, tres y cuatro que representa los dendrogramas de la figura 3.

Iteración 2:

[1] 0.1322872 0.1544466 0.1568090 0.1589904 0.1601788 0.1611567 0.1614936 0.1668220
[9] 0.1697903 0.1716675 0.1718215 0.1770497 0.1773851 0.1801870 0.1806976 0.1814182
[17] 0.1955013 0.1986530 0.2113122 0.2144644 0.2166725 0.2234201 0.2235752 0.2238994
[25] 0.2278911 0.2285260 0.2395586 0.2412119 0.2427746 0.2427772 0.2442587 0.2559144
[33] 0.2569847 0.2664453 0.2690449 0.2726581 0.2792568 0.2796412 0.2803175 0.2863452
[41] 0.2871000 0.2935726 0.2996724 0.3040317 0.3054362 0.3066601 0.3256143 0.3267365
[49] 0.3309531 0.3356942 0.3362180 0.3395267 0.3489901 0.3511214 0.3654090 0.3734701
[57] 0.3752495 0.3766578 0.3821809 0.3936618 0.3998914 0.4304844 0.4652088 0.4676343
[65] 0.4872875 0.5142826 0.5246000 0.6548871 0.8134861

Orden:

[1] 4 20 56 7 30 14 17 18 55 13 35 49 60 62 43 67 40 24 25 5 6 44 31 23 15 22 36
[28] 41 65 10 66 54 50 61 16 21 34 48 53 58 64 2 70 1 68 46 69 3 19 47 59 29 38 42
[55] 52 51 57 12 8 39 63 28 45 11 32 33 9 37 26 27

Iteración 3:

[1] 0.1322649 0.1543878 0.1569035 0.1590023 0.1601714 0.1610788 0.1614916 0.1667345
[9] 0.1697564 0.1716416 0.1718114 0.1770930 0.1774411 0.1801344 0.1806906 0.1814097
[17] 0.1955468 0.1986456 0.2113635 0.2144626 0.2166706 0.2233918 0.2235313 0.2238747
[25] 0.2279304 0.2285866 0.2411230 0.2415964 0.2427837 0.2428186 0.2433013 0.2559745
[33] 0.2569875 0.2664584 0.2690697 0.2728010 0.2792954 0.2796389 0.2803218 0.2862971
[41] 0.2870755 0.2994150 0.2994389 0.3018670 0.3054588 0.3090105 0.3256424 0.3268288
[49] 0.3334144 0.3361630 0.3382785 0.3412681 0.3511918 0.3651344 0.3652788 0.3731667
[57] 0.3752499 0.3814116 0.3834089 0.3998552 0.4032467 0.4304939 0.4652088 0.4676951
[65] 0.4873072 0.5142824 0.5245943 0.6548881 0.8134833

Orden:

[1] 4 20 56 7 30 14 17 18 55 13 35 49 60 62 40 24 25 5 6 44 31 23 15 22 43 67 41
[28] 10 65 66 54 50 61 36 16 21 34 48 53 58 64 2 70 1 68 3 19 46 69 47 59 29 38 42
[55] 52 51 57 12 8 39 63 28 45 9 11 32 33 37 26 27

Iteración 4:

[1] 0.1321623 0.1542516 0.1566848 0.1595259 0.1595303 0.1612629 0.1616267 0.1662975
[9] 0.1698943 0.1712353 0.1715433 0.1768123 0.1775846 0.1805434 0.1812367 0.1814098
[17] 0.1949558 0.1985498 0.2110767 0.2144847 0.2171329 0.2232382 0.2235273 0.2235792
[25] 0.2280716 0.2289602 0.2394932 0.2413199 0.2428153 0.2429311 0.2446911 0.2559220
[33] 0.2568152 0.2662492 0.2717540 0.2724908 0.2762099 0.2793913 0.2796101 0.2798851
[41] 0.2861921 0.2870409 0.3026236 0.3046395 0.3059244 0.3087775 0.3257937 0.3272822
[49] 0.3326064 0.3347346 0.3370778 0.3403514 0.3509238 0.3625610 0.3660444 0.3733559
[57] 0.3752279 0.3816354 0.3836205 0.4003960 0.4027860 0.4303872 0.4653240 0.4675033
[65] 0.4872751 0.5141424 0.5246732 0.6548816 0.8134868

