

Informe Final

(Primera Etapa)

Proyecto: Análisis genómico de especies del genero *Phaseolus* para el mejoramiento del frijol común (*Phaseolus vulgaris*).

Responsable: Dr. Alfredo Heriberto Herrera Estrella

Unidad de Genómica Avanzada (LANGEBIO)–CINVESTAV

Km 9.6 Libramiento Norte Carretera León, 36821 Irapuato, Guanajuato

Teléfonos: 52 462 166 3000

Correo-e: alfredo.herrera@cinvestav.mx

Resumen

El frijol en México se considera un producto estratégico en el desarrollo rural y social del país, con una gran tradición productiva y de consumo, y que cumple diversas funciones tanto de carácter alimentario como para el desarrollo socioeconómico. Existen 570 mil productores dedicados al cultivo de frijol, la superficie cultivada promedio del frijol es de 2.3 millones de hectáreas por año, ocupando el segundo lugar en importancia dentro de la superficie sembrada total a nivel nacional, sólo después del maíz. Los principales países productores del mundo llegan a obtener un rendimiento de hasta 1.86 ton/ha, mientras que en México se obtienen rendimientos promedio inferiores a una tonelada por hectárea. Entre los factores que limitan su producción en el país se encuentran las enfermedades, las cuales causan pérdidas en el rendimiento que van desde un 25 hasta un 50%. Sin embargo, la principal limitante en su producción, la constituye la escasa disponibilidad de agua. Por ejemplo, durante el año 2011, debido a la sequía que afectó a las principales entidades productoras, la cosecha de frijol se redujo a 45.3% y la producción cayó 50.9% con relación a 2010 (SAGARPA, 2012).

La fuente de información para el desarrollo futuro de programas de mejoramiento de cultivos y por tanto para la alimentación humana, reside en la enorme diversidad genética que existe entre las especies vegetales silvestres y cultivadas. Por ello, para resolver esta problemática pretendemos aprovechar el hecho de que México cuenta con algunos de los reservorios de biodiversidad silvestre y domesticada más importante para la preservación del patrimonio biológico y de la seguridad alimentaria de la humanidad.

Por ello, propusimos explorar la diversidad genética existente en diferentes especies del género *Phaseolus* para el desarrollo de nuevas variedades de *Phaseolus vulgaris* (frijol común). Con estos resultados se podrá acelerar la incorporación de información genética derivada de especies silvestres en especies cultivadas. Así se podrían usar caracteres de tolerancia a la carencia de agua, a altas temperaturas o a enfermedades, presentes en la gran variedad de especies silvestres.

Palabras clave: Frijol, genoma, mejoramiento, cambio climático

Introducción

Los cultivos críticos para la alimentación humana, como el frijol común, la leguminosa de grano más importante del mundo, generalmente se adaptan a temperaturas más frías y a condiciones moderadas de humedad del suelo, mientras que las temperaturas más altas y la sequía reducen dramáticamente su rendimiento. Teniendo en cuenta el rápido cambio en el clima mundial, que se prevé dará lugar a aumentos de temperatura en todo el mundo de 2 a 6 ° C para 2100 (Rowlands et al., 2012, IPCC, 2007) que resultarían en la reducción dramática del rendimiento en algunos países productores de frijol Para 2030 (Palomino, 2012), el mejoramiento para la tolerancia al calor y a la sequía parecen tener importancia crítica para los cultivos sensibles al calor como el frijol común. Además, la investigación y el desarrollo de especies alternativas de leguminosas de grano, como el frijol tepari, son necesarios para ambientes de producción futuros o para ambientes marginales existentes que no soportan las especies de cultivos actuales. Por ello, resulta de gran importancia elucidar el genoma del frijol Tepari (*P. acutifolius* A. Grey), y de la especie cultivada más estrechamente relacionada a este (*P. vulgaris*).

