

# Caracterización de germoplasma vegetal: la piedra angular en el estudio de los recursos fitogenéticos

Characterization of plant germplasm: the cornerstone in the study of plant genetic resources

Carlos Alberto Núñez-Colín<sup>1\*</sup>, Diana Escobedo-López<sup>1</sup>

---

## RESUMEN

La caracterización de germoplasma vegetal es una herramienta ideal para estudiar los recursos fitogenéticos, especialmente en un país con la megadiversidad con la que cuenta México, sobre todo en su flora. Sin embargo, existe desconocimiento sobre cómo hacer estos estudios. En esta revisión se hace un recorrido por el significado de la caracterización de germoplasma vegetal, se presentan los objetivos y una breve descripción de los análisis estadísticos más utilizados para este fin.

## PALABRAS CLAVE

Estadística multivariada, recursos fitogenéticos, germoplasma, México

## ABSTRACT

The characterization of plant germplasm is a fundamental tool to make studies about plant genetic resources. Mainly in a mega-diverse country as Mexico, especially because of its flora. However, there is lack of knowledge on how to make in a good way these studies. In this review, we make a short journey about the meaning of characterization of plant germplasm, its objectives and a brief description of the most popular statistical analyses used for this purpose.

## KEYWORDS

Multivariate statistic, plant genetic resources, germplasm, Mexico

---

<sup>1</sup> Campo Experimental Bajío; Centro de Investigación Regional Centro; Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias. Km. 6.5 Carretera Celaya-San Miguel de Allende. 38110 Celaya, Guanajuato, México.

\* Autor para correspondencia. Correo electrónico: lit007a@gmail.com

## INTRODUCCIÓN

México es considerado un país megadiverso, especialmente en su flora (Ramamoorthy *et al.*, 1993) y en el continente americano es quizá el territorio con más endemismos registrados (Villaseñor, 2003). Nuestro país forma parte de un tratado internacional para conservar los recursos fitogenéticos para la alimentación y la agricultura, que suscribió en 2001 la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO, por sus siglas en inglés). Gracias a este documento, muchos países latinoamericanos han implementado programas nacionales de conservación de sus recursos fitogenéticos, especialmente el nuestro por ser considerado como centro de origen de diversos cultivos (Vavilov, 1992).

Una de las principales dificultades que se presentan en este tema es que muchos de estos recursos no están totalmente domesticados, por lo que se desconoce su diversidad y potencial agroalimentario. En este sentido, las técnicas estadísticas de caracterización de germoplasma son indispensables para conocer estos recursos fitogenéticos. Por lo anterior, la presente revisión dará un panorama general de qué es la caracterización de germoplasma vegetal.

## MARCO TEÓRICO

### Justificación del estudio de los recursos genéticos

Desde el inicio de la agricultura, en el neolítico, las distintas culturas y razas que han poblado la Tierra han domesticado especies vegetales para su aprovechamiento y han seleccionado, a lo largo de las generaciones, aquellos caracteres que mejor se adaptan a sus necesidades. Paralelamente a este proceso de domesticación y diversificación, ha existido siempre una preocupación por la conservación de los recursos biológicos disponibles (Iriondo-Alegría, 2001).

La conservación y evaluación de la diversidad biológica no sólo aporta beneficios en el área de la producción alimentaria, sino que puede contribuir sustancialmente a otras áreas de la actividad humana, como la medicina e industria. Por ello, el estudio de los recursos genéticos es rentable para la sociedad y la industria (Iriondo-Alegría, 2001).

### Conceptos básicos sobre caracterización

Para obtener una definición de caracterización relacionada con la agricultura, primero es in-

dispensable conocer el término común, como lo ofrece el diccionario de la Real Academia Española (2001): "Determinar los atributos peculiares de alguien o algo, de modo que claramente se distinga de los demás". Partiendo de este concepto, caracterizar un recurso fitogenético es determinar los atributos peculiares de dicho recurso, de modo que podamos distinguirlo claramente de cualquier otro.

