



CAPÍTULO 4.

Diversidad genética de la batata en Colombia

▶ JHON ALEXANDER BERDUGO-CELY, JAZMÍN VANESSA PÉREZ-PAZOS Y
ELVIA AMPARO ROSERO ALPALA

El origen de la batata a nivel genético no es del todo claro (Roullier et al., 2013). Hasta la fecha, se considera que esta es una especie autopoliploide, con un posible origen en *I. trifida* como único pariente (Muñoz-Rodríguez et al., 2018). Sin embargo, otra hipótesis es que *I. batatas* es un aloautohexaploide ($2n = 6x = 90$) generado a partir de un posible cruce entre genotipos tetraploides y diploides, seguido de un evento de duplicación del genoma completo (Yang et al., 2017). Independientemente de su origen, el genoma de *I. batatas* es hexaploide y altamente heterocigoto, lo que le confiere una complejidad genética. En los últimos diez años, las bases genómicas con información para esta especie han ido en aumento, con la identificación de transcriptomas ensamblados *de novo* para *I. batatas* y especies relacionadas (Ponniah et al., 2017; Solis et al., 2014). También, se cuenta con recursos como marcadores moleculares tipo microsatélite, el fragmento amplificado de longitud específica (SLAF) y el fragmento amplificado de polimorfismo de longitud (AFLP), entre otros, que se han utilizado para caracterizar la diversidad genética de esta especie en varios países del mundo (Roullier et al., 2013; Su et al., 2017). Con los avances en las tecnologías de secuenciación, los genomas nucleares y de cloroplasto de referencia publicados para esta especie ya se encuentran disponibles (Muñoz-Rodríguez et al., 2018; Yang et al., 2017).

Realizar mejoramiento genético de la batata mediante el fitomejoramiento tradicional es difícil debido a su naturaleza poliploide, su complejidad genética, su alta variabilidad con respecto a la producción de flores y su incompatibilidad (Wadl et al., 2018). Generar recursos genómicos

adicionales y aprovechar los existentes ayudaría a identificar las bases moleculares de la variación fenotípica y avanzar en el diseño de estrategias de mejoramiento asistidas por marcadores eficientes y efectivos (Wadl et al., 2018). El mejoramiento asistido por marcadores permite la evaluación de plantas jóvenes en la etapa de plántula para detectar múltiples rasgos de interés, lo que reduce en gran medida los costos asociados con el crecimiento de las plantas hasta la madurez. Este enfoque es especialmente valioso en la batata, en la que el costo de la evaluación en campo a largo plazo es un factor limitante importante en los esfuerzos de mejoramiento, sumado a que no es factible realizar retrocruzamientos para introducir rasgos. Además, los datos genómicos podrían proporcionar una base para dilucidar las relaciones genéticas entre líneas parentales y potencialmente identificar nuevas fuentes de variación genética asociadas con la tolerancia ambiental, la resistencia a plagas y enfermedades, y otros rasgos de alto valor (Wadl et al., 2018).

Diversidad genética en Colombia

El Banco de Germoplasma Vegetal de Colombia (BGV), administrado por AGROSAVIA, posee una colección de aproximadamente ochenta genotipos de batata, los cuales han sido pobremente caracterizados y utilizados. Además, en los últimos diez años se han realizado nuevas colectas para esta especie en múltiples departamentos de las regiones Andina y Caribe de Colombia, con el fin de conocer la variabilidad genética de esta especie y potencializar su uso. Más de 450 genotipos, que ahora conforman una colección de trabajo (Rosero et al., 2022), han venido siendo caracterizados a nivel fenotípico en las condiciones del Centro de Investigación Turipaná de AGROSAVIA. A partir de este germoplasma, se logró liberar la primera variedad de batata registrada en Colombia: Agrosavia Aurora. Sin embargo, hasta la fecha no existían reportes del análisis del germoplasma de batata de Colombia a nivel genético usando herramientas moleculares.