El frijol Tepari (*Phaseolus acutifolius* A. Grey) está adaptado a zonas agroecológicas de alta temperatura y áridas y se origina en la región mesoamericana que comprende el suroeste de los Estados Unidos, México y Centroamérica (Freeman, 1912, 1913, Vavilov, 1931) El desierto de Sonora (Thomas et al., 1983). El frijol común (*Phaseolus vulgaris*), proveniente de los acervos genéticos mesoamericano y andino, es la especie ampliamente cultivada más estrechamente relacionada, al frijol Tepari. Las especies de *P. acutifolii* consisten en tres variedades-*acutifolius*, *latifolius* y *tenuifolius*-mientras que el frijol tepari cultivado pertenece a la variedad *latifolius* Freeman (Freytag y Debouck, 2002). Los nativos americanos han utilizado tepari bean en su producción en zonas áridas por más de 5000 años, aprovechando su adaptación al desierto y su alto valor nutricional (revisado por Nabhan y Felger, 1978). Sin embargo, los esfuerzos limitados de comercialización, el pequeño tamaño de su semilla y sus características culinarias, así como el desarrollo de sistemas de riego en el suroeste de los Estados Unidos, pueden haber limitado su diseminación y crecimiento para cultivo en tierras secas.

Objetivo General

Elucidación del genoma de especies del genero Phaseolus para su utilización en el descubrimiento de genes novedosos y la introgresión de caracteres de interés agronómico en nuevas variedades comestibles.

Objetivos Particulares

- 1) Establecer una secuencia genómica de referencia para dos especies del genero Phaseolus.

Resultados Obtenidos

A pesar de que contamos con dos genomas de genotipos cultivados de *P. vulgaris*, planteamos secuenciar con mejor tecnología la línea BAT93, para obtener un verdadero genoma de referencia que nos diera total certidumbre para su utilización en programas de mejoramiento genético. Para ello sembramos trescientas plantas en condiciones in vitro para asegurar un alto porcentaje de germinación y que al menos de arranque estuvieran libres de patógenos. Se logró obtener doscientas noventa y cinco plántulas, mismas que fueron transferidas a suelo “esterilizado” con vapor.



Después de tres semanas de crecimiento las plantas fueron transferidas a condiciones de obscuridad por 48 h y se colectaron hojas jóvenes hasta obtener aproximadamente 200 g de tejido, mismo que fue congelado de inmediato en N₂ líquido. A partir de este tejido se purificaron núcleos para obtener DNA de alto peso molecular. La calidad del DNA fue verificada por medio de electroforesis en campo pulsado, observándose un tamaño mínimo de aproximadamente 150 kb y tamaño promedio de 200 kb. El DNA obtenido fue fragmentado con el sistema BluePippin y se prepararon dos bibliotecas. Las bibliotecas fueron secuenciadas a una profundidad estimada de 60 equivalentes del genoma en la plataforma PacBio Sequel, con base en el tamaño esperado del genoma de 600 Mpb. Utilizando esta información y el software Canu 1.6 obtuvimos tres ensamblados con los siguientes resultados (Tabla 1).

Tabla 1. Estadísticas de ensamblados del genoma de *P. vulgaris* BAT93

Versión	No. de Contigs	Bases ensambladas (Mpb)	Contig más largo (Mpb)	N50 (tamaño/número)	N90 (tamaño/número)	Error de ensamblado
1.1	562	564.502	24.388	6,325,151/28	1,284,168/101	0.045
1.2	536	562.170	19.246	6,698,136/28	1,136,406/105	0.075
1.3	566	565.660	19.977	6,009,418/29	1,300,736/99	0.045

Los tres ensamblados obtenidos son muy superiores al reportado anteriormente (Vlasova et al., 2016), ya que dicho ensamblado cubría 549.6 Mpb fragmentado en 68,379 scaffolds/contigs, con N50(tamaño/número) de 433,759/324 y N90(tamaño/número) 2,023/8,894. Estos ensamblados también son muy superiores a los reportados para la raza de *P. vulgaris* (G19833), de origen Andino, cuyo ensamblado cubría solamente 472.5 Mpb con N50(tamaño) de 39,000pb (Schmutz et al., 2014). Consideramos que la versión 1.3 es nuestro mejor ensamblado, ya que tiene una ligeramente mayor cobertura del genoma que el 1.1 y el 90% del genoma esta cubierto en menos de 100 contigs de tamaños mayores a 1.3 Mpb. Debemos resaltar que a pesar de que para calcular la profundidad consideramos un tamaño de genoma de 600 Mpb, en la actualidad el tamaño real del genoma de *P. vulgaris* se estima es de 587 Mpb, por lo que nuestra cobertura es del 96.4%.