Orden:

[1] 4 20 56 7 30 14 17 18 55 13 35 49 60 62 40 24 25 41 65 10 66 54 50 61 16 21 53
[28] 58 64 36 34 48 43 67 5 6 44 31 23 15 22 2 70 1 68 3 19 46 69 47 59 29 38 42
[55] 52 51 57 12 8 39 63 28 45 9 11 32 33 37 26 27

En la figura 4 se puede observar que los dendrogramas consensos obtenido por el método *manhattan* en múltiples iteraciones difieren con variaciones más pronunciadas en la estructura de los grupos, altura y orden de las accesiones que el método consensos *euclidean*.

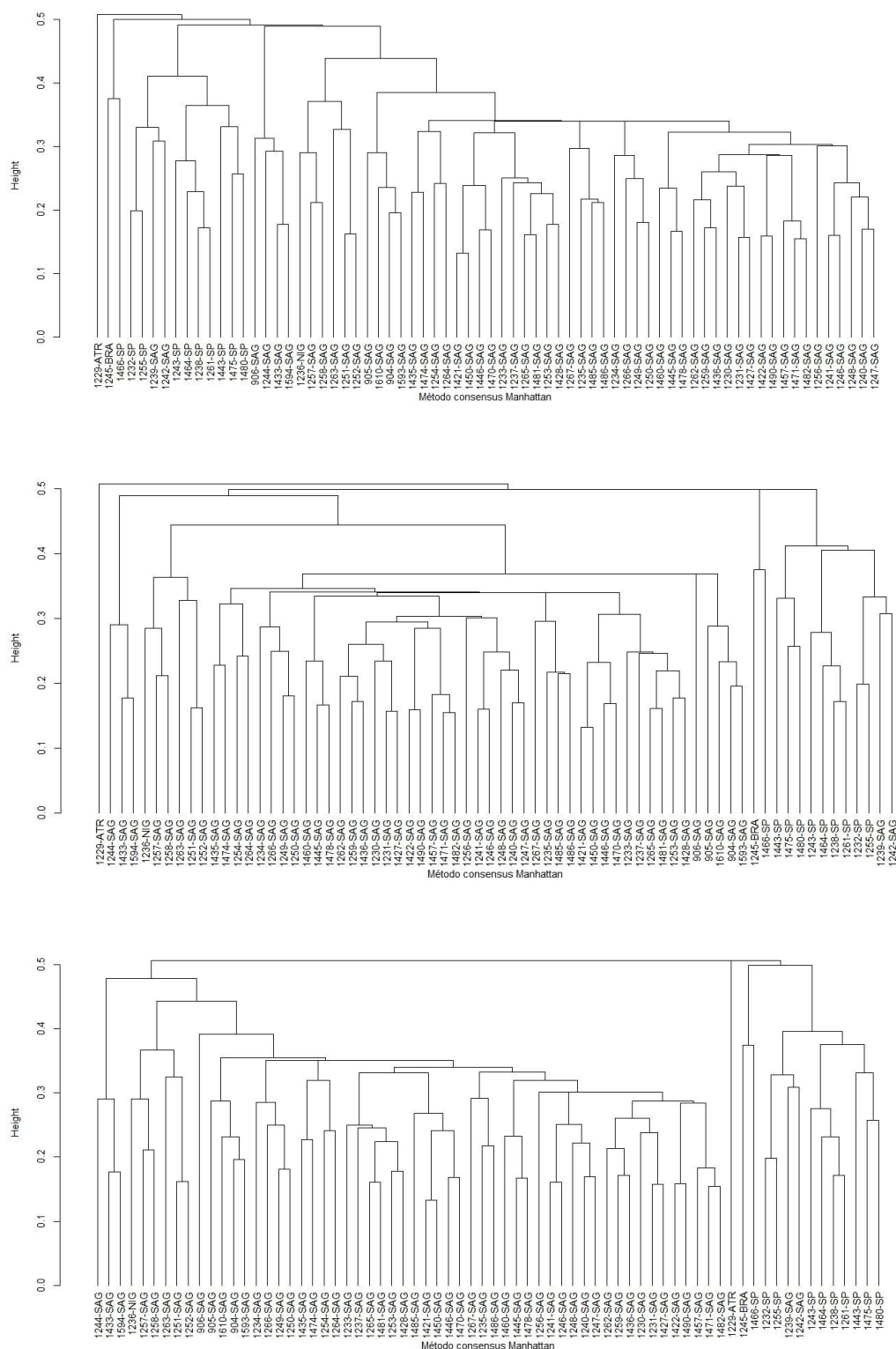


Figura 4. Dendrogramas consensos con el método *manhattan*, en una segunda, tercera y cuarta iteración. Base de datos de la colección del germoplasma de malanga (*Xanthosoma* spp.).

A continuación, se puede observar los cambios en cuanto a la altura y el orden para las iteraciones dos, tres y cuatro que representan los dendrogramas de la figura 4.

Iteración 2:

```
[1] 0.1322754 0.1543929 0.1569062 0.1590017 0.1601743 0.1610782 0.1615001 0.1667392
[9] 0.1680507 0.1697482 0.1716512 0.1718114 0.1771005 0.1774456 0.1806920 0.1826523
[17] 0.1955443 0.1986444 0.2105114 0.2113576 0.2149364 0.2169211 0.2184849 0.2196802
[25] 0.2263922 0.2272478 0.2316183 0.2326994 0.2342189 0.2344268 0.2411345 0.2456487
[33] 0.2475163 0.2484228 0.2491522 0.2569852 0.2603429 0.2785145 0.2842250 0.2850558
[41] 0.2863986 0.2873357 0.2901711 0.2942736 0.2951412 0.3008747 0.3025280 0.3064988