Procedencia de las accesiones colectadas en Colombia

A nivel genético-molecular, se analizaron 438 genotipos de batata, de los cuales el 95% (415) son de Colombia, mientras que el resto de los genotipos (23) son de origen desconocido (5%). Este germoplasma se encuentra conservado en el C. I. Turipaná de AGROSAVIA, ubicado en el municipio de Cereté, departamento de Córdoba, en Colombia (8° 51' 2.158" N, 75° 49' 10.809" W). Por otro lado, del total de los genotipos analizados, el 14,3% (63) pertenece al BGV, administrado por AGROSAVIA, y el 85,6% (375) pertenece a una colección de trabajo generada a partir de una colecta de genotipos de batata realizada entre 2013 y 2016 en departamentos de las regiones Andina (Boyacá, Caldas, Norte de Santander, Santander, Tolima y las áreas andinas de los departamentos de Nariño y Valle del Cauca) y Caribe (Atlántico, Bolívar, Cesar, Córdoba, La Guajira, Magdalena, Sucre y el norte de Antioquia) (tabla 4.1).

Tabla 4.1. Origen geográfico de las accesiones de batata analizadas en este estudio

País	Número y porcentaje de muestras por país	Región	Número y porcentaje de muestras por región	Departamento	Número y porcentaje de muestras por región
Colombia	415 (95%)	Andina	120 (27%)	Boyacá	6 (1%)
				Caldas	1 (0,22%)
				Norte de Santander	12 (3%)
				Santander	22 (5%)
				Tolima	37 (8%)
				Nariño (área andina)	1 (0,22%)
				Valle del Cauca (área andina)	41 (9%)

País	Número y porcentaje de muestras por país	Región	Número y porcentaje de muestras por región	Departamento	Número y porcentaje de muestras por región
Colombia	415 (95%)	Caribe	191 (44%)	Antioquia (área Caribe)	61 (14%)
				Atlántico	23 (5%)
				Bolívar	24 (5%)
				Cesar	29 (7%)
				Córdoba	29 (7%)
				La Guajira	15 (3%)
				Magdalena	4 (1%)
				Sucre	6 (1%)
Desconocido	23 (5%)	Desconocido	127 (29%)	Desconocido	127 (29%)
Total	438 (100%)	—	—	—	438 (100%)

Fuente: Elaboración propia

Caracterización genética con marcadores moleculares

Para el análisis genético-molecular, por cada genotipo se sembraron cinco plantas en condiciones de vivero en el C. I. Turipaná. Dos meses después del establecimiento de los materiales, se colectó tejido foliar de una sola planta, a partir del cual se obtuvo ADN. Todas las muestras de ADN fueron sometidas a una doble digestión química utilizando las enzimas de restricción CviAII (a 25 °C) y Tsel (a 65 °C). A partir de cada muestra de ADN digerido, se construyeron bibliotecas genómicas, las cuales se agruparon en grupos de 96 muestras (*pool*), cada una marcada con una secuencia única de identificación. Cada *pool* fue secuenciado de manera independiente.

La identificación de variantes tipo SNP se realizó a través del programa Genome Analysis Toolkit (GATK), versión 4.2.6 (Khalfan, 2020), utilizando secuencias previamente alineadas al genoma de referencia del cultivar hexaploide Taizhong6 de *I. batatas* (Yang et al., 2017). El filtrado final de

los marcadores se realizó en el programa VCFtools (Danecek et al., 2011) utilizando los siguientes parámetros de filtrado: frecuencia mínima de alelos (MAF) del 2 %, porcentaje máximo de datos perdidos por marcador del 30 % y profundidad media de secuenciación (DS) de 2. Por otro lado, se excluyeron del análisis las muestras con datos perdidos superiores al 30 %, las variantes SNP multialélicas y los tipos InDels (inserciones y deleciones).

Con la matriz de datos obtenida, se implementaron diferentes análisis multivariados, como el análisis de componentes principales (ACP) y el análisis de conglomerados (AC), para determinar la estructura genética de esta colección. Una vez identificado el número de subpoblaciones, se estimó la diversidad y estructura genética de la población, utilizando estadísticas como la heterocigosidad (H_o) y un análisis de varianza molecular (AMOVA). En todos los casos, se buscó una posible asociación entre el número de subpoblaciones identificado y los datos de pasaporte disponibles. Finalmente, usando el programa CoreHunter3 (De Beukelaer et al., 2018), se propuso una colección núcleo del 23,5 %, seleccionando muestras que representaran el espectro de la diversidad genética presente en la totalidad de la población analizada.