Y entonces surge otra interrogante: ¿qué es un recurso fitogenético? Como sucede con otros conceptos "nuevos" no hay una única definición válida, pero normalmente suele hablarse de individuos vegetales con características específicas que pueden utilizarse en un programa de mejoramiento (Hidalgo, 2003). No solamente los recursos vegetales que le sirven al hombre son recursos fitogenéticos, no en estricto sentido, aunque todos los recursos vegetales tengan un posible uso antropocéntrico, aun aquellos a los que actualmente no se les conozca ninguno.

De ahí que un recurso fitogenético es todo aquel individuo o grupo de individuos vegetales que presentan características específicas que permiten distinguirlos del resto de los recursos genéticos vegetales, ya sea un genotipo, una accesión de germoplasma, una variedad cultivada, una población natural, una especie (en el sentido taxonómico) o un género, entre otros (Núñez-Colín y Escobedo-López, 2011).

### Objetivos de los estudios de caracterización de recursos fitogenéticos

Esta herramienta puede tener una amplia gama de objetivos, aunque hay quien realiza caracterizaciones sólo por conocer todos los atributos de todos los individuos que se tienen en una colección vegetal (banco de germoplasma, colección de trabajo, programa de mejoramiento genético) y, aunque esto sea un avance en la materia, debe haber objetivos más claros al desempeñar esta tarea para poder sacar provecho de la información y dar una interpretación clara y precisa al recurso fitogenético que se pretenda caracterizar (González-Andrés, 2001; Hidalgo, 2003; Núñez-Colín y Escobedo-López, 2011). En este sentido, los principales objetivos de la caracterización vegetal son los siguientes:

**Taxonomía y sistemática.** Se trata de clarificar la identidad taxonómica de un individuo o grupo de individuos a través de sus características, con el uso de técnicas de taxonomía numérica (que están dentro de la estadística multivariada), para tener clasificaciones taxonómicas más claras y a la vez hacer una

identificación más precisa del individuo o del grupo con el que se está trabajando.

**Variabilidad y diversidad genética.** Cuando se tiene claridad sobre la taxonomía del recurso, estas técnicas de caracterización pueden servir para identificar y medir qué tan variable (si se trata de una especie) o qué tan diverso (si se trata de género o un rango taxonómico superior a especie) es dicho recurso en comparación con otros. Estas técnicas se utilizan sin preclasificar a los individuos por región geográfica o alguna otra variable clasificatoria, hasta cierto punto, artificial.

**Variabilidad y diversidad ecológica.** Al igual que la caracterización anterior, ésta permite determinar qué tan variable (si se trata de una especie) o qué tan diverso (si se trata de un género o un rango taxonómico superior a especie) es dicho recurso, pero en este caso sí se pre-clasifica por hábitat o región geográfica para conocer las diferencias que éste presenta, más que la clasificación o el parecido entre individuos.

**Estructura genética.** Esta técnica es fácil de explicar partiendo de que todas las características tienen un origen genético, por lo que al conocer las peculiaridades de cada individuo (o población) se puede hacer (mediante las técnicas y características apropiadas) una estructura de la población o de los individuos con base en su parecido y predecir, mediante genética cuantitativa y de poblaciones, sus características genéticas básicas.

**Gestión de un programa de mejoramiento genético.** Es fundamental para conocer los atributos particulares de cada individuo del programa (padres y descendencia), para planear qué características se van a mejorar, qué individuos presentan dicha característica y cuáles presentan otras importantes, así como para conocer las diferencias entre ellos y establecer cuáles serían las cruza más promisorias o si es mejor realizar hibridaciones o selecciones.

**Gestión de bancos de germoplasma.** En los bancos de germoplasma la caracterización es el primer paso para conocer, justamente, el germoplasma, así como para identificar duplicados, medir la variabilidad o diversidad de lo que se está guardando y si se necesitan más accesiones o no. Con una buena caracterización se pueden iniciar colecciones nucleares o ampliar una colección de acuerdo con las necesidades fundamentales de cada banco. Incluso, es posible seleccionar los individuos representativos para hacer una colección de trabajo para un programa de mejoramiento genético y ahorrar recursos con esta planeación.