Para predecir el número de genes en el genoma de *P. vulgaris* BAT93 utilizamos el programa Augustus, entrenado datos transcriptómicos de *P. vulgaris* disponibles en bases de datos publicas. Esta predicción nos arrojó un total de 84,816 genes, de los cuales eliminamos redundancia utilizando

el programa cd-hit, resultando en 74,414 genes no redundantes. Para verificar que tan completo está el ensamblado del genoma, utilizamos Augustus alimentado con *embryophyta_odb9*, la cual tiene 1440 grupos BUSCO, donde los parámetros de especie por default es *Arabidopsis*. Sin embargo, también utilizamos arroz como referencia. De los 1440 BUSCO esperados, al utilizar *Arabidopsis* como referencia encontramos 1307 completos, con 1229 singles y 78 duplicados, indicando una cobertura del 90.76% (Fig. 1). Al utilizar arroz como referencia, estos números cambian ligeramente, obteniéndose 1299 completos, con 1221 singles y 78 duplicados, indicando una cobertura del 90.20% (Fig. 1).

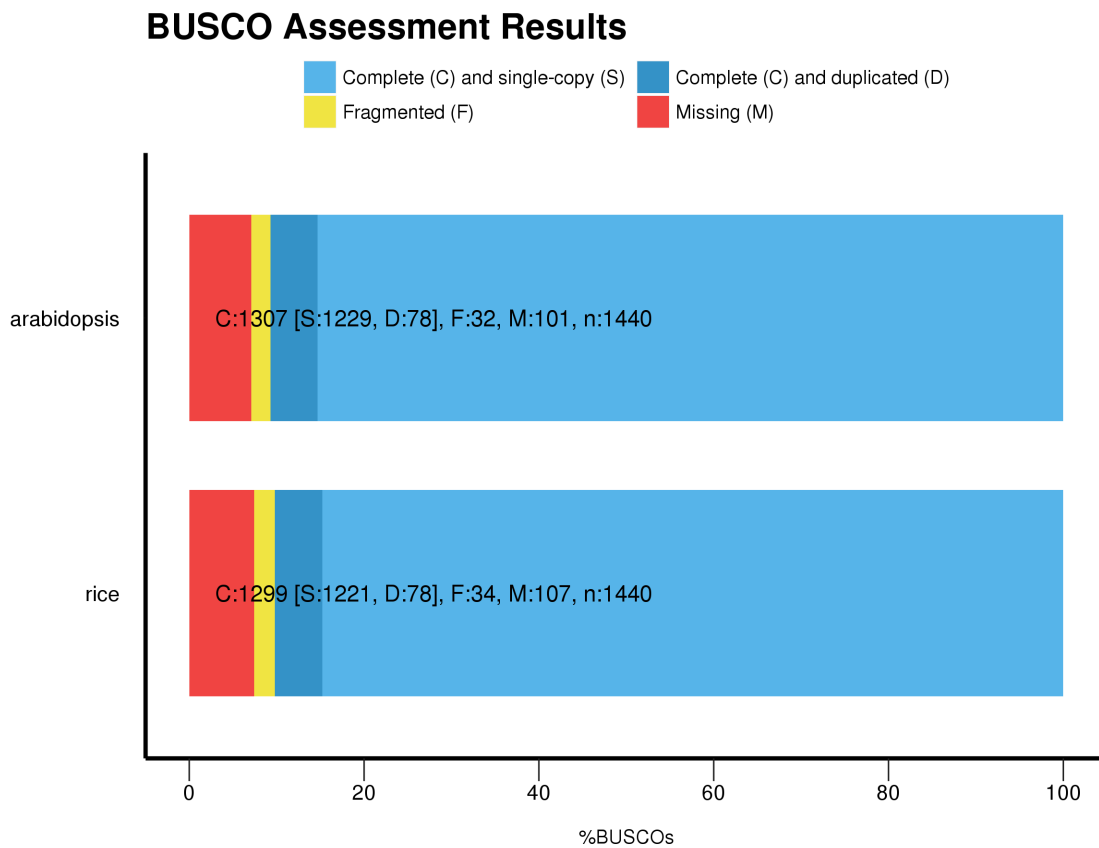


Figura 1. Análisis de grupos BUSCO para el genoma de *P. vulgaris* BAT93.

Así mismo, en esta primera etapa, prometimos secuenciar el genoma completo de *P. acutifolius* con una profundidad de sesenta equivalentes (60X) del genoma, con un tamaño de genoma estimado en 734 Mbp. Para ello a través del Dr. Caspar Chater, colaborador de la Dra. Alejandra Covarrubias, obtuvimos las líneas TARS-Tep 22 (semilla blanca), TARS-Tep 23 (semilla negra) y TARS-Tep 32

(semilla amarilla) del “USDA-ARS National Plant Germplasm System” de los Estados Unidos. Tep 22 (Núm. De registro GP-288, PI 666350) y Tep32 (Núm. De registro GP-289, PI 666351) son líneas de germoplasma de frijol tepari tolerantes a estrés abiótico y biótico, adaptadas a regiones templadas y al trópico húmedo. Estas líneas de germoplasma