Estructura poblacional de la colección de batata

Una vez realizado el análisis genético de las 438 muestras de batata, se identificaron un total de 26.152 variantes tipo SNP usando los parámetros de filtrado seleccionados. Del total de las muestras analizadas, 33 presentaron un porcentaje de datos perdidos superior al 30 % y fueron excluidas del análisis. Por lo anterior, en los resultados que se presentarán sobre la caracterización molecular, se incluyen 405 muestras con 37.014 SNP distribuidos en los 15 cromosomas del genoma de batata usado como referencia. La estructura genética de las 405 accesiones analizadas se indagó desde diferentes niveles, usando los dos primeros componentes principales del PCA, que explicaron el 20,3 % y el 9,4 % de la variabilidad genética, donde el 100 % de la variabilidad se explicó en 425

Los seis grupos genéticos identificados con las variantes tipo SNP presentan características diferenciables asociadas con la estructura genética detectada. A continuación, se realiza una breve descripción de cada agrupamiento para las variables de pasaporte más contrastantes o con mayor representatividad.

Genetic_Cluster_K6_1: en este grupo se ubican 35 accesiones, con un promedio de diversidad genética (H_o) de 0,07 ($\pm 0,05$) y altos valores de endogamia (F_i : $0,41 \pm 0,37$) (tabla 4.2). Estos materiales son principalmente de la región Caribe (77%), de los cuales se resaltan los departamentos de Antioquia (zona Caribe) (49 %) y Córdoba (23 %). A nivel geográfico, aquí se encuentran materiales colectados desde los 8 hasta los 155 m.s.n.m., y en cuanto a tipo de cultivo, se ubicó el 34 % de los materiales nativos, 23 % de materiales de agricultores y el 20 % de materiales catalogados como malezas.

Tabla 4.2. Estadísticos de diversidad genética determinados para los seis agrupamientos detectados en la colección de batata

Agrupamiento	Número de muestras	H_o (DE)	F_i (DE)
Genetic_Cluster_K6_1	35	0,07 ($\pm 0,05$)	0,41 ($\pm 0,37$)
Genetic_Cluster_K6_2	36	0,06 ($\pm 0,02$)	0,54 ($\pm 0,19$)
Genetic_Cluster_K6_3	32	0,05 ($\pm 0,02$)	0,63 ($\pm 0,15$)
Genetic_Cluster_K6_4	19	0,04 ($\pm 0,02$)	0,65 ($\pm 0,18$)
Genetic_Cluster_K6_5	73	0,01 ($\pm 0,01$)	0,89 ($\pm 0,06$)
Genetic_Cluster_K6_6	56	0,05 ($\pm 0,02$)	0,64 ($\pm 0,18$)
<i>Admixtures</i>	154	0,04 ($\pm 0,02$)	0,68 ($\pm 0,18$)

H_o : heterocigosidad observada; F_i : coeficiente de endogamia.

Fuente: Elaboración propia

Genetic_Cluster_K6_2: en este grupo se ubican 36 accesiones, con un promedio de diversidad genética (H_o) de 0,06 ($\pm 0,02$) y altos valores de endogamia (F_i : $0,54 \pm 0,19$) (tabla 4.2). Estos materiales son principalmente de la región Andina (72 %), de los departamentos de Tolima (42 %) y Santander (22 %). A nivel geográfico, aquí se ubican principalmente los materiales muestreados entre los 672 y los 2.209 m.s.n.m., y en cuanto a tipo de cultivo, estos materiales fueron catalogados como silvestres (50%).

Genetic_Cluster_K6_3: en este grupo se ubican 32 accesiones, con un promedio de diversidad genética (H_o) de 0,05 ($\pm 0,02$) y altos valores de endogamia (F_i : $0,63 \pm 0,15$) (tabla 4.2). Estos materiales fueron colectados principalmente en la región Andina (63%), en el departamento del Valle del Cauca (59%); la mayoría de ellos fueron colectados entre los 1.000 y los 1.401 m s. n. m. En cuanto al tipo de cultivo, estos materiales son líneas de mejoramiento genético (66%).