**Filogenia.** Con los resultados previos o parciales de los objetivos anteriores se puede buscar también la filogenia del recurso fitogenético, es decir, su árbol evolutivo para contestar las interrogantes de su origen o parentesco con respecto al de los otros individuos emparentados, hacer una hipótesis evolutiva del mismo y predecir su historia evolutiva.

### Unidad básica de caracterización

Como su nombre lo indica, este término se refiere a la unidad básica de lo que se va a caracterizar. Los taxónomos la llaman unidad taxonómica operativa, pero, dado que no todos los trabajos son de taxonomía, esto no puede aplicarse a todas las investigaciones de caracterización. La unidad básica de caracterización (UBC) puede estar constituida por un único individuo si éste presenta características peculiares, por ejemplo, si se obtiene de un programa de mejoramiento o de una única colecta o accesión de germoplasma. También puede tratarse de un conjunto de individuos que presentan ciertas características en común, por ejemplo, individuos de una población silvestre (para caracterizar dicha población y compararla con otras), una variedad cultivada (puede ser clonal o no) o incluso una especie taxonómica completa (si ésta es la que se va a caracterizar y comparar con otras). La UBC tiene que determinarse de acuerdo con el objetivo específico de la caracterización y con el interés del investigador (González-Andrés, 2001; Núñez-Colín y Escobedo-López, 2011).

Las características homólogas son las que se deben medir en cada UBC, es decir, las que tengan la misma utilidad y sean evolutivamente homogéneas. Estos caracteres fueron los que Darwin tomó en cuenta para su teoría de la evolución de las especies. Se deben considerar estructuras o caracteres con modificaciones que puedan ser interpretados de manera evolutiva, como la estructura ósea de las extremidades superiores de los vertebrados, que en peces, reptiles, anfibios, aves y mamíferos está presente pero siempre con modificaciones. Caso contrario, esto es, una característica no homóloga, es la presencia de cola, porque la rana y el hombre no la presentan, pero muchos mamíferos sí y éstos están más relacionados con el hombre que la rana (González-Andrés, 2001; Hidalgo, 2003; Núñez-Colín y Escobedo-López, 2011).

Sobre este aspecto el uso de marcadores moleculares (principalmente de ADN) aleatorios (RAPD, CAP, entre otros) puede no dar caracteres homólogos si se analizan individuos evolutivamente muy alejados, por

lo que hay que tener cuidado con las características y con el tipo de técnicas estadísticas con las que se están analizando para evitar cometer errores en la caracterización (Núñez-Colín y Valadez-Moctezuma, 2010; Núñez-Colín y Escobedo-López, 2011).

### Metodologías usadas en estudios de caracterización de germoplasma vegetal

La mayoría de técnicas utilizadas en la caracterización de recursos genéticos vegetales se ubica en la estadística multivariada (también llamada multivariable), técnica por la que se evalúa más de una variable respuesta a la vez (Hair *et al.*, 2001; Johnson, 1998).

Este tipo de metodologías son utilizadas con distintos objetivos y cada una requiere ciertos supuestos estadísticos o modificaciones para poder emplearse.

### Objetivos de los métodos multivariados

Los principales objetivos de los métodos multivariados son:

**Hacer clasificaciones.** Es el principal objetivo de la caracterización de recursos genéticos vegetales (aunque no el único). Este tipo de técnicas permite al investigador obtener clasificaciones de acuerdo con las similitudes o disimilitudes de todas las características medidas entre sus UBC, para formar grupos más o menos homogéneos. Lo anterior permite corroborar la identificación taxonómica, ver la variabilidad (o diversidad) que presentan las UBC, para hacer hipótesis de la estructura genética e incluso una hipótesis filogenética (González-Andrés, 2001; Núñez-Colín y Escobedo-López, 2011).

Objetivos de este tipo no sólo se logran con métodos de agrupamiento (análisis clúster), sino también con otras metodologías, como los análisis factoriales (análisis en componentes principales, análisis de correspondencias simples y múltiples, análisis en coordenadas principales o análisis por factores).