fueron desarrolladas en una la colaboración de la Estación de Investigación Agrícola Tropical del USDA-ARS (TARS), la Estación Experimental Agrícola de la Universidad de Puerto Rico (UPR) y la Universidad Estatal de Colorado con el objetivo de desarrollar germoplasma con mejores características agronómicas y de tolerancia a estrés (Porch et al., 2013). En particular, Tep 32 es altamente tolerante a sequía, y tiene un alto rendimiento en comparación con otros teparies bajo las condiciones en que fueron probados por el Dr. Chater. Para obtener el material vegetal necesario para la extracción de suficiente DNA, fue necesarios propagar la semilla de las líneas seleccionadas. Desafortunadamente, nunca logramos hacer germinar a las semillas de la línea Tep22. Sin embargo, no tuvimos problemas con las otras dos líneas, una vez obtenida suficiente semilla sembramos doscientas plantas de cada línea en suelo “esterilizado” con vapor.



Después de cuatro a cinco semanas de crecimiento las plantas fueron transferidas a condiciones de obscuridad por 48 h y se colectaron hojas jóvenes hasta obtener aproximadamente 150 g de tejido, mismo que fue congelado de inmediato en N₂ líquido. A partir de este tejido procedimos a la extracción de DNA nuclear de la línea Tep32. La calidad del DNA fue verificada por medio de electroforesis en campo pulsado, observándose un tamaño mínimo de aproximadamente 150 kb y tamaño promedio de 220 kb. El DNA obtenido fue fragmentado con el sistema BluePippin y se prepararon dos bibliotecas. Las bibliotecas fueron secuenciadas a una profundidad estimada de 65 equivalentes del genoma en la plataforma PacBio Sequel, con base en el tamaño esperado del genoma de 734 Mbp. Utilizando esta información y el software Canu 1.6 obtuvimos tres ensamblados con los siguientes resultados (Tabla 2).