Genetic_Cluster_K6_4: en este grupo se ubican 19 accesiones, con un promedio de diversidad genética (H_o) de 0,04 ($\pm 0,02$) y altos valores de endogamia (F_i : $0,65 \pm 0,18$) (tabla 4.2). La mayoría de estos materiales provienen de la región Andina (63%), con datos de pasaporte del departamento de Norte de Santander (58%), y fueron colectados a altitudes entre 1.000 y 1.216 m s. n. m. Según el tipo de cultivo, estos materiales son en su mayoría nativos (68%).

Genetic_Cluster_K6_5: en este grupo se ubican 73 accesiones, con un promedio de diversidad genética (H_o) de 0,01 ($\pm 0,01$) y altos valores de endogamia (F_i : $0,89 \pm 0,06$) (tabla 4.2). La mayoría de estos materiales provienen de la región Caribe (64%), con datos de pasaporte de los departamentos de Antioquia (33%) y Córdoba (16%). También se identificaron algunas muestras de la región Andina (21%), específicamente del departamento de Tolima (15%). Los materiales de este agrupamiento fueron colectados entre los 6 y los 1.118 m s. n. m. en la región Caribe, y entre los 431 y los 1.647 m s. n. m. en la región Andina. En su mayoría, estos materiales están catalogados como silvestres (30%) y nativos (30%).

Genetic_Cluster_K6_6: en este grupo se ubican 56 accesiones, con un promedio de diversidad genética (H_o) de 0,05 ($\pm 0,02$) y altos valores de endogamia (F_i : $0,64 \pm 0,18$) (tabla 4.2). La mayoría de estos materiales provienen de la región Caribe (98%), con datos de pasaporte de los departamentos de Atlántico (27%), Bolívar (25%) y La Guajira (16%). Los materiales de este agrupamiento fueron colectados entre los 6 y los 526 m s. n. m., y en su mayoría están catalogados como materiales nativos (63%).

Comparación entre las muestras conservadas en el banco de germoplasma y las de la colección de trabajo

A partir de los análisis de estructura genética obtenidos, se realizó una comparación entre las muestras conservadas en el banco de germoplasma y las de la colección de trabajo del C. I. Turipaná (figura 4.2). Según el número de muestras por cada grupo genético, se observa que la colección de batata del banco presenta, en su mayoría, representatividad de materiales de la región Andina (grupos genéticos 2 y 5). Sin embargo, el banco de germoplasma presenta vacíos para los grupos genéticos 1 y 6 identificados en este análisis; lo anterior sugiere que en la colección de trabajo hay materiales con diversidad genética que no se encuentra actualmente conservada en el banco de germoplasma. A partir de estos resultados, se incrementa la necesidad de obtener una colección núcleo que capture la diversidad genética de la colección aquí analizada, la cual puede ser postulada para su conservación en el banco de germoplasma (tabla 4.3).

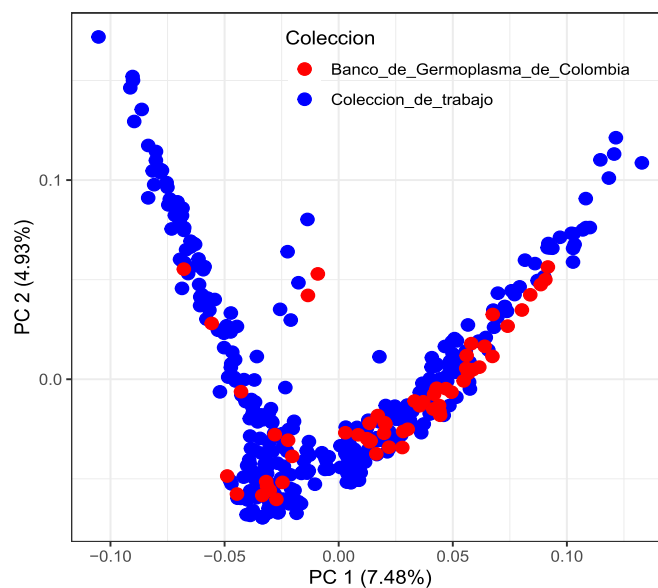


Figura 4.2.

PCA: distribución de las muestras de batata según el tipo de colección a la que pertenecen.