[49] 0.3073434 0.3223530 0.3273753 0.3312808 0.3331231 0.3342434 0.3396239 0.3396252
[57] 0.3406290 0.3461156 0.3628232 0.3680715 0.3688980 0.3743942 0.4047739 0.4111865
[65] 0.4432573 0.4884793 0.4984315 0.4984912 0.5070490
```

Orden:

```
[1] 4 19 46 69 11 32 33 37 26 27 47 59 29 38 9 40 24 25 54 50 61 36 34 48 5 6 44
[28] 43 67 53 58 64 31 16 21 23 15 22 41 10 65 66 42 52 51 57 8 12 39 63 28 45 3 2
[55] 70 1 68 20 56 49 60 62 18 55 13 35 7 30 14 17
```

Iteración 3:

```
[1] 0.1322754 0.1543928 0.1569061 0.1590016 0.1601744 0.1610785 0.1615000 0.1667409
[9] 0.1680505 0.1697482 0.1716511 0.1718114 0.1771005 0.1774456 0.1806920 0.1832199
[17] 0.1955443 0.1986443 0.2113575 0.2133610 0.2171375 0.2220162 0.2235151 0.2272535
[25] 0.2314255 0.2317825 0.2323833 0.2382252 0.2408019 0.2411260 0.2456341 0.2492293
[33] 0.2495605 0.2505283 0.2569851 0.2603425 0.2677103 0.2758355 0.2837595 0.2854284
[41] 0.2874039 0.2874310 0.2904297 0.2904628 0.2913318 0.3007978 0.3008747 0.3085247
[49] 0.3196460 0.3198642 0.3244213 0.3283036 0.3308100 0.3310287 0.3317669 0.3396228
[57] 0.3500848 0.3508409 0.3543550 0.3664032 0.3743950 0.3751792 0.3915550 0.3952873
[65] 0.4430715 0.4779709 0.4989035 0.5056357 0.5060150
```

Orden:

```
[1] 19 46 69 11 32 33 37 26 27 3 2 70 1 68 9 40 24 25 47 59 29 38 8 12 39 63 28
[28] 45 65 42 52 51 57 41 10 66 54 50 61 31 16 21 23 15 22 36 34 48 5 6 44 43 67 53
[55] 58 64 4 20 56 7 30 14 17 18 55 13 35 49 60 62
```

