



UNIVERSIDAD DE TALCA
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA CIVIL EN BIOINFORMÁTICA

**Estudio de las faseolinas presentes en el endosperma
de las variedades locales de *Phaseolus vulgaris L.* de
la raza Chile mediante herramientas bioinformáticas.**

NATALIA FRANCISCA RODRÍGUEZ CABELLO

Profesor Tutor: BÁRBARA ARÉVALO RAMOS

Memoria para optar al título de
Ingeniero Civil en Bioinformática

Talca – Chile
Julio, 2022

CONSTANCIA

La Dirección del Sistema de Bibliotecas a través de su unidad de procesos técnicos certifica que el autor del siguiente trabajo de titulación ha firmado su autorización para la reproducción en forma total o parcial e ilimitada del mismo.



Talca, 2022



UNIVERSIDAD DE TALCA
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA CIVIL EN BIOINFORMÁTICA

**Estudio de las faseolinas presentes en el endosperma
de las variedades locales de *Phaseolus vulgaris L.* de
la raza Chile mediante herramientas bioinformáticas.**

NATALIA FRANCISCA RODRÍGUEZ CABELLO

Profesor Tutor: BÁRBARA ARÉVALO RAMOS

Profesor Co-Tutor: BASILIO CARRASCO

Profesor Informante: WENDY GÓNZALEZ

Memoria para optar al título de
Ingeniero Civil en Bioinformática

Talca – Chile
Julio, 2022

*Dedicado a Olaya, mi abuela.
Los brazos que me cobijaron hasta los 5 años.*

AGRADECIMIENTOS

Agradezco a mis padres Cecilia y Osvaldo, por ser mi soporte y principal apoyo. A Cecilia, mi hermana, que sin duda me hace ver que tengo capacidad para lograrlo todo. A la Magda, mi sobrina que llegó hace ocho meses a alegrarnos la vida. A la Coca, nino, Cata y Lore, mi segunda familia, me han acompañado toda la vida. A mi cuñado Matías que quería un lugar en estos agradecimientos. A la Gory, que me peinaba y daba la comida antes de ir al colegio todos los días con gran paciencia. A trini, que tan 'chiquitita' le gustaba acompañarme a la u.

A la profesora Bárbara Arévalo, mi tutora. Quien guió mi primera investigación en el año 2019 (me retaba por tratarla de usted) y no ha dejado de acompañarme desde entonces. Gracias por enseñarme gran parte de lo que sé ahora, por incentivar me a descubrir más y dejar desarrollar mis propias ideas. Al profesor Basilio Carrasco, co-tutor, por darme la oportunidad de realizar esta investigación en el Ceap y haberme enseñado la diversidad de porotos existentes.

A Sandra Gutiérrez, mi sandri. Me dió oportunidades dentro de la universidad cuando era completamente otra persona. Vió mi proceso de crecimiento, mi lucha por subir mis notas para mantenerme en admisión. Gracias por dejar quedarme en su oficina por horas cuando me daba temor venir a la u.

A mis profesores por darme los conocimientos y herramientas para ser una buena profesional. A los funcionarios, que también son parte de este camino. Gracias a la tía Carmen, profesor Raúl, Jaz, Josy, Elena, Erick. A mis compañeros, Gabriel, Fran S y Chicha. Mis amigos de admisión Pauli y Dami. A mi amiga de la vida Javi. A mis amigos del colegio, Felipe, Meli, Yis, Caro y Jupi. A los que se fueron, porque también aprendí de ellos y a los que llegaron.

...Infinitamente gracias a la vida por haberme enseñado tanto todos estos años.

ÍNDICE DE CONTENIDOS

	página
Dedicatoria	I
Agradecimientos	II
Índice de Contenidos	III
Índice de Figuras	v
Índice de Tablas	vii
Resumen	viii
Abstract	ix
1. INTRODUCCIÓN	1
2. HIPÓTESIS	9
3. OBJETIVOS	10
3.1. Objetivo General	10
3.1.1. Objetivos Específicos	10
4. MATERIALES Y MÉTODOS	11
4.1. Análisis de expresión genética de transcritos de faseolinas presentes en <i>Phaseolus vulgaris L.</i> , sobre datos de <i>RNAseq</i>	12
4.2. Relaciones evolutivas de las secuencias de faseolinas presentes en <i>Phaseolus vulgaris L.</i> mediante un análisis filogenético.	14
4.2.1. Obtención de secuencias de transcripción para el gen de faseolina.	14
4.2.2. Alineamiento múltiple de secuencias de ARN de <i>Phaseolus vulgaris L.</i>	15
4.2.3. Construcción de árbol filogenético.	17
4.3. Construcción de modelos por homología de las faseolinas de las variedades locales de la raza Chile, y caracterización estructural mediante herramientas bioinformáticas.	17
4.3.1. Construcción de modelos desde secuencias reportadas.	18
4.3.2. Estabilización energética de los modelos.	20
4.3.3. Producción de simulaciones de dinámica molecular.	21

	IV
4.3.4. Análisis estructural de las simulaciones de dinámica molecular.	22
4.3.5. Construcción de modelos prospectivos con muestras propias.	22
5. RESULTADOS Y DISCUSIÓN	24
5.1. Análisis de expresión genética de transcritos de faseolinas presentes en <i>Phaseolus vulgaris L.</i> , sobre datos de <i>RNAseq</i>	24
5.2. Relaciones evolutivas de las secuencias de faseolinas presentes en <i>Phaseolus vulgaris L.</i> mediante un análisis filogenético.	28
5.3. Construcción modelos por homología de las faseolinas de las variedades locales de la raza Chile, y caracterización estructural mediante herramientas bioinformáticas.	30
6. CONCLUSIONES	49
7. BIBLIOGRAFÍA	51
8. ANEXOS	60

ÍNDICE DE FIGURAS

	página
1.1. Especies domesticas de <i>Phaseolus</i>	2
1.2. Porcentaje de contenido de proteína en variedades.	5
1.3. Estructuras de faseolinas reportadas.	7
4.1. Variedad de porotos de <i>Phaseolus vulgaris L.</i> incorporadas en esta investigación.	11
4.2. Metodología general para realizar el análisis de expresión genética de transcritos de faseolinas presentes en <i>Phaseolus vulgaris L.</i> , sobre datos de <i>RNAseq</i>	12
4.3. Metodología general para realizar el análisis filogenético de las secuencias de ARN que codifican a diversas proteínas de <i>Phaseolus vulgaris L.</i>	16
4.4. Metodología general para la construcción de modelos a partir de secuencias ya reportadas.	18
4.5. Metodología general para minimización de energía y producción de dinámica molecular(DM).	20
4.6. Metodología general para la construcción de modelos por homología con muestras propias.	23
5.1. Análisis de expresión diferencial con respecto al gen Phvul.007G059725.v2.1 que codifica a faseolina en las muestras de poroto coscorrón.	24
5.2. Análisis de expresión diferencial con respecto al gen Phvul.007G059725.v2.1 que codifica a faseolina en las muestras de poroto enriqueta.	25
5.3. Análisis de expresión diferencial con respecto al gen Phvul.007G059725.v2.1 que codifica a faseolina en las muestras de poroto tórtola.	25
5.4. Análisis de expresión diferencial con respecto al gen Phvul.007G059725.v2.1 que codifica a faseolina en las muestras de poroto negro argel.	26
5.5. Réplicas biológicas escogidas que poseen mayor nivel de expresión con respecto al gen Phvul.007G059725.v2.1 que codifica a faseolina.	27
5.6. Árbol filogenético de secuencias de ARN que codifican a diversas proteínas de <i>Phaseolus vulgaris L.</i>	29
5.7. Estructura cristalina de <i>Adzuki Bean 7S Globulin-1</i> (PDB 2EA7).	30
5.8. Trímeros ensamblados a partir de los monómeros alfa y beta modelados. . .	31
5.9. Evaluación en <i>ProSA</i> de los trímeros ensamblados antes de equilibrio. . . .	32
5.10. Evaluación en <i>ProSA</i> de los trímeros ensamblados después de equilibrio. . .	33

5.11. Evaluación de la energía potencial en minimización.	35
5.12. <i>RMSD</i> del backbone de las proteínas.	36
5.13. Radio de giro (<i>Rg</i>) del <i>backbone</i> de las proteínas.	37
5.14. Análisis de <i>sitemap</i> para el trímero de tres subunidades alfa luego de la DM.	38
5.15. Análisis de <i>sitemap</i> para el trímero de tres subunidades beta luego de la DM.	39
5.16. Análisis de <i>sitemap</i> para el trímero de dos subunidades alfa y una subunidad beta luego de la DM.	40
5.17. Análisis de <i>sitemap</i> para el trímero de dos subunidades beta y una subunidad alfa luego de la DM.	41
5.18. Análisis de potencial electrostático para el trímero compuesto por tres subunidades alfa.	43
5.19. Análisis de potencial electrostático para el trímero compuesto por tres subunidades beta.	44
5.20. Análisis de potencial electrostático para el trímero compuesto por dos subunidades alfa y una subunidad beta.	45
5.21. Análisis de potencial electrostático para el trímero compuesto por dos subunidades beta y una subunidad alfa.	46
5.22. Alineamiento estructural de los modelos por homología de faseolina y los modelos por homología de una fracción de faseolina a partir de muestras propias.	47
8.1. Figura suplementaria 1. Análisis de expresión diferencial con respecto al gen Phvul.007G059725.v2.1 que codifica a faseolina en las muestras de poroto coscorrón.	60
8.2. Figura suplementaria 2. Análisis de expresión diferencial con respecto al gen Phvul.007G059725.v2.1 que codifica a faseolina en las muestras de poroto enriqueta.	61
8.3. Figura suplementaria 3. Análisis de expresión diferencial con respecto al gen Phvul.007G059725.v2.1 que codifica a faseolina en las muestras de poroto tórtola.	62
8.4. Figura suplementaria 4. Análisis de expresión diferencial con respecto al gen Phvul.007G059725.v2.1 que codifica a faseolina en las muestras de poroto negro argel.	63
8.5. Figura suplementaria 5. Alineamiento de secuencias de faseolinas alfa y beta.	64

ÍNDICE DE TABLAS

	página
1.1. Carbohidratos y análogos presentes en <i>Phaseolus vulgaris</i> L.	4
4.1. Muestras de las variedades de porotos recolectadas por el Ceap para realizar el análisis de expresión diferencial.	13
4.2. Códigos de acceso para secuencias de ARN de <i>Phaseolus vulgaris</i> L.	17
5.1. Promedio de valores obtenidos en alineamiento local de secuencia reportada con <i>RNAseqs</i> sin procesar de todas las muestras de las variedades de porotos.	28
5.2. Resultado alineamiento local de secuencia de ARN de faseolina reportada con secuencias de ARN las muestras estudiadas.	28
5.3. Resultado búsqueda de templado en <i>Blastp</i> para secuencias alfa y beta. . .	31
5.4. Evaluación de la calidad de los trímeros antes y después del equilibrio. . . .	34
5.5. Valores de evaluación obtenidos para los sitios de cada trímero.	42
5.6. Alineamiento en <i>Blastp</i> de muestras propias con monómeros alfa y beta. . .	48

RESUMEN

El poroto (*Phaseolus vulgaris L.*), es la leguminosa alimenticia más importante para el consumo humano en el mundo. Sus granos secos están compuestos principalmente de carbohidratos, además de ser una buena fuente de nitrógeno y proteínas. También contiene magnesio, potasio, calcio, cobre, hierro, zinc, manganeso y azufre. Es una leguminosa abundante en compuestos bioactivos como inhibidores de enzimas, lectinas, fitatos, oligosacáridos y fenoles, que involucran funciones metabólicas tanto en humanos y animales (Suárez-Martínez et al., 2015).