Fuente: Elaboración propia

Tabla 4.3. Representatividad de las muestras del banco de germoplasma y de la colección de trabajo en cada grupo genético detectado

Agrupamiento	Origen	Número de muestras	Banco de germoplasma	Colección de trabajo
Genetic_Cluster_K6_1	Región Caribe	35	4	31
Genetic_Cluster_K6_2	Región Andina	36	10	26
Genetic_Cluster_K6_3	Región Andina	32	3	29
Genetic_Cluster_K6_4	Región Andina	19	5	14
Genetic_Cluster_K6_5	Regiones Caribe y Andina	73	10	63
Genetic_Cluster_K6_6	Regiones Caribe	56	1	55
<i>Admixtures</i>	Regiones Caribe y Andina	154	25	129

Fuente: Elaboración propia

Colección núcleo de batata determinada a partir de variantes tipo SNP

Una vez establecidas la diversidad y la estructura genética de las 405 accesiones de batata, se buscó seleccionar 100 accesiones (que corresponden al 23,5 % del total de la colección analizada), pues este es el número mínimo de muestras aceptadas para la implementación de estudios de mapeo por asociación. Las accesiones postuladas para conformar la colección núcleo a través de esta estrategia se pueden visualizar en el PCA de la figura 4.3; estas accesiones representan la diversidad genética de la colección de batata, incluyendo accesiones representativas de los cinco grupos genéticos establecidos a través del análisis de estructura genética y el espectro de las variables de pasaporte asociadas con la estructura poblacional.

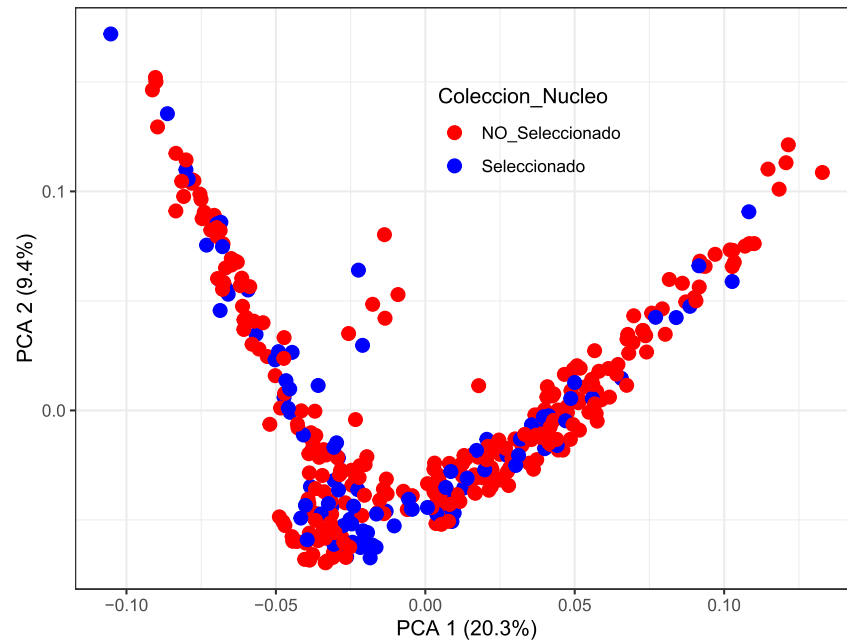


Figura 4.3.

PCA: distribución de la variabilidad genética de la colección de batata y de las muestras seleccionadas como colección núcleo.

Fuente: Elaboración propia

En general, la diversidad colectada y determinada mediante análisis genéticos demuestra la gran riqueza que tiene Colombia en cuanto a la batata, siendo este un insumo importante para generar progreso en el desarrollo de nuevas variedades, lo cual demuestra que la oferta ambiental en diferentes regiones favorece el cultivo de batata. Iniciativas relacionadas con el uso de esta gran diversidad deben ser emprendidas con el fin de aprovechar la riqueza genética de esta especie.

Los resultados aquí presentados forman parte del artículo *Genetic Diversity and Population Structure of Colombian Sweet Potato Genotypes Reveal Possible Adaptations to Specific Environmental Conditions*, publicado en la revista *Crop Science* en 2025, con el número e70091, por Berdugo-Cely et al. (2025).