Tabla 2. Estadísticas de ensamblados del genoma de *P. acutifolius* Tep32

Versión	No. de Contigs	Bases ensambladas (Mpb)	Contig más largo (Mpb)	N50 (tamaño/número)	N90 (tamaño/número)	Error de ensamblado
1.1	1,160	520.643	20.190	6,620,273/24	379,781/107	0.045
1.2	1,068	517.897	26.153	5,999,395/26	303,213/133	0.075
1.3	1,189	521.256	16.469	6,501,879/26	325,273/114	0.045

Consideramos que la versión 1.3 es nuestro mejor ensamblado, ya que tiene una ligeramente mayor cobertura que la 1.1 y el 90% del genoma esta cubierto en tan solo 114 contigs de tamaños mayores a 325 Kpb. A pesar de que las estimaciones del tamaño del genoma de *P. acutifolius* disponibles para calcular la cobertura necesaria era de 734 Mpb, recientemente se reporto que el tamaño más preciso estimado por citometría de flujo es de sólo 537 Mpb (Lobaton et al., 2018). Lo anterior es consistente con nuestros ensamblados y nos indica que hemos alcanzado a ensamblar el 97% de su genoma con una cobertura de 85X.

Para predecir el número de genes en el genoma de *P. acutifolius* Tep32 utilizamos el programa Augustus, entrenado datos transcriptómicos de *P. vulgaris* disponibles en bases de datos publicas. Esta predicción nos arrojó un total de 74,037 genes, de los cuales eliminamos redundancia utilizando el programa cd-hit, resultando en 67,548 genes no redundantes. Como en el caso de *P. vulgaris*, para verificar que tan completo está el ensamblado del genoma, utilizamos Augustus alimentado con embryophyta_odb9. En este caso, de los 1440 BUSCO esperados, al utilizar Arabidopsis como referencia encontramos 1296 completos, con 1222 singles y 74 duplicados, indicando una cobertura

del 90% (Fig. 2). Al utilizar arroz como referencia, estos números cambian ligeramente, obteniéndose 1291 completos, con 1225 singles y 66 duplicados, indicando una cobertura del 89.65% (Fig. 2).