Las proteínas del poroto están divididas en proteínas metabólicas y de almacenamiento (Osborne, 1924). Las globulinas son proteínas de almacenamiento, mientras que las albúminas pueden cumplir la función de almacenar y tener un rol metabólico. Dentro de las globulinas, se encuentran las proteínas 7S y 11S de naturaleza oligomérica, con pesos moleculares entre 150.000 a 190.000 PM y 300.000 a 400.000 PM, respectivamente. La faseolina es una proteína 7S compuesta por tres subunidades, cada subunidad tiene un peso molecular de alrededor de 50.000 PM y con ausencia de enlaces disulfuro (Tang, 2008; Rui et al., 2011).

Algunos estudios han demostrado que la faseolina cumple un rol biológico importante y se encuentra relacionada con la prevención de enfermedades crónicas no transmisibles, reportándose actividad antioxidante, antihipertensiva y disminuyendo los niveles de colesterol (Carrasco-Castilla et al., 2012a; Carrasco-Castilla et al., 2012b; Martino et al., 2012). Aun así, siendo muy abundante en el poroto, pocas investigaciones describen las actividades biológicas para esta proteína en razas locales chilenas.

De acuerdo a esta problemática, es necesario hacer uso de herramientas proteómicas y bioinformáticas con el objetivo de estudiar la composición de las faseolinas y su posible rol en la prevención de enfermedades crónicas no transmisibles.

ABSTRACT

The bean (*Phaseolus vulgaris L.*) is the most relevant food legume for human consumption worldwide. Its dry grains are composed mainly of carbohydrates and are a good source of nitrogen and protein. It also contains magnesium, potassium, calcium, copper, iron, zinc, manganese, and sulfur. It is a legume abundant in bioactive compounds such as enzyme inhibitors, lectins, phytates, oligosaccharides, and phenols, which involve metabolic functions in both humans and animals (Suárez-Martínez et al., 2015).

Bean proteins classify into metabolic and storage proteins (Osborne, 1924). Globulins are storage proteins, while albumins can fulfill storage functions and have a metabolic role. The 7S and 11S proteins within the globulins are oligomeric, with molecular weights between 150,000 to 190,000 MW and 300,000 to 400,000 MW, respectively. Phaseolin is a 7S protein composed of three subunits, each with a molecular weight of around 50,000 MW and with no disulfide bonds (Tang, 2008; Rui et al., 2011).

Some studies have shown that phaseolin plays an essential biological role and is related to the prevention of chronic non-communicable diseases, reporting antioxidant and antihypertensive activity and lowering cholesterol levels (Carrasco-Castilla et al., 2012a; Carrasco-Castilla et al. al., 2012b; Martino et al., 2012). Even so, being abundant in the bean, few investigations describe the biological activities of this protein in Chilean landraces of common beans.

According to this problem, it is necessary to use proteomic and bioinformatic tools to study the composition of phaseolins and their possible role in the prevention of chronic non-communicable diseases.

1. INTRODUCCIÓN

El uso de productos naturales como las plantas, con propiedades terapéuticas, se remonta hasta las antiguas civilizaciones humanas, donde los productos vegetales fueron una de las principales fuentes de drogas (Rates, 2001). Actualmente, las leguminosas han tomado un rol importante en investigaciones para determinar sus propiedades terapéuticas (Oliveira et al., 2018).

Las leguminosas son plantas pertenecientes a la familia Fabaceae. Comúnmente contienen fragmentos subterráneos desarrollados, hojas compuestas, flores papilionáceas sobre las ramas y frutos con vainas. Dentro de sus características se encuentra la gran capacidad que tienen para desarrollar nódulos en las raíces y fijar nitrógeno en simbiosis con bacterias compatibles como *Rhizobium*. Esta característica las hace ser plantas aptas para volver a colonizar los suelos alterados, aquellos que se encuentren con déficit de nitrógeno, entregando un material vegetal abundante en proteínas (Suttie, 1999). Además de ser importantes al ser usadas como alimento, tienen otras aplicaciones, entre las que se incluye los biocombustibles y la fitorremediación (Prasad y Mathew, 2008).

El tipo de alimento que suministran las leguminosas para los humanos es de tres tipos: tubérculos comestibles, encontrados en los géneros *Pachyrhizus*, *Psophocarpus* y *Sphenostylis*, hojas, vainas verdes, y semillas inmaduras encontradas en *Cajanus*, *Pisum*, *Psophocarpus* y *Vigna* o como grano seco encontrados en el género *Phaseolus* (Allen y Lenne, 1998). El género *Phaseolus* (Figura 1.1), ha sido caracterizado como único, ya que algunas de sus 70 especies han sido domesticadas (*Phaseolus vulgaris*, *P. coccineus* y *P. lunatus*). *Phaseolus vulgaris* L. (comúnmente llamado poroto en Chile) en su forma silvestre, se distribuye tanto en Mesoamérica como en América del Sur. Actualmente, se sabe que sufrió dos eventos aislados e independientes de domesticación lo que han conducido a la formación del *pool* genético Mesoamericano y el *pool* genético Andino (Bitocchi et al., 2017).

El poroto es una planta anual, que se cultiva mayoritariamente en regiones templadas y semitropicales debido a sus semillas secas comestibles que poseen diferentes formas como los porotos blancos, porotos rojos, porotos negros, porotos pintos y porotos arándanos. Hace unos 8.000 años atrás, sus primeros cultivos fueron desarrollados en Perú y México. Actualmente su cultivo puede encontrarse en todo el mundo (FAOSTAT, 2021). Se cultivaba en suelos con abundantes precipitaciones, luego su cultivo se extendió a regiones donde la sequía es endémica, como en el noreste de Brasil, y en las zonas más bajas de África oriental (Singh y Jauhar, 2006).

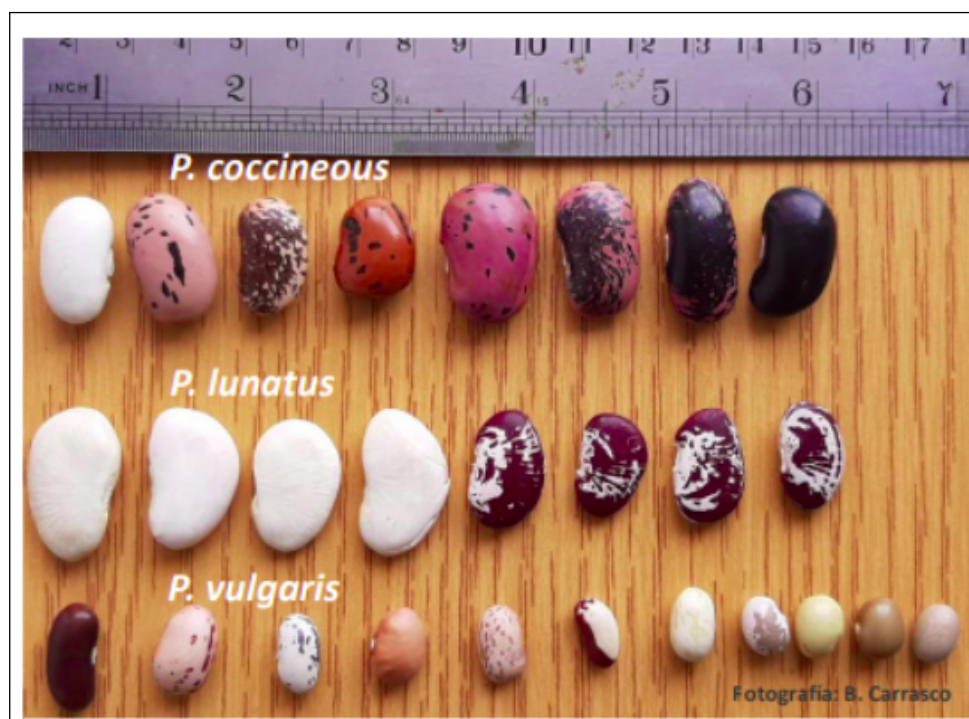


Figura 1.1: Tres especies domesticadas del género *Phaseolus*. *Phaseolus coccineus*, *Phaseolus lunatus* y *Phaseolus vulgaris* (Fotografía B. Carrasco, 2021).

En Chile esta especie se ubica prácticamente en todo el territorio, abarcando desde la Región de Arica y Parinacota hasta la de Los Lagos, exactamente en la provincia de Chiloé. El área cultivada dedicada a la producción de porotos equivale a unas 13 mil hectáreas anualmente (INIA, 2017). Las variedades utilizadas en la producción comercial son mayoritariamente genotipos mejorados, abarcando muchos tipos, unos son destinados para el comercio dentro del país y una parte pequeña de la producción es destinada a la exportación. Esta leguminosa puede ser consumida en estado fresco, que se refiere a su vaina verde y semilla inmadura (poroto granado), también en estado seco como grano seco.

El uso del estado fresco, es determinado por la coloración y la forma de las vainas, teniendo desde tamaños medianos a grandes, coloración verde, son planas y usualmente consumidas completamente (Herrera y Bascur, 1986). Las vainas verdes inmaduras tienen mayores cantidades de vitamina C y fibra dietética, comúnmente se venden enlatadas o congeladas en EE. UU, mientras que en China se venden como verduras frescas. En cuanto al consumo de grano seco en nuestro país, se enfoca en granos con características suaves, los más consumidos corresponden al tipo tórtola y coscorrón. Otros consumidos en menor proporción son los que presentan una sola coloración llamados bayos, mantecas y azufrados y otros de diversos colores como los de tipo frutillas, araucano y sapito (Bascur y Tay, 2005).

Los granos secos se consideran importantes recursos alimenticios siendo una fuente rica en proteínas, almidón, ácidos grasos insaturados, fibras dietéticas, vitaminas y minerales. Normalmente se remojan y cocinan durante unas horas, y se sirven como sopas, guisos y platos con carne (Ganesan y Xu, 2017). Su consumo frecuente (3 porciones a la semana) ha sido asociado con la prevención de enfermedades crónicas no transmisibles. Al respecto, en un estudio realizado con variedades brasileñas de poroto, se demostró que al alimentar a ratas hipercolesterolémicas con las variedades porotos negros, rojos y carioca, disminuían los niveles de colesterol total en sangre en un 16 %, 12 % y 11 %, respectivamente (Martino et al., 2012). Además de su valor nutritivo, llevar una ingesta regular de porotos reduce enfermedades comunes como enfermedades coronarias, la diabetes tipo II y el cáncer. Sin embargo, el valor nutricional del poroto común se ve afectado por una serie de factores que incluyen bajos niveles de aminoácidos que contienen azufre (metionina y cisteína) y triptófano, baja digestibilidad de las proteínas y la presencia de distintos agentes antinutricionales (Krupa, 2008).

Los porotos contienen principalmente carbohidratos (55-65 % en peso seco), de estos, los polisacáridos derivados y no derivados del almidón (fibra dietética) son los componentes primarios con 50 a 68 % (Tabla 1), con cantidades menores pero significativas de mono, di y oligosacáridos. Contienen carbohidratos de digestión lenta y una gran proporción de carbohidratos no digeribles, los cuales pueden fermentar en el intestino grueso. Los carbohidratos no digeridos que llegan al colon incluyen almidón resistente, fibra dietética soluble (pectinas, glucanos) e insoluble (celulosa, lignina) y oligosacáridos no digeribles (Guevara-González y Torres-Pacheco, 2006).