Referencias

- Berdugo-Cely, J. A., Pérez-Pazos, J. V., & Rosero, A. (2025). Genetic diversity and population structure of Colombian sweet potato genotypes reveal possible adaptations to specific environmental conditions. *Crop Science*, *65*, artículo e70091. <https://doi.org/10.1002/csc2.70091>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2.156-2.158. <https://doi.org/10.1093/BIOINFORMATICS/BTR330>
- De Beukelaer, H., Davenport, G. F., & Fack, V. (2018). Core Hunter 3: Flexible core subset selection. *BMC Bioinformatics*, *19*(1), artículo 203. <https://doi.org/10.1186/s12859-018-2209-z>
- Khalfan, M. (2020). Variant Calling Pipeline using GATK4. *Genomics Core at NYU CGSB*. <https://gencore.bio.nyu.edu/variant-calling-pipeline-gatk4/>
- Muñoz-Rodríguez, P., Carruthers, T., Wood, J. R. I., Williams, B. R. M., Weitemier, K., Kronmiller, B., Ellis, D., Anglin, N. L., Longway, L., Harris, S. A., Rausher, M. D., Kelly, S., Liston, A., & Scotland, R. W. (2018). Reconciling conflicting phylogenies in the origin of sweet potato and dispersal to Polynesia. *Current Biology*, *28*(8), 1.246-1.256. <https://doi.org/10.1016/J.CUB.2018.03.020>
- Ponniah, S. K., Thimmapuram, J., Bhide, K., Kalavacharla, V., & Manoharan, M. (2017). Comparative analysis of the root transcriptomes of cultivated sweetpotato (*Ipomoea batatas* [L.] Lam) and its wild ancestor (*Ipomoea trifida* [Kunth] G. Don). *BMC Plant Biology*, *17*(1), artículo 9. <https://doi.org/10.1186/s12870-016-0950-x>
- Rosero, A., Rodríguez, E., Aguilera-Arango, G., Rosero, M.-G., Granda, L., Pastrana, I., Martínez, R., Perez, J.-L., Espitia, L., Gomez, E., Rodríguez, T., & Sieber, S. (2022). Assessment of the current state of *in situ* conservation and use of sweet potato (*Ipomoea batatas* L.) in Colombia. *Culture, Agriculture, Food and Environment*, *44*(1), 76-89. <https://doi.org/10.1111/CUAG.12293>
- Roullier, C., Duputié, A., Wennekes, P., Benoit, L., Fernández Bringas, V. M., Rossel, G., Tay, D., McKey, D., & Lebot, V. (2013). Disentangling the origins of cultivated sweet potato (*Ipomoea batatas* (L.) Lam.). *PLOS ONE*, *8*(5), artículo e62707. <https://doi.org/10.1371/journal.pone.0062707>
- Solis, J., Villordon, A., Baisakh, N., LaBonte, D., & Firon, N. (2014). Effect of drought on storage root development and gene expression profile of sweetpotato under greenhouse and field conditions. *Journal of the American Society for Horticultural Science*, *139*(3), 317-324. <https://doi.org/10.21273/JASHS.139.3.317>
- Su, W., Wang, L., Lei, J., Chai, S., Liu, Y., Yang, Y., Yang, X., & Jiao, C. (2017). Genome-wide assessment of population structure and genetic diversity and development

of a core germplasm set for sweet potato based on specific length amplified fragment (SLAF) sequencing. *PLOS ONE*, 12(2), artículo e0172066-. <https://doi.org/10.1371/journal.pone.0172066>

Wadl, P. A., Olukolu, B. A., Branham, S. E., Jarret, R. L., Yencho, G. C., & Jackson, D. M. (2018). Genetic diversity and population structure of the USDA sweetpotato (*Ipomoea batatas*) germplasm collections using GBSpoly. *Frontiers in Plant Science*, 9. <https://doi.org/10.3389/fpls.2018.01166>

Yang, J., Moeinzadeh, M.-H., Kuhl, H., Helmuth, J., Xiao, P., Haas, S., Liu, G., Zheng, J., Sun, Z., Fan, W., Deng, G., Wang, H., Hu, F., Zhao, S., Fernie, A. R., Boerno, S., Timmermann, B., Zhang, P., & Vingron, M. (2017). Haplotype-resolved sweet potato genome traces back its hexaploidization history. *Nature Plants*, 3(9), 696-703. <https://doi.org/10.1038/s41477-017-0002-z>