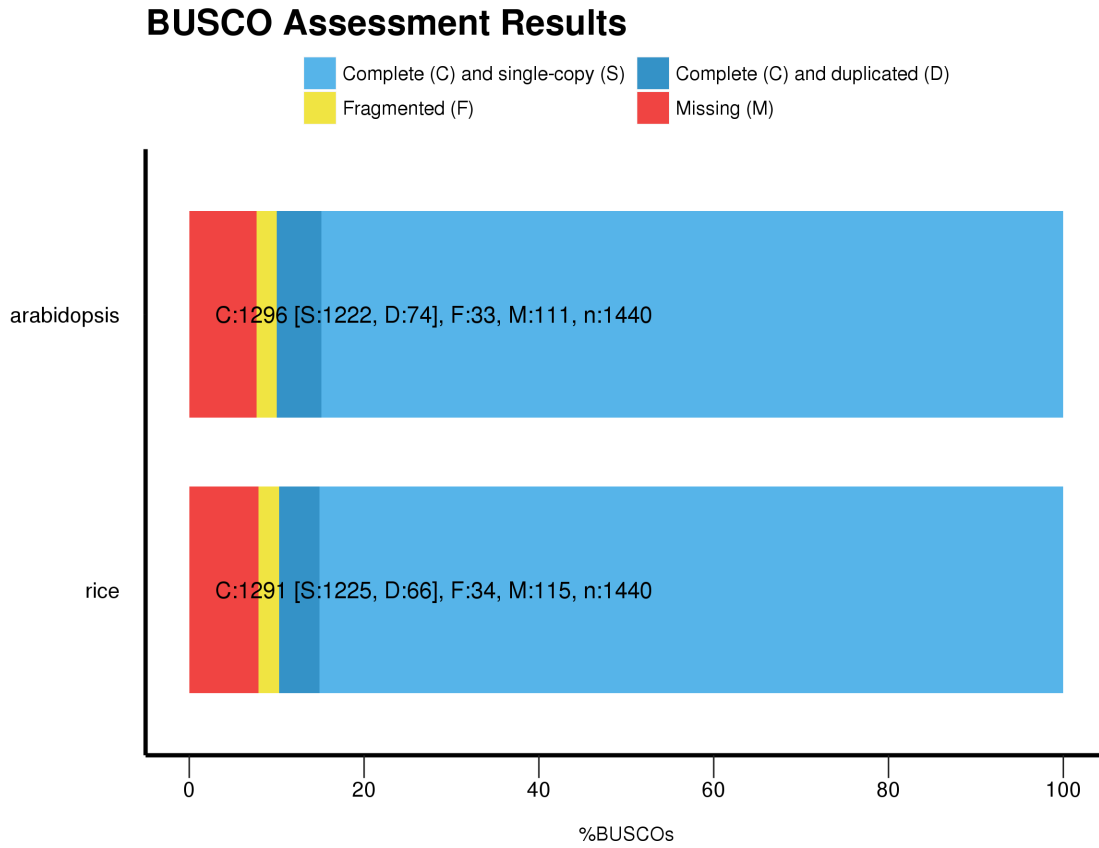


Figura 2. Análisis de grupos BUSCO para el genoma de *P. acutifolius* Tep32.

Referencias

- Freeman, G.F. 1912. Southwestern beans and teparies. Univ. Ariz. Agric. Exp. Stn. Bull. 68. Univ. of Arizona, Tucson.
- Freeman, G.F. 1913. The tepary, a new cultivated legume from the Southwest. Bot. Gaz. (Chicago) 56:395–417. doi:10.1086/331183
- Freytag, G.F., and D.G. Debouck. 2002. Taxonomy, distribution and ecology of the genus *Phaseolus* (Leguminosae–Papilionoideae) in North America, Mexico and Central America. BRIT (Bot. Res. Inst. Texas), Ft. Worth.
- Lobaton J. D., Miller T., Gil J., Ariza D., de la Hoz J. F., Soler A., Beebe S., Duitama J., Gepts P., and Raatz B. 2018. Resequencing of common bean identifies regions of inter-gene pool introgression and provides comprehensive resources for molecular breeding. *The Plant Genome* 11:170068.
- Nabhan, G.P., and R.S. Felger. 1978. Teparies in southwestern North America. *Econ. Bot.* 32:3–19.
- Rowlands, D., D.J. Frame, D. Ackerley, T. Aina, B.B.B. Booth, C. Christensen, et al. 2012. Broad range of 2050 warming from an observationally constrained large climate model ensemble. *Nat. Geosci.* 5:256–260.
- Thomas, C.V., R.M. Manshardt, and J.G. Waines. 1983. Teparies as a source of useful traits for improving common beans. *Desert Plants* 5:43–48.
- Vavilov, N.I. 1931. Mexico and Central America as the principal centre of origin of cultivated plants of the New World. *Trudy po Prikladnoi Botanike, Genetike i Seleksii* [Bull. Appl. Bot., Genet. Plant Breed.] 26:179–199.
- Porch T. G., Beaver J. S., and Brick M. A. 2013. Registration of Tepary Germplasm with Multiple-Stress Tolerance, TARS-Tep 22 and TARS-Tep 32. *J. Plant Regist.* 7: 358-364.
- Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, et al. 2014. A reference genome for common bean and genome-wide analysis of dual domestications. *Nat Genet.* 46:707–13.
- Vlasova A., Capella-Gutiérrez S., Rendón-Anaya M., Hernández-Oñate M., Minoche A., Erb I., Câmara-Ferreira F., Prieto-Barja P., Corvelo A., Sanseverino W., Westergaard G., Dohm J.C., Pappas Jr G.J., Saburido-Alvarez S., Kedra D., Gonzalez I., Cozzuto L., Andreu N., Aguilar M., Garcia-Mas J., Zehnsdorf M., Vázquez M.P., Delgado-Salinas A., Delaye L., Lowy E., Mentaberry A., Vianello-Brondani R.P., Aguilar O.M., García J.L., Alioto T., Sánchez F., Himmelbauer H.,

Santalla M., Notredame C., Gabaldón T., Herrera-Estrella A., & Guigó R. 2016. Genome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and temporal specialization of genes. *Genome Biol.* 17: 32.