Tabla 1.1: Carbohidratos y análogos presentes en *Phaseolus vulgaris L.*

Carbohidratos y análogos	
Almidón(50 %-60 %)	Amilasa Amilopectina
Fibra dietetica insoluble (10 %-17 %)	Celulosa Hemicelulosa Lignina
Fibra dietetica soluble (5 %-7 %)	Pectinas Oligosacáridos Glucanos Galactomananos

Los principales cationes de los porotos domesticados son el magnesio, el potasio y el calcio, siendo este último el que se encuentra en mayor proporción. Mientras que los minerales que presentan concentraciones medias más altas en genotipos silvestres son el azufre, cobre y manganeso. A pesar de que existen diferencias en los cationes en mayor concentración entre porotos silvestres y domesticados, se observa también que existen cationes en común como lo son: zinc y hierro, los cuales presentan una misma concentración al comparar grano seco entre porotos domesticados y silvestres (Paredes et al., 2009).

Los porotos son una excelente fuente de nitrógeno y proteínas, abarcando entre 20 a 30 % de contenido proteico. Un adulto de 70 kg, puede suplir el 15 % de la ingesta recomendada al comer una porción (70 g) de porotos, ya que aporta aproximadamente 8 g de proteína (Paredes et al., 2009). Los aminoácidos predominantes en el poroto son la histidina, isoleucina, lisina, fenilalanina, treonina, valina y arginina, pero es deficitario en metionina y cisteína. Los cultivos domesticados generalmente presentan niveles más bajos de aminoácidos con respecto a las especies silvestres (Baldi y Salamini, 1973).

Las proteínas son un factor nutricional importante en el poroto común, estas son clasificadas en dos grupos, el primero corresponde a las proteínas metabólicas y el segundo a proteínas de almacenaje. Según Osborne (1924), esta clasificación está basada en la solubilidad de las proteínas de acuerdo a diferentes disolventes. Por lo tanto, las proteínas solubles en sales diluidas son denominadas globulinas, mientras que las proteínas solubles en agua se denominan albúminas. El poroto puede contener entre un 40 % y un 60 % de globulinas, un 12 % a 30 % de albúminas, entre un 20 % a 30 % de glutelinas y un 2 a 4 % de prolaminas.

La función que cumplen las globulinas son netamente proteínas de almacenamiento, mientras el porcentaje presente de albúmina cumple tanto la función almacenamiento como de proteínas metabólicas. En algunas variedades generalmente se estima un porcentaje entre 20 % y 30 % de proteínas (Figura 1.2), como las variedades palo, sapito y tórtola, con un contenido proteico de 30,3 %, 24,1 % y 23,2 %, respectivamente.

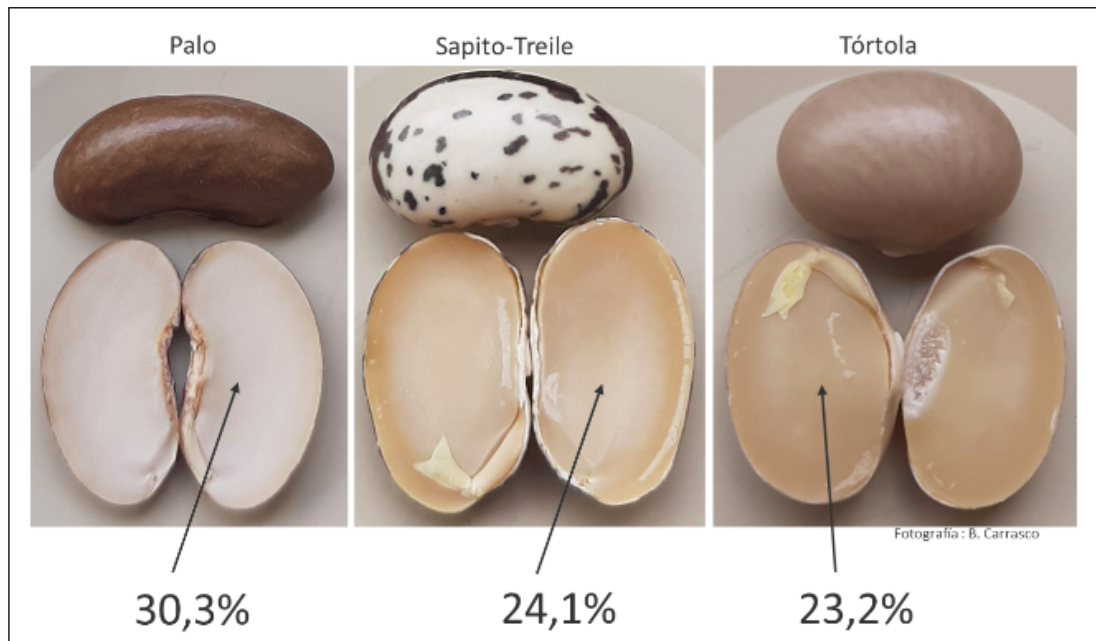


Figura 1.2: Porcentaje de contenido de proteína en las variedades palo, sapito y tórtola (Fotografía B. Carrasco, 2021).

El nombre de las proteínas de almacenaje de los porotos en estado seco es diverso, haciendo que su nomenclatura sea un poco confusa. Si se utiliza el coeficiente de sedimentación por ultracentrifugación (S), los porotos presentan en su contenido las proteínas de almacenamiento 7S y 11S (Tang y Sun, 2011). De acuerdo a la variedad de poroto que se tenga en cuenta, el contenido de un tipo u otro de proteína de almacenamiento cambiará ampliamente.

Las proteínas de tipo 11S comúnmente son proteínas no glicosiladas con un peso molecular aproximado entre 300.000 a 400.000 PM. Estas proteínas están constituidas de seis pares de subunidades unidas por enlaces no covalentes. A su vez, cada par de subunidades se une por un único enlace disulfuro, y está formado por una subunidad ácida con un peso molecular aproximadamente de 40.000 PM y una subunidad básica con un peso molecular aproximado de 20.000 PM (Shewry et al., 1995).

Las globulinas 7S usualmente son proteínas triméricas con un peso molecular entre 150.000 PM a 190.000 PM con ausencia de residuos que contienen cisteína, impidiendo la formación de enlaces disulfuro. En comparación con las globulinas 11S, estas proteínas difieren considerablemente con respecto a la composición de sus subunidades, como consecuencia de los diferentes grados de procesamiento postraducciona (proteólisis y glicosilación) (Sathe, 2016).

Se han identificado tres principales tipos de proteínas 7S, que han sido purificadas y caracterizadas bioquímicamente, el primer tipo corresponde a la faseolina, el segundo a lectina y el tercer tipo a arcelina. Al respecto, se han informado diferentes globulinas 7S donde la glicosilación es más extensa pero no se produce proteólisis, una de ellas es la faseolina 7S de *Phaseolus vulgaris* L. (Hall et al., 1977).

La faseolina es una proteína que presenta un azúcar neutro, y está constituida por tres subunidades, cada subunidad tiene un peso molecular de alrededor de 50.000 PM y con ausencia de enlaces disulfuro (Tang, 2008; Rui et al., 2011). Las diferencias entre polipéptidos son el reflejo de las diferencias en las secuencias de ADN, codificando dos subfamilias de polipéptidos diferentes, uno de ellas es la subfamilia de polipéptidos de α -faseolina, que contienen entre 435 a 444 aminoácidos y la subfamilia de los polipéptidos de β -faseolina, que tienen 421 aa aproximadamente (Slightom et al., 1985).

Una faseolina, podría ser homotrimérica o heterotrimérica al estar constituida por tres subunidades alfa, tres subunidades beta, dos subunidades alfa y una subunidad beta o dos subunidades beta y una subunidad alfa. Solo se pueden dar en estas cuatro conformaciones, ya que si se hace alguna otra conformación, de igual manera corresponderá a dos subunidades alfa y una subunidad beta o dos subunidades beta y una subunidad alfa debido a su conformación trimérica.

La diversidad molecular de la faseolina usualmente ha servido como un indicador evolutivo en el poroto común, tanto en Centroamérica como en todo el territorio Andino. La faseolina entrega mucha información histórica, además de botánica, ya que posee polimorfismo y alta complejidad bioquímica (Gepts, 1988).

En los cultivares de *Phaseolus vulgaris* se han distinguido tres tipos distintos de faseolinas de acuerdo a su movilidad electroforética. Dichas faseolinas se han denominado *Tendergreen* (T), *Sanilac* (S) y *Contender* (C) (Montoya et al., 2008). Sathe (2016) señala

que el análisis de 107 cultivares ha demostrado que las faseolinas de tipo S representan el 69 %, mientras que las faseolinas T el 25 %, y por último las faseolinas de tipo C el 6 % del total de cultivares. Los tipos de faseolinas se pueden identificar utilizando electroforesis en gel unidimensional o bidimensional o en electroforesis en gel de poliacrilamida.

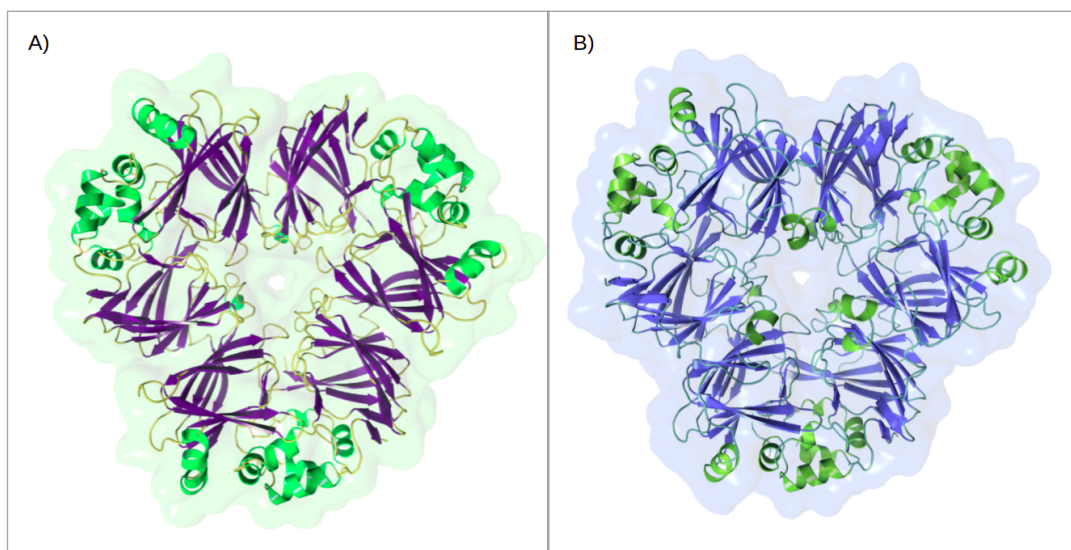


Figura 1.3: A) Estructura de la faseolina de *Phaseolus vulgaris* L. en variedad Francesa (PDB: 2PHL). B) Estructura de globulina 7S de la especie *Vigna angularis* (PDB: 2EA7).

En cuanto a datos cristalográficos de la faseolina, se ha reportado una estructura proveniente de *Phaseolus vulgaris* L. de una variedad francesa. En este cristal se observan mayoritariamente hojas beta (57 %), giros (33 %) y bajo contenido de alfa hélices (10 %) (Figura 1.3A). Otra estructura cristalográfica ha sido reportada como una globulina 7S de la especie *Vigna angularis* correspondiente a un tipo de poroto (Figura 1.3B). Presentando un 49,3 % de hojas beta, un 37 % de giros y un 10 % de alfa hélices. Actualmente, no existen estudios asociados a la estructura de la faseolina para variedades chilenas.