**Selección de variables.** Algunas metodologías multivariadas permiten discriminar o seleccionar variables para evitar repetir información con más de una variable (por ejemplo, utilizar variables con correlaciones lógicas como el largo del fruto con su peso, que en su mayoría están enfocadas al tamaño y fuertemente correlacionadas matemáticamente) o para eliminar variables (características) que no aporten a la caracterización ya que todas las UBC son similares (González-Andrés, 2001; Núñez-Colín

y Barrientos-Priego, 2006; Núñez-Colín y Valadez-Moctezuma, 2010).

Estos pasos son importantes en ensayos preliminares porque de este modo, cuando se incorpore un número mayor de UBC, se evaluarán sólo aquellas variables que hayan sido importantes o discriminantes en los ensayos preliminares y se evita trabajo innecesario.

**Estructuración de la varianza.** Muchos de los métodos multivariados se basan en la estructura de la varianza de todas y cada una de las variables evaluadas, por lo que con esta técnica se puede predecir cuáles variables aportan más a la variación total de las UBC y se puede saber entonces cuáles son las variables y las UBC más variantes dentro de la caracterización (González-Andrés, 2001; Johnson, 1998; Núñez-Colín y Barrientos-Priego, 2006).

**Comparación de grupos.** Otra de las funciones de los métodos multivariados es la comparación de grupos. Éstos pueden o no ser resultado de técnicas de agrupación debido a que muchas veces el investigador tiene sus muestras preclasificadas de acuerdo con múltiples razones. El fin último es tener grupos preestablecidos y compararlos tanto en su variación con otros (variación intergrupos) como la variación dentro del mismo en comparación con los demás (variación intragrupo), para detectar las diferencias que presentan y poder caracterizar al grupo (Núñez-Colín y Barrientos-Priego, 2006; Núñez-Colín *et al.*, 2008).

### Principales análisis estadísticos multivariados

Los principales análisis estadísticos que se utilizan en este tópico son:

**Análisis clúster.** Éste es un método matemático principalmente utilizado para la formación de grupos de UBC con características similares a partir de las similitudes o disimilitudes que se presentan entre pares de estas unidades en características evaluadas (Johnson, 1998; Núñez-Colín y Escobedo-López, 2011). Este tipo de análisis está compuesto por dos métodos interrelacionados e igualmente importantes.

El primero es el cálculo de los índices de similitud o de disimilitud entre pares de UBC; existe un sinnúmero de investigaciones publicadas al respecto. No obstante, estos índices deben ser aplicados de acuerdo con la naturaleza de los datos y el objetivo de la caracterización (Núñez-Colín y Escobedo-López, 2011).

El segundo es la aplicación del método de aglomeración adecuado, también se han desarrollado valiosos trabajos de este tipo. A partir de los índices de

similitud o disimilitud se pueden generar las gráficas de árbol o dendrogramas (representaciones gráficas) para que el investigador pueda ver, de manera esquemática, el parecido que presentan los grupos de UBC. El método de aglomeración adecuado se define principalmente por el objetivo de la caracterización (Núñez-Colín y Escobedo-López, 2011).

Ambos métodos, englobados en el análisis clúster, a pesar de que presentan sólidas bases matemáticas, no son tan restrictivos al respecto de sus bases estadísticas (Johnson, 1998). Por lo anterior, el principal problema que presenta este análisis es la elección del índice de similitud o disimilitud y del método de aglomeración más apropiados. La elección de la mejor combinación de los métodos del análisis clúster depende principalmente de la naturaleza de los datos y el objetivo de la caracterización (Núñez-Colín y Escobedo-López, 2011).

**Análisis factoriales.** Éstos son utilizados para condensar o reducir información contenida en una serie de variables originales, en una serie más pequeña de dimensiones compuestas o valores teóricos nuevos (factores) con una pérdida mínima de información (Hair *et al.*, 2001; Núñez-Colín y Valadez-Moctezuma, 2010).

Los más utilizados son el análisis en componentes principales (ACP), el de coordenadas principales (Acop) y el análisis factorial de correspondencias simple (AFC), que se utilizan de acuerdo con el tipo de datos que se van a evaluar (Núñez-Colín y Valadez-Moctezuma, 2010).