Iteración 4:

```
[1] 0.1322726 0.1543925 0.1569059 0.1590016 0.1601744 0.1610847 0.1615000 0.1667410
[9] 0.1680551 0.1697470 0.1716508 0.1718022 0.1771007 0.1774456 0.1806920 0.1825222
[17] 0.1955423 0.1986442 0.2106226 0.2113577 0.2170906 0.2198950 0.2212539 0.2255193
[25] 0.2272520 0.2310412 0.2317862 0.2326748 0.2350642 0.2379460 0.2411269 0.2486497
[33] 0.2492045 0.2509508 0.2569851 0.2603774 0.2794419 0.2838482 0.2844581 0.2852987
[41] 0.2879456 0.2899067 0.2938322 0.2977591 0.3008746 0.3058066 0.3085247 0.3203517
[49] 0.3227643 0.3249456 0.3256091 0.3256554 0.3260450 0.3320985 0.3359948 0.3363648
[57] 0.3372970 0.3396228 0.3650613 0.3743950 0.3804213 0.3820095 0.3980681 0.4325285
[65] 0.4388027 0.4944091 0.4964234 0.4982356 0.5073551
```

Orden:

```
[1] 4 20 56 7 30 14 17 18 55 13 35 49 60 62 19 46 69 2 70 1 68 47 59 29 38 42 52
[28] 51 57 8 12 39 63 28 45 10 65 41 66 36 34 48 5 6 44 43 67 53 58 64 31 16 21 23
[55] 15 22 54 50 61 9 40 24 25 3 11 32 33 37 26 27
```

Se puede afirmar que el árbol consensos con el método *manhattan* representa mejor la estructura de la base de datos de la colección del germoplasma de malanga (*Xanthosoma* spp.), aunque no logra una estabilidad superior que el obtenido por el método consensos *euclidean*.

CONCLUSIONES

La combinación de agrupamientos para lograr un dendrograma consenso es una técnica efectiva en aras de lograr un mejor rendimiento y calidad de los resultados alcanzados por los algoritmos individuales y compensar posibles errores cometidos en el desempeño de cada uno sobre la clasificación del germoplasma de malanga (*Xanthosoma* spp.).

BIBLIOGRAFÍA

- CUADRAS, C. M. 1981. Métodos de Análisis Multivariante. *Eunibar. D.L.* XII: 642.
- FRANCO, T. L. y HIDALGO R., Eds. 2003. Análisis Estadístico de Datos de Caracterización Morfológica de Recursos Fitogenéticos. *Boletín técnico IPGRI*. Cali, Colombia, Instituto Internacional de Recursos Fitogenéticos (IPGRI).
- GOWER, J.C. 1967. A comparison of some methods of cluster analysis. *Biometrics*, 23:623-628.
- GOWER, J.C. 1971. A general coefficient of similarity and some of its properties. *Biometrics*, 27:857-871.
- HORNIK, K. 2011. clue: Cluster ensembles. R package version 0.3-41. Consultado: 14 de junio de 2016. Disponible en: <http://CRAN.R-project.org/package=clue>. <http://www.r-project.org/>.
- MILIÁN, M J. (2008). Caracterización de la variabilidad de los cultivares de la colección cubana de germoplasma del género *Xanthosoma* (Araceae). Tesis para aspirar al grado de Doctor en Ciencias Biológicas, Ciudad de la Habana, 123 p.
- R DEVELOPMENT CORE TEAM. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing. R Foundation for Statistical Computing. Consultado: 17 de diciembre de 2016. Disponible en: <http://www.r-project.org/>. 17/12/2016.
- ROUSSEEUW, P.J. 1986. A visual display for hierarchical classification. *Data Analysis and Informatics*, 4:743-748.
- SORENSEN, T.A. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analysis of vegetation on Danish commons. *Biologiske Skrifter*, 5:1-34.
- SOKAL, R.R. y ROHLF, F.J. 1962. The comparisons of dendrograms by objective methods. *Taxon*, 11:33-40.
- SNEATH, P.H.A. y SOKAL, R.R. 1973. Numerical taxonomy. The principles and practice of numerical classification. W. H. Freeman and Co, San Francisco, California, USA.
- VEGA, S.P. y RUIZ, J.S. 2010. A survey of clustering ensemble algorithms, World Scientific Publishing Company.
- WARD, J.H. 1963. Hierarchical grouping to optimize an objective function. *J Amer Statist Assoc*, 58:236-244.