Aunque la faseolina sea muy abundante en el poroto, se ha reportado muy poco sobre su actividad biológica en razas locales chilenas. Carrasco-Castilla y Dávila-Ortiz (2012a) han descrito el resultado de un hidrolizado de proteína a partir de faseolina reportando actividad antioxidante contra un generador de radicales libres, además de actividad quelante férrica y de cobre *in vitro*. Otro estudio de Carrasco-Castilla y Dávila-Ortiz (2012b), informa que análisis bioinformáticos demostraron que las proteínas del poroto contienen muchos péptidos con actividades de tipo antihipertensivo, antioxidante e inhibidor de neuropéptidos que pueden ser liberados por hidrólisis enzimática. Según un nuevo estudio, la

faseolina es un agente con altos efectos antioxidantes y anti genotóxicos, así como con un fuerte potencial anticancerígeno, en el colon y el hígado de ratones a los que se les administró el carcinógeno azoximetano (García-Cordero et al., 2021). Dado esto, el lograr un entendimiento de la faseolina a nivel estructural se torna de gran importancia, para así poder darle un valor agregado a razas locales de poroto y predecir posibles actividades a nivel terapéutico.

Cabe destacar que el uso de las herramientas bioinformáticas es sumamente importante para realizar esta investigación, ya que en diferentes estudios se ha demostrado su utilidad para comprender aspectos evolutivos de la biología molecular, la expresión de los genes a nivel transcripcional, además de ser una ayuda en biología estructural, al poder realizar simulaciones y modelos de estructuras de proteínas. Un ejemplo del uso de técnicas bioinformáticas, es el análisis de las transcripciones mediante la técnica *RNAseq*, revelando las diferencias en el perfil de expresión génica entre genotipos con diferentes tratamientos, con la finalidad de determinar la eficiencia del uso del fósforo en el transcriptoma del poroto común (Silva et al., 2019). Otro estudio bioinformático, analizó la diversidad genética y la estructura poblacional de 219 accesiones de porotos brasileños, la mayoría de las cuales pertenecen al acervo genético mesoamericano (Delfini et al., 2021).

Por lo tanto, considerando las actividades biomédicas beneficiosas del grano seco de *Phaseolus vulgaris L.* y sus diversos constituyentes, así como la relevancia de la faseolina y la ausencia de investigación sobre esta proteína en las razas locales chilenas, es posible que con el uso de herramientas genómicas, proteómicas y bioinformáticas se puedan estudiar diferentes variedades locales determinando diferentes tipos de proteínas de almacenaje (faseolina) en la semilla y su posible implicancia a nivel nutricional y terapéutico.

2. HIPÓTESIS

Las variedades locales de porotos de la raza Chile, presentan diferentes niveles de expresión de faseolinas y variabilidad estructural.

3. OBJETIVOS

3.1. Objetivo General

Caracterizar y analizar las faseolinas presentes en el endosperma de variedades locales de la raza Chile.

3.1.1. Objetivos Específicos

- Realizar análisis de expresión genética *in silico* de transcritos de faseolinas presentes en *Phaseolus vulgaris L.*, sobre datos de *RNAseq*.
- Representar las relaciones evolutivas de las secuencias de faseolinas presentes en *Phaseolus vulgaris L.* mediante un análisis filogenético.
- Construir modelos por homología de las faseolinas de las variedades locales de la raza Chile, y caracterizar estructuralmente mediante herramientas bioinformáticas.

4. MATERIALES Y MÉTODOS

El material vegetal utilizado en esta investigación pertenece al Centro de Estudios de Alimentos Procesados (Ceap).

Corresponden a cuatro tipos de variedades de porotos de la especie *Phaseolus vulgaris* L. (Figura 4.1), cada una de ellas en tres diferentes estados de maduración del grano de poroto. Las variedades coscorrón y tórtola son de raza Chile que pertenece al *pool* genético andino. La variedad negro argel es de raza Durango perteneciente al *pool* mesoamericano. Mientras que la variedad enriqueta es parte de la raza Nueva granada que pertenece al *pool* genético andino. Las razas Nueva granada y Durango en esta investigación se han incorporado a modo de comparación con respecto a la raza Chile.

Es importante señalar que los datos analizados en este estudio corresponde a experimentos de *RNAseq* realizado por el equipo del Dr., Gerardo Tapia del Instituto de Investigaciones Agropecuarias, Estación Experimental Quilamapu, Chillán.

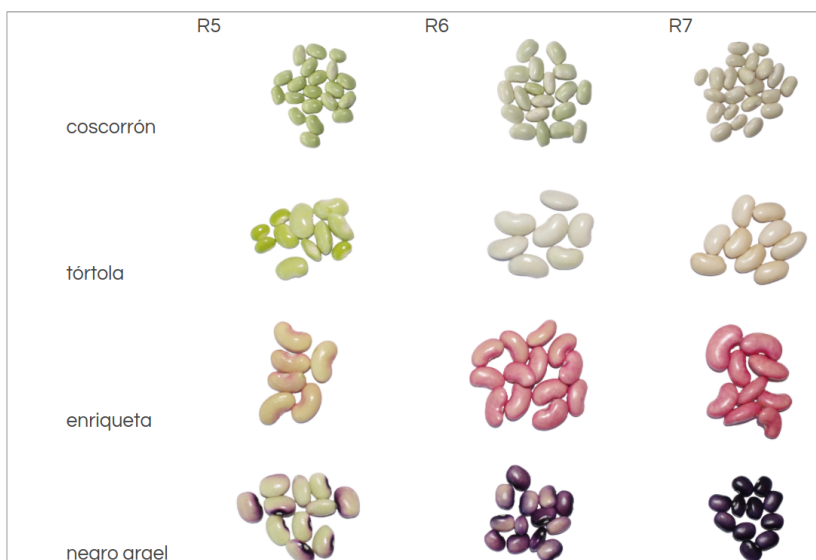


Figura 4.1: Variedad de porotos de *Phaseolus vulgaris* L. incorporadas en esta investigación.

4.1. Análisis de expresión genética de transcritos de faseolinas presentes en *Phaseolus vulgaris* L., sobre datos de *RNAseq*.

La secuenciación de alto rendimiento *RNAseq*, es el uso de tecnología de secuenciación de próxima generación de alto rendimiento para estudiar, caracterizar y cuantificar el transcriptoma de un genoma (Morin et al., 2008). Utilizando esta metodología se puede identificar la expresión diferencial de genes (DEG), genes mutados, genes de fusión e isoformas de genes en estados patológicos (Gao et al., 2017). Actualmente, se ha convertido en el método estándar para medir y comparar los niveles de expresión génica en una amplia variedad de especies y condiciones. Los experimentos de *RNAseq* generan conjuntos de datos muy grandes y complejos que exigen un *software* rápido, preciso y flexible para reducir los datos de lectura sin procesar a resultados comprensibles (Pertea et al., 2016). Contar con una buena metodología, permite a los investigadores alinear las lecturas con un genoma, ensamblar transcripciones que incluyen nuevas variantes de empalme, calcular la abundancia de estas transcripciones en cada muestra y comparar experimentos para identificar genes y transcripciones expresados diferencialmente (Pertea et al., 2016). El protocolo que se ha escogido (Figura 4.2) describe todos los pasos necesarios para manejar un gran conjunto de lecturas de secuenciación sin procesar y crear listas de transcripciones de genes, niveles de expresión y genes, además de transcritos expresados diferencialmente.

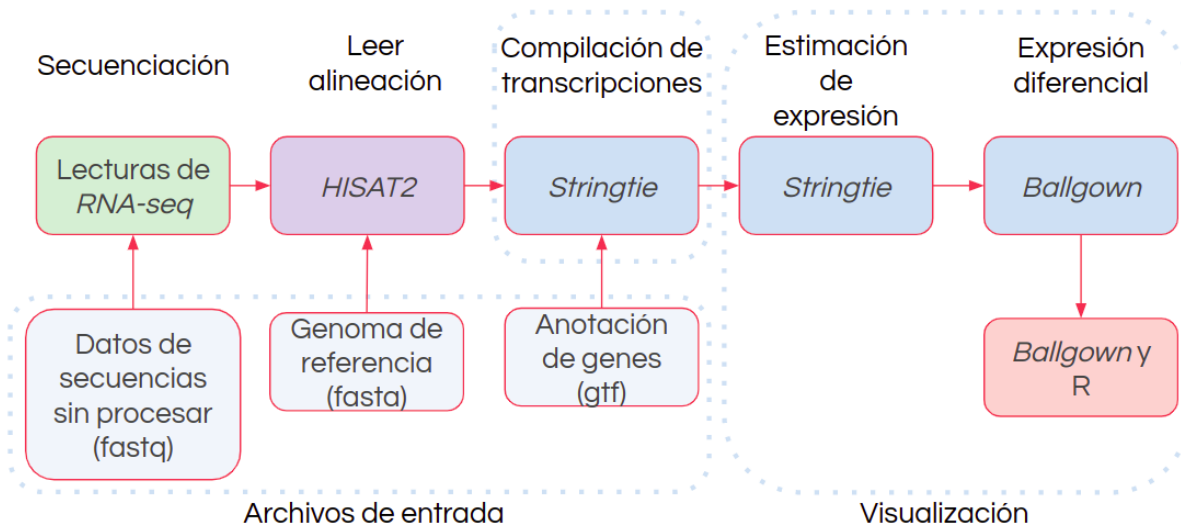


Figura 4.2: Metodología general para realizar el análisis de expresión genética de transcritos de faseolinas presentes en *Phaseolus vulgaris* L., sobre datos de *RNAseq*.

El protocolo comenzó con lecturas de *RNAseq* sin procesar recopiladas de todas las muestras de las variedades de porotos por el Centro de Estudios de Alimentos Procesados (Ceap), que se indican en la tabla 4.1. Obteniendo varios resultados útiles, como la listas de genes, transcritos y niveles de expresión de cada muestra utilizada, también tablas que muestran genes expresados diferencialmente entre las réplicas de cada especie de poroto, junto con medidas estadísticas de importancia.

Tabla 4.1: Muestras de las variedades de porotos recolectadas por el Ceap para realizar el análisis de expresión diferencial.

Tipo/ variedad de poroto	coscorrón	tórtola	negro argel	enriqueta
Raza	Chile	Chile	Durango	Nueva granada
Muestras/ réplicas biológicas	PV24_R5_1	PV48_R5_1	PV132_R5_1	PV217_R5_1
	PV24_R5_2	PV48_R5_2	PV132_R5_2	PV217_R5_2
	PV24_R5_3	PV48_R5_3	PV132_R5_3	PV217_R5_3
	PV24_R6_1	PV48_R6_1	PV132_R6_1	PV217_R6_1
	PV24_R6_2	PV48_R6_2	PV132_R6_2	PV217_R6_2
	PV24_R6_3	PV48_R6_3	PV132_R6_3	PV217_R6_3
	PV24_R7_1	PV48_R7_1	PV132_R7_1	PV217_R7_1
	PV24_R7_2	PV48_R7_2	PV132_R7_2	PV217_R7_2
	PV24_R7_3	PV48_R7_3	PV132_R7_3	PV217_R7_3

PV: *Phaseolus vulgaris* L.

R5;R6;R7: Menor a mayor estado de maduración del grano de poroto.

1;2;3: Réplica biológica.