El ACP es un procedimiento matemático que transforma un conjunto de variables de respuesta correlacionadas a partir de la matriz de correlación de las variables originales, o en su defecto, la matriz de varianzas-covarianzas, en un conjunto menor de variables no correlacionadas llamadas componentes principales, que resultan tantos como variables evaluadas. Estos componentes acumulan una parte de la varianza total y tienen mayor importancia y mayor varianza acumulada entre mayor sea su valor propio: el primer componente es el que tenga más varianza acumulada, y así subsecuentemente (Johnson, 1998; Núñez-Colín y Valadez-Moctezuma, 2010). Este procedimiento es utilizado para datos cuantitativos y mixtos, mayormente cuantitativos; se deben evitar datos binomiales.

Por otra parte, el Acop es una modalidad de los análisis factoriales que se usa exclusivamente para datos binomiales. Se trata de una transformación de un índice de similitud mediante la distancia de Gower en una matriz cuadrada, que hace a su vez de matriz

de varianzas y covarianzas, se obtienen sus valores y sus vectores propios (eigenvalues y eigenvectors) y se usa únicamente para tener una representación bi o tridimensional de las UBC en un plano reducido para datos binomiales, principalmente de huellas genómicas de marcadores dominantes (Hair *et al.*, 2001; Johnson, 1998; Núñez-Colín y Valadez-Moctezuma, 2010).

El AFC presenta el mismo principio del ACP, pero con una transformación de datos con base en distribución  $\chi^2$ . Está desarrollado para datos multinomiales (los binomiales son un caso particular), por lo que esta técnica es la más recomendable para matrices binomiales, ya que se cometen menos errores utilizando el AFC con relación al ACP. La interpretación del AFC es exactamente igual a cualquier otro análisis factorial (Johnson, 1998; Núñez-Colín y Valadez-Moctezuma, 2010).

Todas estas técnicas multivariadas permiten representar los genotipos en gráficos con dimensionalidad reducida y ver el parecido de ellos de acuerdo con su cercanía en un plano dimensional o tridimensional. Asimismo se pueden detectar las posibles características asociadas a cada genotipo, esto es muy útil para obtener características específicas que puedan servir para la identificación rápida de los genotipos que se comparan (Núñez-Colín y Valadez-Moctezuma, 2010).

**Análisis discriminante canónico.** Tiene fuertes bases estadísticas y es solamente utilizable en datos cuantitativos y cuando se tienen grupos preestablecidos. Estos datos pueden ser resultado de un análisis clúster, o bien, estar predefinidos por el investigador (Hair *et al.*, 2001; Johnson, 1998).

Este análisis se utiliza principalmente para la comparación de grupos y se puede decir que es una combinación de un ACP con un análisis de varianza. Al igual que el ACP, se basa en la matriz de correlación, pero en este caso de la correlación canónica entre las matrices de varianzas-covarianzas de los grupos preestablecidos; la interpretación es muy similar a la de los análisis factoriales. No obstante, las variables se denominan raíces o canónicas y muestran las principales variables para diferenciar los grupos que se están comparando. Además se generan dos pruebas más, la distancia Mahalanobis, que permite comparar si dos de los grupos son o no iguales, así como una prueba de resubstitución o a posteriori, que muestra si los elementos de un grupo estuvieron clasificados de acuerdo con sus características (Hair *et al.*, 2001; Johnson, 1998; Núñez-Colín *et al.*, 2008).

## CONCLUSIONES

La caracterización de germoplasma vegetal es una prioridad en investigación en México, un país megadiverso (Ramamoorthy *et al.*, 1993). Sin embargo, existen pocos expertos en el tema, la mayor parte de investigadores que han adoptado esta línea de investigación no ha llevado un curso formal de estadística multivariada y desconoce muchos de los principios de los análisis comúnmente utilizados. Asimismo, en muchas ocasiones no se tiene bien definido el objetivo de la caracterización, es decir, se caracteriza sólo por caracterizar; esto complica la interpretación de los análisis pues se desconoce si se escogieron las características adecuadas y las UBC necesarias.