Primero, las lecturas de cada muestra se asignaron al genoma de referencia con *HISAT-2*, un programa de indexación jerárquica para la alineación empalmada de transcripciones (Kim et al., 2019). Se proporcionó un archivo de posiciones de genes anotados, *HISAT-2* utilizó ese archivo pero también detectó los sitios de empalme que faltan en la anotación. A continuación, las alineaciones se pasaron a *StringTie* para el ensamblaje de la transcripción. Este es un programa que utiliza un enfoque de ensamblaje del transcriptoma utilizando un genoma de referencia junto con conceptos del ensamblaje del genoma *de novo* para mejorar el ensamblaje de la transcripción (Pertea et al., 2015).

StringTie ensambló los genes para cada conjunto de datos por separado, estimó los niveles de expresión de cada gen y cada isoforma a medida que los fue ensamblando. Después de ensamblar cada muestra, el conjunto completo de ensamblajes pasó a la función de fusión de *StringTie*, fusionando todas las estructuras de genes que se encontraron en cualquiera de las muestras. Este paso es necesario porque las transcripciones en algunas de las muestras pueden estar parcialmente cubiertas por las lecturas y, como consecuencia,

solo se ensamblarán versiones parciales de ellas en la ejecución inicial de *StringTie*. El paso de combinación creó un conjunto de transcripciones coherente en todas las muestras, de modo que las transcripciones se pudieron comparar en pasos posteriores. Luego, las transcripciones fusionadas se retroalimentaron a *StringTie* una vez más volviendo a estimar la abundancia de transcripciones utilizando las estructuras fusionadas. La reestimación utilizó el mismo algoritmo que el ensamblaje original. *StringTie* también proporcionó datos adicionales de recuento de lecturas para cada transcripción requerida por *Ballgown*.

Finalmente, en *Ballgown* se tomaron todas las transcripciones y abundancias de *StringTie*, incluidos fragmentos por kilobase de transcripción por millón de lecturas secuenciadas (FPKM), las cuales fueron agrupadas por tipo muestra correspondiente a las especies de porotos descritas en la tabla 4.1 y su número de réplica biológica. *Ballgown* es un paquete de *software* diseñado para facilitar el análisis de expresión diferencial flexible de los datos de *RNASeq*, proporcionando funciones para organizar, visualizar y analizar las medidas de expresión para el ensamblaje de transcriptomas (Frazee et al., 2015). Incluye herramientas de trazado como parte del paquete R/Bioconductor que ayudan a visualizar los resultados, permitiendo general un análisis de expresión bastante detallado.

4.2. Relaciones evolutivas de las secuencias de faseolinas presentes en *Phaseolus vulgaris* L. mediante un análisis filogenético.

4.2.1. Obtención de secuencias de transcripción para el gen de faseolina.

La secuenciación de alto rendimiento es una herramienta nueva y poderosa para obtener información sobre los ácidos nucleicos. *RNAseq* permite tanto el descubrimiento de nuevos genes como una cuantificación computarizada de todos los niveles de ARN (Ira et al., 2015). Aunque las tecnologías de secuenciación de próxima generación han hecho posible generar rápidamente una gran colección de secuencias, se necesitan herramientas disponibles públicamente para trabajar cada tipo de datos. Afortunadamente, se han diseñado una variedad de nuevas herramientas de alineación para realizar un mapeo de lectura eficiente contra grandes secuencias de referencia (Langmead et al., 2009).

Para obtener las secuencias de transcripción de las muestras de ARN que codifican al gen de faseolina se escogió una réplica biológica de cada tipo de poroto a partir del análisis de expresión ejecutado, con el objetivo de seguir trabajando solo con las muestras

que presentaron mayor nivel de expresión diferencial. Se utilizó el programa *Samtools*, que admite lecturas de extremo único y emparejado con combinación de lecturas de diferentes tipos, siendo una biblioteca y un paquete de *software* para analizar y manipular alineaciones en el formato *SAM/BAM* (Li et al., 2009). En este caso, se usaron los archivos en formato *BAM* de las muestras de ARN, que es la representación binaria de *SAM* y mantiene exactamente la misma información que *SAM* (Li et al., 2009). De esta manera los archivos en formato *BAM* se convirtieron a formato *fasta*.

Luego, se usó *BLAST* (*Basic Local Alignment Search Tool*) (Altschul et al., 1990), una herramienta eficaz para realizar alineamientos locales. Se creó una base de datos con la aplicación *makeblastdb* a partir de cada archivo *fasta* de cada muestra. Posteriormente, se utilizó *blastn* para realizar los alineamientos. El archivo de consulta en formato *fasta* para buscar las secuencias similares, fue una secuencia de ARN de faseolina (Phvul.007G059725.1) extraída desde la base de datos PCB (*Pulse Crop Database*, https://www.pulsedb.org/bio_data/1526347) (Humann et al., 2019). De este alineamiento se obtuvieron en promedio 500 coincidencias las que pasaron por un proceso de agrupamiento con el programa *CD-hit*, que se creó especialmente para agrupar y comparar secuencias de proteínas o nucleótidos (Li y Godzik, 2006). De esta manera, los grupos formados corresponden a las secuencias de ARN que codifican al gen de faseolina para cada uno de los tipos de porotos estudiados.

4.2.2. Alineamiento múltiple de secuencias de ARN de *Phaseolus vulgaris* L.

Un alineamiento múltiple de secuencias, es un proceso en el cual se alinean más de dos secuencias que están relacionadas para obtener un emparejamiento óptimo de estas. Las secuencias relacionadas son identificadas de acuerdo a la búsqueda de similitudes utilizando diferentes métodos para alinear múltiples secuencias, los cuales se basan en diferentes algoritmos, teniendo en consideración eventos de carácter evolutivos (Chatzou et al., 2015). Este proceso permite que se puedan identificar patrones y motivos de secuencia conservados en toda una familia de secuencias, que muchas veces no han podido ser detectados cuando sólo se han comparado dos secuencias. Cuando las secuencias a utilizar corresponden a secuencias aminoacídicas, el alineamiento es capaz de identificar residuos de aminoácidos conservados que pueden ser funcionalmente importantes. Además, es de suma importancia que el método a escoger tenga como atributos la precisión biológica y complejidad computacional. Esta última obtiene mayor relevancia a causa del rápido crecimiento de las bases de datos de secuencias, que cada vez contienen familias de proteínas más grandes (Edgar, 2004). Obtener alineamientos múltiples precisos de

secuencias biológicas es todo un desafío y es un paso previo esencial antes de realizar un análisis filogenético (Xiong, 2019). En la figura 4.3, se muestra el flujo de trabajo para realizar el análisis de las relaciones evolutivas de las secuencias de faseolinas presentes en *Phaseolus vulgaris L.*

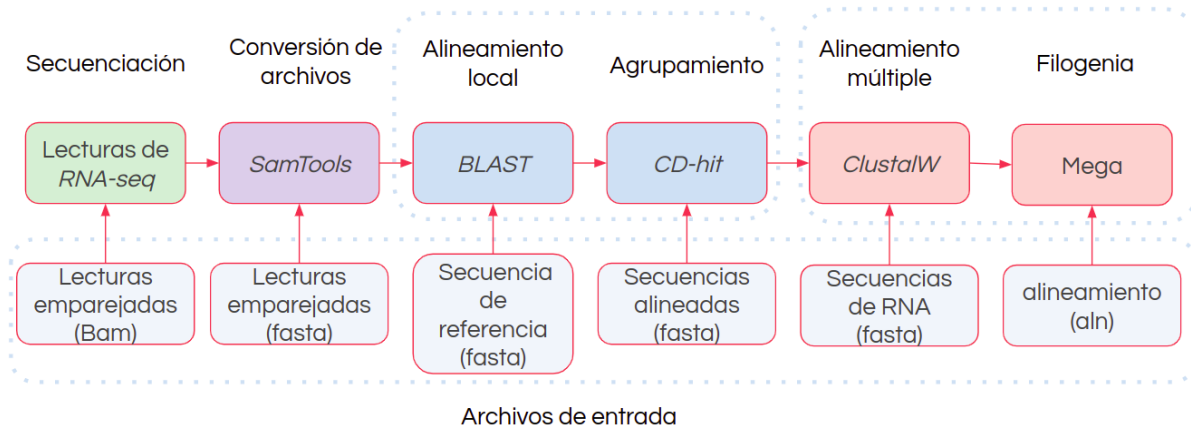


Figura 4.3: Metodología general para realizar el análisis filogenético de las secuencias de ARN que codifican a diversas proteínas de *Phaseolus vulgaris L.*

Una vez obtenidas las secuencias de ARN de *Phaseolus vulgaris L.*, que codifican al gen de faseolina correspondientes a las variedades locales de raza Chile, coscorrón y tórtola, y las variedades extranjeras negro argel y enriqueta. Se agregaron cuatro secuencias de ARN por cada tipo de proteína incluida (albúminas, prolaminas y glutelinas), además como globulina se agregó una secuencia de faseolina que ya ha sido reportada. Trece secuencias de ARN correspondientes a albúminas, prolaminas, glutelinas y una faseolina más las cuatro secuencias de ARN de las variedades estudiadas en esta investigación. Teniendo este conjunto de 17 secuencias de ARN (Tabla 4.2), se realizó un alineamiento múltiple de secuencias con el programa *ClustalW* integrado en el programa *MEGA* (Kumar et al., 2018). *ClustalW* es un método sensible de alineamiento progresivo de múltiples secuencias que se encuentra disponible gratuitamente (Thompson et al., 1994).

Tabla 4.2: Códigos de acceso para secuencias de ARN de *Phaseolus vulgaris* L.

Tipo de proteína	Albúminas	Globulinas	Prolaminas	Glutelinas
Accesiones	Phvul.003G218600.1 Phvul.011G203800.1 Phvul.011G204466.1 Phvul.011G203600.1	Phvul.007G059725.1	Phvul.001G071700.1 Phvul.006G037100.1 Phvul.002G268400.1 Phvul.006G064400.1	Phvul.002G239900.1 Phvul.011G067800.1 Phvul.002G249800.1 Phvul.L001743.1
Muestras propias		PV24_R5_2 PV48_R5_2 PV132_R7.1 PV217_R5_2		

4.2.3. Construcción de árbol filogenético.

La filogenia corresponde al estudio de las relaciones evolutivas de las especies. El análisis filogenético es el medio por el cual podemos estimar estas relaciones evolutivas (Choudhuri, 2014). La relación evolutiva obtenida del análisis filogenético generalmente se describe como un diagrama ramificado en forma de árbol: el árbol filogenético. Un árbol filogenético es una estimación de las relaciones entre secuencias o taxones y sus supuestos ancestros en común (Hall, 2013). Actualmente, muchos de los árboles filogenéticos son construidos a partir de datos moleculares, utilizando secuencias de ADN o proteínas.

Una vez obtenido el alineamiento múltiple de las 17 secuencias de ARN integradas utilizando *clustalW* en el mismo *software MEGA*, se realizó la construcción del árbol filogenético usando el método estadístico *neighbor-joining* y un método de *bootstrap*, entregando un número de 5.000 repeticiones, que corresponden a la proporción de filogenias replicadas que recuperan un clado particular de la filogenia original construida utilizando la alineación original.

4.3. Construcción de modelos por homología de las faseolinas de las variedades locales de la raza Chile, y caracterización estructural mediante herramientas bioinformáticas.

El modelado por homología, corresponde a una metodología que se basa en el principio de alta similitud e identidad que pueda existir entre dos secuencias de proteínas. Si esto existe las proteínas tendrán estructuras similares. Se construye el modelo molecular de una proteína usando su secuencia aminoacídica, basándose en una estructura conocida

perteneciente a la misma familia de proteínas. Esta estructura conocida se denomina como molde o templado. Un modelo por homología es una buena aproximación hacia la estructura de una proteína de interés biológico cuando aún no ha sido cristalizada. Teniendo en cuenta que la cristalografía de rayos X, la resonancia magnética nuclear (RMN) y la criomicroscopía electrónica son procedimientos mucho más lentos, se pueden utilizar métodos computacionales para predecir la estructura de proteínas relacionadas (Bishop, 2008). Este procedimiento es una buena manera de obtener información estructural en caso de que no se puedan desarrollar técnicas experimentales (Weissig y Bourne, 2003). Los pasos que se siguieron para realizar los modelos por homología son mostrados en la figura 4.4, dentro de ellos incluyeron los cuatro pasos rigurosos que deben realizarse: (I) Búsqueda y selección de templados homólogos a la proteína de interés (target), (II) alineación del target con el templado, (III) construcción de uno o más modelos para la proteína diana, (IV) evaluación de la calidad del modelo(s) obtenido(s) (Orry y Abagyan, 2016).

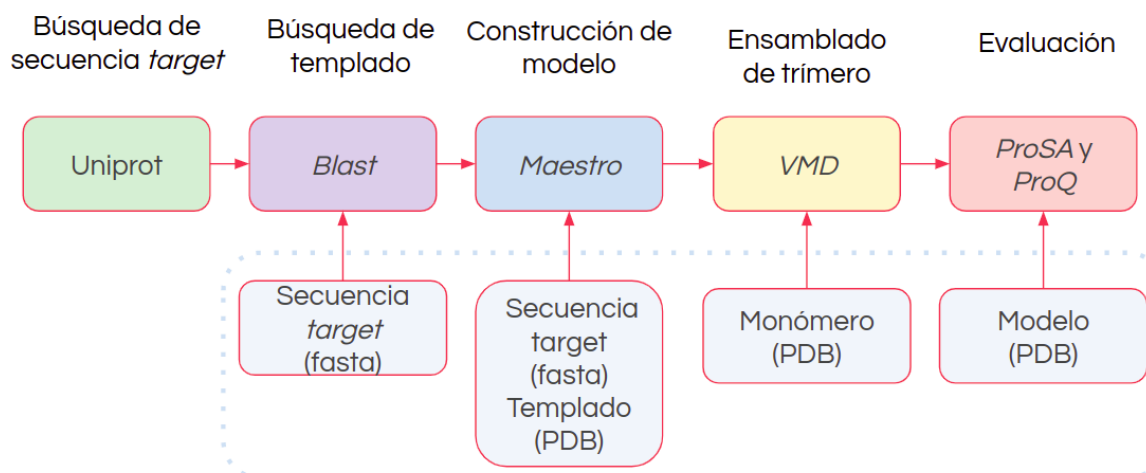


Figura 4.4: Metodología general para la construcción de modelos a partir de secuencias ya reportadas.

4.3.1. Construcción de modelos desde secuencias reportadas.

Esta etapa comenzó con la construcción de modelos por homología de faseolinas en sus cuatro conformaciones a partir de monómeros de secuencias alfa y beta ya reportadas. Ejecutando la misma metodología (Figura 4.4) tanto para la subunidad alfa como beta.

La secuencia de faseolina de subunidades alfa y beta de *Phaseolus vulgaris L.* ya reportadas, se obtuvieron en formato fasta desde la base de datos *UniProt*

(<https://www.uniprot.org/>)(Bateman et al., 2020). El código para la secuencia alfa corresponde a P07219, mientras que el código de la secuencia beta es P02853. Para buscar el templado de cada secuencia, se utilizó *BLAST*, seleccionando en la página de inicio (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) la opción *blastp*. Se cargó en formato fasta la secuencia *target*, para buscar el molde con la base de datos *Protein Data Bank* (pdb) (Berman, 2000).

Para seleccionar el templado desde *Protein Data Bank*, se seleccionó la secuencia que cumpliera con los criterios de alto porcentaje de identidad con la secuencia diana, alto porcentaje de cobertura sobre la secuencia de la proteína, y un valor E que indicara que las coincidencias fueran significativas. Además, se consideró la resolución o calidad del templado.

Para la construcción de los modelos se utilizó el *software Maestro* (Schrödinger, 2021). Con la herramienta *advanced homology modeling*, utilizando como archivos de entrada la secuencia *target* en formato fasta y el templado en formato PDB, primero se generó un alineamiento para posteriormente construir el modelo por homología correspondiente a un monómero. Obteniendo un monómero alfa y un monómero beta.

Luego de la construcción de monómeros alfa y beta, fue necesario realizar la construcción de los modelos completos. Los trímeros fueron ensamblados utilizando la herramienta *stamp* de *VMD* (Humphrey et al., 1996), de acuerdo a su composición de tres subunidades alfa, tres subunidades beta, dos subunidades alfa y una subunidad beta, o dos subunidades beta y una subunidad alfa, generando dos homotrímeros y dos heterotrímeros.

Para evaluar la calidad de los modelos de faseolina construidos, se utilizaron herramientas como *ProSA (Protein Structure Analysis)*, que es una herramienta *web* ampliamente utilizada para comprobar modelos 3D de estructuras de proteínas, basándose en el análisis estadístico de todas las estructuras proteicas disponibles (Wiederstein y Sippl, 2007). Los modelos, además fueron evaluados en *ProQ Server* (<https://proq.bioinfo.se/cgi-bin/ProQ/ProQ.cgi>), que es un predictor que se basa en redes neuronales, obteniendo diferentes características estructurales para predecir la calidad de un modelo de proteína (Wallner y Elofsson 2003). Estos análisis fueron realizados antes del equilibrio y posteriormente para realizar una comparación.

4.3.2. Estabilización energética de los modelos.

Luego de obtener las cuatro conformaciones de faseolinas, los modelos fueron sometidos a refinamiento mediante minimización de energía, siguiendo la misma metodología para los cuatro modelos. Esto se llevó a cabo utilizando el *software Maestro* (Schrödinger, 2021). Los pasos generales se muestran en la figura 4.5.

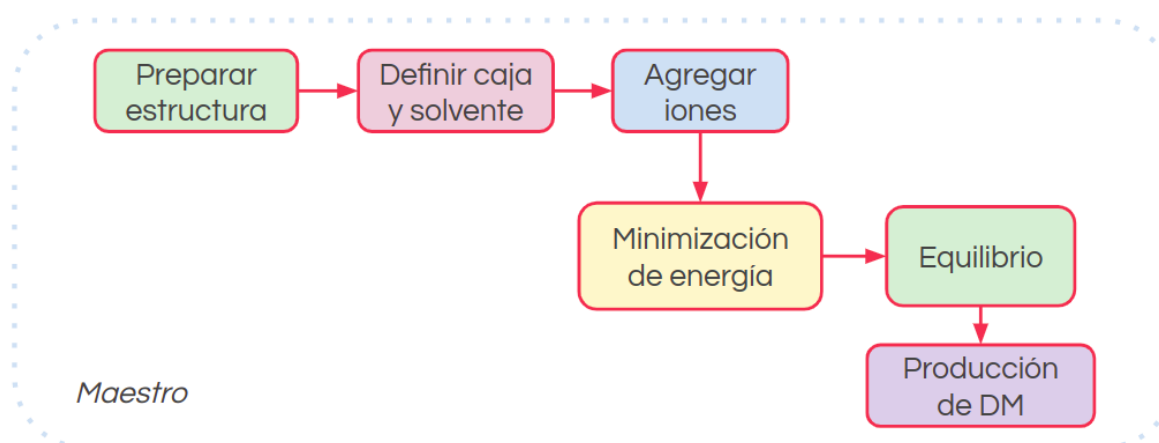


Figura 4.5: Metodología general para minimización de energía y producción de dinámica molecular(DM).

Para ello, primero se preparó la estructura de cada proteína con la herramienta *Protein Preparation Wizard*, con el objetivo de verificar si faltaban átomos de hidrógeno, estados de carga u orientaciones de grupos incorrectos y así obtener un modelo refinado para su posterior análisis.

En segundo lugar, se preparó el sistema con la herramienta *System Builder*, utilizando el campo de fuerza *OPLS3e* (Roos, 2019) que se encuentra validado e incluido dentro del *software Maestro* (Lu et al., 2021), definiendo la caja de solvente la cual rodeó la proteína. Se llenó la caja con agua usando el modelo TIP3P, que especifica una molécula de agua rígida de 3 sitios con cargas y parámetros de *Lennard-Jones* asignados a cada uno de los 3 átomos (Pullman, 2013). El tipo de caja se definió como ortorrómbica, centrando la proteína en la caja, con una distancia de 10 Å con respecto a los bordes periódicos.

En tercer lugar, se agregaron los iones necesarios para neutralizar el sistema con la misma herramienta *System Builder*. Se adicionaron iones de sodio para neutralizar calculando automáticamente la cantidad para cada sistema, además se agregó una concentración de 0,15 M de NaCl con el objetivo de simular las condiciones naturales del ambiente de la

proteína.

Por último, se realizó el proceso de minimización energética, con el fin de obtener sistemas sin impedimentos estéricos o con una geometría inapropiada. Para llevar a cabo la minimización se usó la herramienta *minimization*, definiendo 20.000 pasos de minimización. Finalizada la minimización se realizó un gráfico de energía para verificar que la energía potencial es negativa y estable determinando el éxito de la minimización.

4.3.3. Producción de simulaciones de dinámica molecular.

Las simulaciones de dinámica molecular (DM) son métodos utilizados para comprender las relaciones macromoleculares entre la estructura y su función, durante un tiempo determinado. Las primeras simulaciones de dinámica molecular se desarrollaron a fines de los años 70 (McCammon et al., 1977), pero no fue hasta hace poco se han conseguido escalas de tiempo similares a procesos biológicos. Actualmente, las simulaciones se acercan a la escala de los microsegundos (Gelpi et al., 2015), facilitando mejor la comprensión de la función y estructura de las macromoléculas.

Luego de asegurar tener una estructura inicial en su mínimo energético, se equilibró cada sistema completo, utilizando el *software Maestro*. Se realizó en una sola fase, restringiendo el *backbone* de la proteína. Este equilibrado se llevó a cabo bajo un conjunto NPT, en el que el número de partículas, la presión y la temperatura son constantes. El periodo de tiempo también fue de 25 ns, mediante el acoplamiento a un termostato de *Nosé-Hoover chain* (Nosé, 1984) y al barostato *Martyna-Tobias-Klein* (Martyna et al., 1994) esperando alcanzar una meseta para la presión del sistema en 1.01325 bar. Además, se repitieron los análisis de evaluación de los modelos que fueron realizados antes del equilibrio para poder comparar los resultados obtenidos.

Una vez finalizado el equilibrado, se eliminaron las restricciones posicionales del *backbone* y se procedió a la producción mediante una simulación de dinámica molecular de 100 ns por sistema utilizando el conjunto NPT (isotérmico-isobárico). Esta simulación se realizó para cada sistema con el *software Maestro* (Schrödinger, 2021) y el campo de fuerza *OPLS3e* (Roos, 2019).

4.3.4. Análisis estructural de las simulaciones de dinámica molecular.

Se realizaron análisis de las simulaciones moleculares generadas, con el objetivo de establecer diferencias estructurales para las faseolinas modeladas. Teniendo la finalidad de que estos análisis sirvan en investigaciones futuras, para ayudar a determinar la importancia biológica de la faseolina de las razas chilenas y su relación con los efectos antioxidantes, anti genotóxicos o con un fuerte potencial anti cancerígeno, ya demostrado en otras razas (García-Cordero et al., 2021).