Por lo anterior, es necesario hacer extensivos los principios básicos de estas técnicas, así como la finalidad de cada uno de ellos y, de esta manera, conservar mejor nuestros recursos genéticos vegetales mexicanos. Un recurso genético que no se conoce no se utiliza; si no se utiliza, no se conserva y, si no se conserva, está destinado a la extinción. En este caso adquiere relevancia la frase “la estadística es demasiado importante como para dejarla sólo a los estadísticos, así como la ciencia es demasiado importante como para dejarla solamente a los científicos” (Johnson, 1998).

## LITERATURA CITADA

- González-Andrés, F. 2001. La caracterización vegetal: Objetivos y enfoques. pp. 189-198. En: González-Andrés, F. & J. M. Pita Villamil (Eds.). Conservación y caracterización de recursos filogenéticos. Publicaciones Instituto Nacional de Educación Agrícola. Valladolid, España.
- Hair, J. F. Jr., R. E. Anderson, R. L. Tatham & W. C. Black. 2001. Análisis multivariante. 5a edición. Traducido por Prentice E, Cano D. Prentice Hall Iberia. Madrid. 799 pp.
- Hidalgo, R. 2003. Variabilidad genética y caracterización de especies vegetales. pp. 2-26. En: Franco, T. L. & R. Hidalgo (Eds.). Análisis estadístico de datos de caracterización morfológica de recursos filogenéticos. Boletín Técnico Núm. 8 Instituto Internacional de Recursos Fitogenéticos. Cali, Colombia.
- Iriondo-Alegría, J. M. 2001. Conservación de recursos fitogenéticos. pp. 15-31. En: González-Andrés, F. & J. M. Pita Villamil (Eds.). Conservación y caracterización de recursos filogenéticos. Publicaciones Instituto Nacional de Educación Agrícola. Valladolid, España.
- Johnson, D. E. 1998. Métodos multivariados aplicados al análisis de datos. Traducido por H. Pérez-Castellanos. International Thomson Editores. Ciudad de México, México. 566 pp.
- Núñez-Colín, C. A. & A. F. Barrientos-Priego. 2006. Estimación de la variabilidad interna de muestras poblacionales, mediante el análisis de componentes principales. *Interciencia* 31(11): 802-806.
- Núñez-Colín, C. A. & D. Escobedo-López. 2011. Uso correcto del análisis clúster en la caracterización de germoplasma vegetal. *Agronomía Mesoamericana* 22(2): 415-427.
- Núñez-Colín, C. A., R. Nieto-Ángel, A. F. Barrientos-Priego, J. Sahagún-Castellanos, S. Segura & F. González-Andrés. 2008. Variability of three regional sources of germplasm of tejacote (*Crataegus* spp.) from Central and Southern Mexico. *Genetic Resources and Crop Evolution* 55(8): 1159-1165.
- Núñez-Colín, C. A. & E. Valadez-Moctezuma. 2010. Análisis estadístico de huellas genómicas, un uso práctico de los paquetes computacionales más populares. Libro Científico Núm. 1. Campo Experimental Bajío, Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias. Celaya, México. 103 pp.
- Ramamoorthy, T. P., R. Bye, A. Lot & J. Fa. 1993. Biological diversity of Mexico. Origins and distribution. Oxford University Press. Nueva York. 812 pp.
- Real Academia Española. 2001. Diccionario de la lengua española. 22ª edición. Real Academia Española, Madrid. En línea: <http://lema.rae.es>
- United Nations Food and Agriculture Organization (FAO). 2001. Tratado Internacional sobre los Recursos Fitogenéticos para la Alimentación y la Agricultura. Versión en español. Resolución 3/2001. Roma. 27 pp.
- Vavilov, N. I. 1992. Origin and geography of cultivated plants. Cambridge University Press. Cambridge, Inglaterra. 505 pp.
- Villaseñor, J. L. 2003. Diversidad y distribución de las Magnoliophyta de México. *Interciencia* 28: 160-168.