Para los primeros análisis luego de las simulaciones, se utilizó la herramienta *Simulation Event Analysis* de *Maestro* para calcular la desviación cuadrática media (*RMSD*) del *backbone* con el objetivo de medir la diferencia entre las cadenas principales de la proteína en evaluación, desde su conformación estructural inicial hasta su posición final, determinando este cálculo durante el curso de su simulación (Aier et al., 2016). Cuanto menores son las desviaciones, más estable es la estructura de la proteína. También se calculó el radio de giro (*Rg*) de la proteína durante la simulación con la misma herramienta en *Maestro*. Si la proteína se pliega de manera estable, probablemente mantendrá un valor relativamente estable de *Rg*.

Cada trímero modelado también fue sometido a un análisis de *Sitemap* y potencial electrostático. La herramienta *SiteMap* (Halgren, 2009) del *software Maestro* (Schrödinger, 2021) permitió detectar las cavidades de los trímeros de las faseolinas modeladas y sus propiedades fisicoquímicas. También, utilizando el mismo *software*, se usó el panel del potencial electrostático (*Electrostatic Potential Surfaces*), el cual se deriva de las cargas atómicas parciales de la proteína resolviendo la ecuación de *Poisson Boltzmann* en una matriz de puntos en tres dimensiones para luego evaluar con un modelo de solvatación continúa, con la finalidad de generar un volumen que contiene los valores del potencial electrostático para cada faseolina.

4.3.5. Construcción de modelos prospectivos con muestras propias.

En esta etapa se construyeron modelos por homología de faseolinas de las muestras propias correspondientes a las variedades coscorrón, tórtola, negro argel y enriqueta. Ejecutando la misma metodología para cada uno de los modelos (Figura 4.6). Es importante detallar que las secuencias se obtuvieron a partir de *RNAseq* completo. Por lo tanto, la construcción corresponde a sólo una fracción modelada de cada monómero de faseolina.

En primer lugar, las secuencias de ARN de cada tipo de poroto estudiado fue traducida en el servidor *web Augustus* (<https://bioinf.uni-greifswald.de/augustus/references>), el cual predice genes *ab initio* en secuencias genómicas de eucariontes, basándose en un modelo oculto de *Markov* (Stanke et al., 2004).

En segundo lugar, cada secuencia de aminoácidos fue consultada en *BLAST* para realizar un alineamiento con las secuencias alfa y beta utilizadas para la construcción de los trímeros, con la finalidad de que el mejor resultado fuera usado como templado. Una vez obtenidos los alineamientos, en el *software Maestro* utilizando la herramienta *advanced homology modeling*, se cargó la secuencia aminoacídica en formato *fasta* y el trímero con el monómero alfa o beta mejor evaluado como templado construyendo el modelo por homología correspondiente a un fragmento de faseolina de cada tipo de poroto, respectivamente.

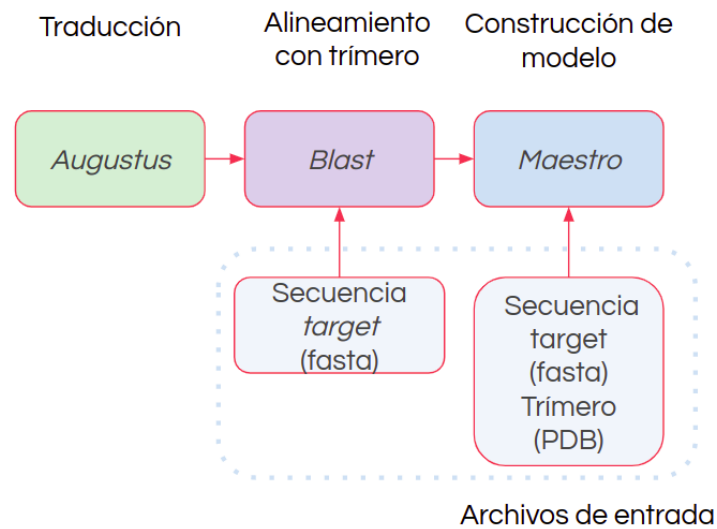


Figura 4.6: Metodología general para la construcción de modelos por homología con muestras propias.

5. RESULTADOS Y DISCUSIÓN

5.1. Análisis de expresión genética de transcritos de faseolinas presentes en *Phaseolus vulgaris L.*, sobre datos de *RNAseq*.

Los niveles de expresión van desde un color rojo, amarillo y blanco indicando mayor a menor expresión determinados por el valor de FPKM. El eje X corresponde a la posición del genoma en la que se encuentra el gen analizado.

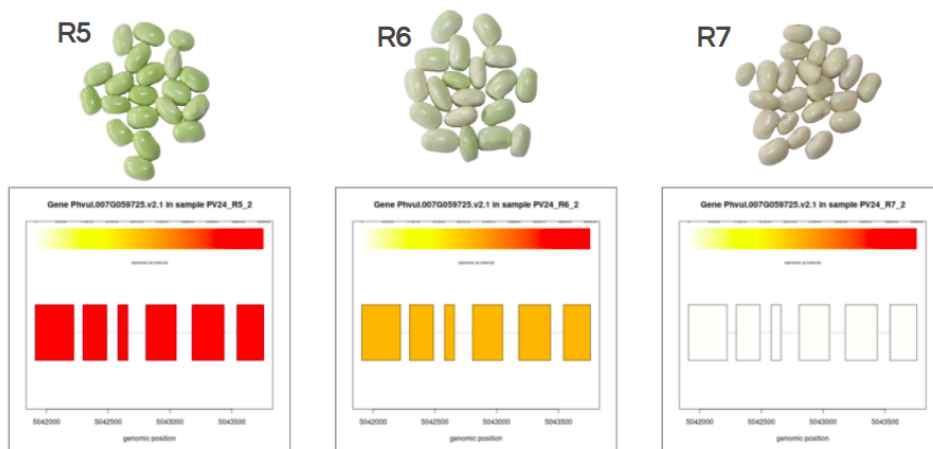


Figura 5.1: Análisis de expresión diferencial con respecto al gen Phvul.007G059725.v2.1 que codifica a faseolina en las muestras de poroto coscorrón.

En la figura 5.1, se muestra el análisis de expresión diferencial con respecto al gen que codifica a faseolina en las muestras de poroto coscorrón utilizando la réplica biológica 2 para cada estado. En él se puede visualizar que el estado R5 presenta una mayor expresión con respecto al estado R6, disminuyendo completamente en el estado R7.

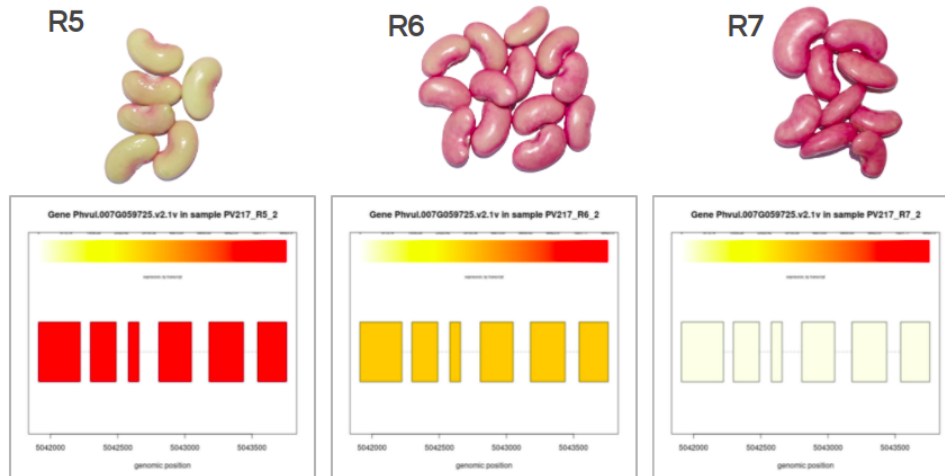


Figura 5.2: Análisis de expresión diferencial con respecto al gen *Phvul.007G059725.v2.1* que codifica a faseolina en las muestras de poroto enriqueta.

En la figura 5.2 se muestra el análisis de expresión diferencial con respecto al gen que codifica a faseolina en las muestras de poroto enriqueta utilizando la réplica biológica 2 para cada estado. En este caso se puede visualizar que el estado R5 presenta una mayor expresión con respecto al estado R6, disminuyendo fuertemente en el estado R7.

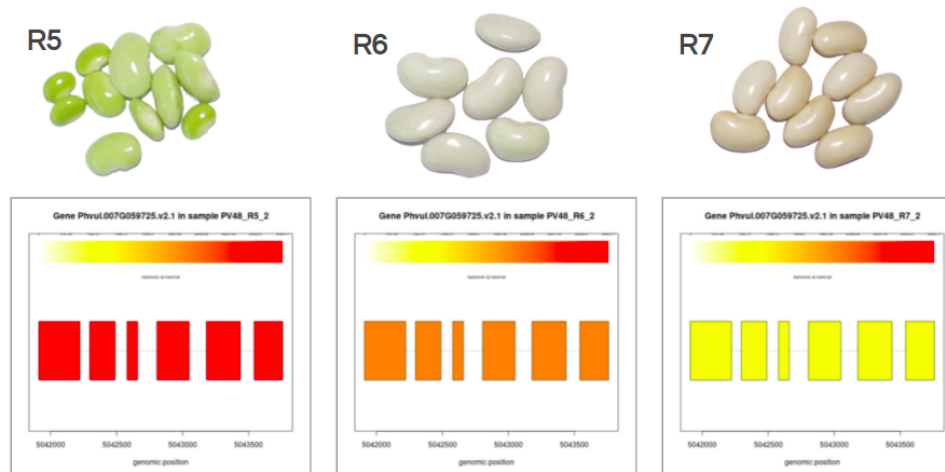


Figura 5.3: Análisis de expresión diferencial con respecto al gen *Phvul.007G059725.v2.1* que codifica a faseolina en las muestras de poroto tórtola.

En la figura 5.3 se muestra el análisis de expresión diferencial con respecto al gen que codifica a faseolina en las muestras de poroto tórtola utilizando la réplica biológica 2 para

cada estado. En la imagen se muestra que el estado R5 presenta una mayor expresión con respecto al estado R6, disminuyendo levemente en el estado R7 frente a R6.

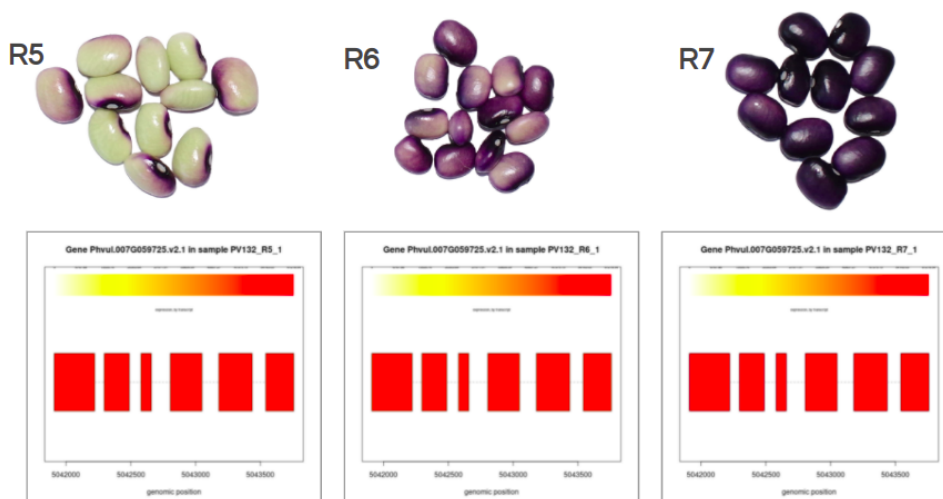


Figura 5.4: Análisis de expresión diferencial con respecto al gen Phvul.007G059725.v2.1 que codifica a faseolina en las muestras de poroto negro argel.

En la figura 5.4 se muestra el análisis de expresión diferencial con respecto al gen que codifica a faseolina en las muestras de poroto negro argel utilizando la réplica biológica 1 para cada estado. En este caso se puede visualizar que en los tres estados se mantiene prácticamente constante el nivel de expresión.

Estos análisis DEG determinan las diferencias genotípicas con respecto a la faseolina entre los tres estados de maduración para cada tipo de poroto. En los casos de poroto coscorrón, enriqueta y tórtola, la acumulación del gen que codifica a faseolina va disminuyendo. Mientras que en negro argel la acumulación no disminuye, se mantiene constante, pudiendo ser necesario el análisis a estados de grano más avanzados para determinar si permanece constante o disminuye.

La metodología utilizada, ha permitido obtener las abundancias específicas de las características estudiadas, poniendo a disposición directa tablas de expresión característica por muestra. Esto facilita la aplicación de pruebas de expresión diferencial alternativas o la evaluación de otras propiedades estadísticas del ensamblaje en caso de ser requeridos en análisis posteriores.

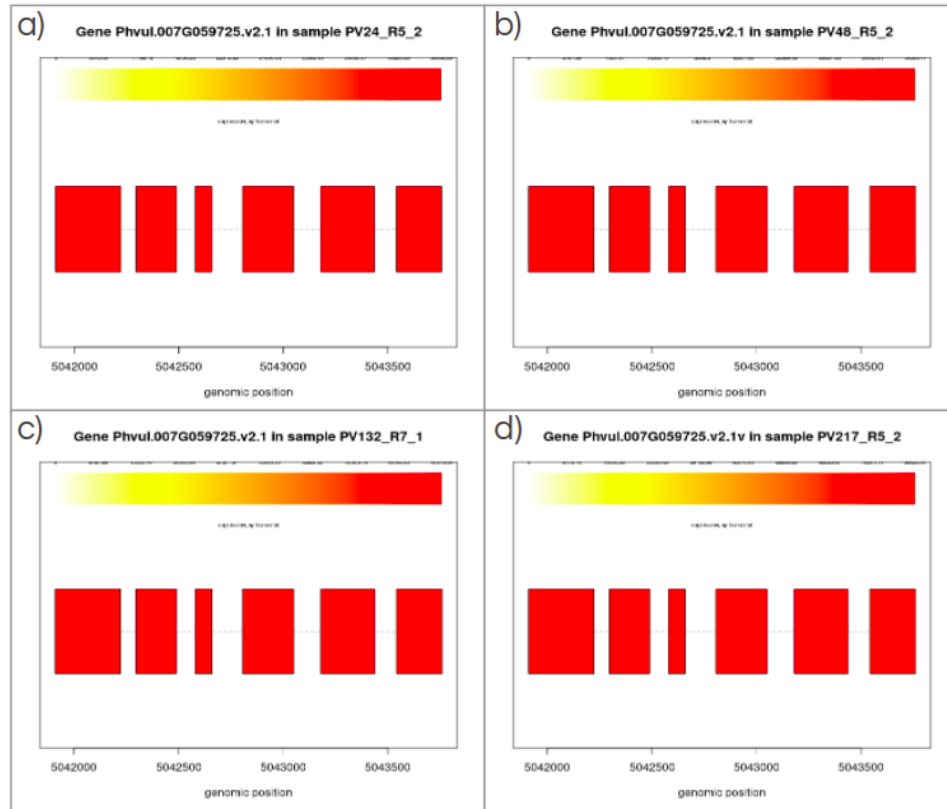


Figura 5.5: Réplicas biológicas escogidas que poseen mayor nivel de expresión con respecto al gen Phvul.007G059725.v2.1 que codifica a faseolina. A) Coscorrón PV24_R5_2. B) Tórtola PV48_R5_2. C) Negro argel PV132_R7_1. D) Enriqueta PV217_R5_2.

Por lo tanto a partir de las réplicas biológicas que presentan mayor nivel de expresión con respecto al gen Phvul.007G059725.v2.1 que codifica a faseolina, se ha escogido una para cada tipo de poroto, con la finalidad de seguir trabajando solo con estas 4 muestras (Figura 5.5). El análisis de expresión diferencial de las muestras con todas sus réplicas biológicas puede ser visualizado en los anexos (Figuras anexas 8.1,8.2,8.3,8.4).

5.2. Relaciones evolutivas de las secuencias de faseolinas presentes en *Phaseolus vulgaris* L. mediante un análisis filogenético.

Tabla 5.1: Promedio de valores obtenidos en alineamiento local de secuencia reportada con *RNAseqs* sin procesar de todas las muestras de las variedades de porotos.

Variedad	Código	Puntaje	Valor E	Identidad	Gaps	Número de secuencias	Coincidencias
Coscorrón	PV24_R5_2	246	4e-62	100 %	0 %	38,222,127	485
Tórtola	PV48_R5.2	246	4e-62	100 %	0 %	39,762,186	598
Negro argel	PV132_R7.1	246	4e-62	100 %	0 %	39,167,585	503
Enriqueta	PV217_R5.2	246	5e-62	100 %	0 %	47,773,224	519

Tabla 5.2: Resultado alineamiento local de secuencia de ARN de faseolina reportada con secuencias de ARN las muestras estudiadas.

Nombre	Puntaje máximo	Puntaje total	Query cover	Valor E	Porcentaje identidad	Largo secuencia
PV24_R5.2	250	1736	53 %	2e-69	100 %	931
PV48_R5.2	250	1237	35 %	2e-69	100 %	665
PV132_R7.1	246	800	33 %	2e-68	100 %	532
PV217_R5.2	250	2478	47 %	2e-69	100 %	1330

Los archivos en formato *BAM* de las muestras escogidas en el análisis de expresión diferencial de cada variedad de poroto fueron convertidos a formato *fasta*, para luego realizar un alineamiento local en *BLAST* y así obtener secuencias coincidentes a partir de una secuencia de referencia consultada. En la tabla 5.1, se observa un promedio de los valores obtenidos luego de realizar los alineamientos locales. El alineamiento con cada variedad de poroto obtuvo cerca de 500 coincidencias en promedio. Como las cuatro variedades de poroto provienen desde *RNAseq* sin procesar que contienen secuencias de largos pequeños, los alineamientos han resultado con valores similares y con resultados significativos en todos los casos, ya que al comparar entre la secuencia de consulta y los archivos *fasta* de los *RNAseq* de cada variedad de poroto, se encontraron fragmentos pertenecientes a la secuencia de consulta. Posteriormente, estas secuencias coincidentes pasaron por un proceso de agrupamiento con el programa *CD-hit* para obtener las secuencias correspondientes al gen que codifica a faseolina.

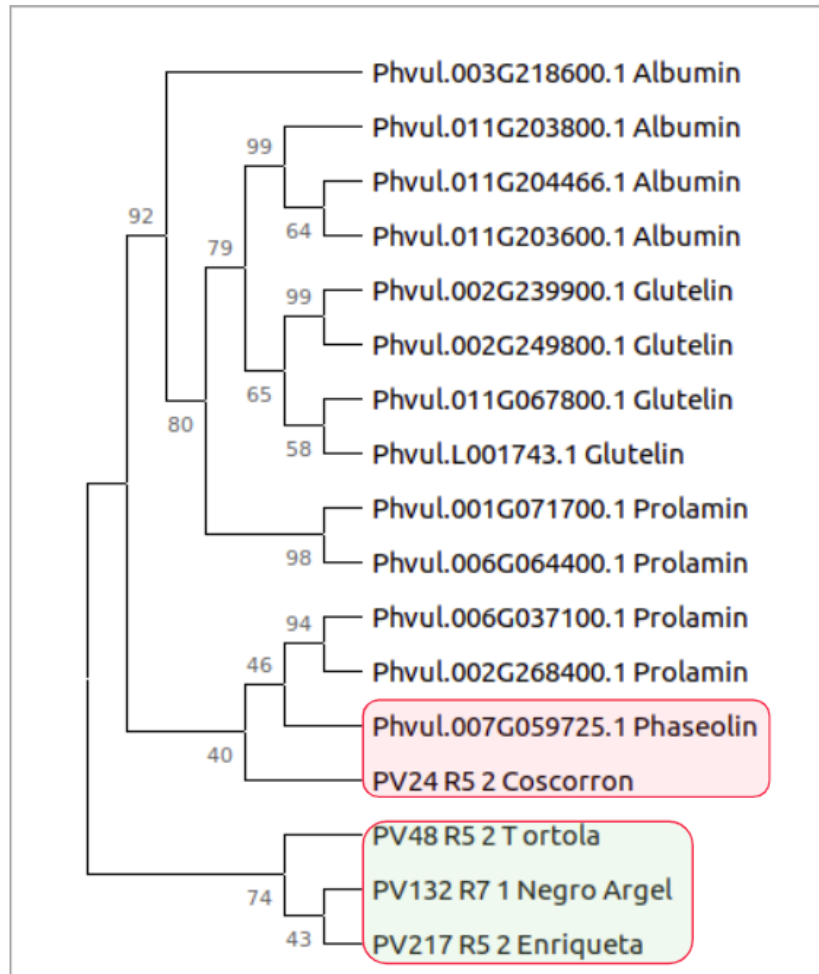


Figura 5.6: Árbol filogenético de secuencias de ARN que codifican a diversas proteínas de *Phaseolus vulgaris* L.

Una vez que se obtuvieron las secuencias de ARN que codifican al gen de faseolina para las muestras utilizadas de las cuatro variedades de poroto, se realizó un alineamiento múltiple de secuencias para poder realizar el análisis filogenético. Todas las accesiones de las secuencias incorporadas en este análisis fueron mostradas en la tabla 4.2. En la figura 5.6, se visualiza el árbol filogenético con las secuencias de ARN que codifican a albúminas, prolaminas, glutelinas y faseolinas. Las secuencias de ARN se agrupan de acuerdo a su clasificación. Mientras que las secuencias estudiadas se agrupan en conjunto a excepción de coscorrón PV24_R5_2 que se encuentra agrupada con la faseolina integrada. Esto podría deberse al porcentaje de cobertura que presenta coscorrón PV24_R5_2 frente a la faseolina consultada (Tabla 5.2).

En cuanto a los valores de confianza obtenidos el clado superior obtiene un valor de 92 que quiere decir que el 92% de todas las réplicas construidas mantiene la misma formación

desde el árbol original. Mientras que las muestras propias solo el 74 % de todas las réplicas construidas mantienen la formación del árbol original.

Este árbol filogenético entrega una estimación de las relaciones entre las secuencias de ARN que codifican a las diversas proteínas de *Phaseolus vulgaris L* y las secuencias de ARN desde *RNAseq* completo de las muestras propias. Esto da una primera aproximación a lo que sería la relación evolutiva de las proteínas y las muestras propias en caso de obtener el ARN aislado de las faseolinas de coscorrón, tórtola, enriqueta y negro argel.

5.3. Construcción modelos por homología de las faseolinas de las variedades locales de la raza Chile, y caracterización estructural mediante herramientas bioinformáticas.

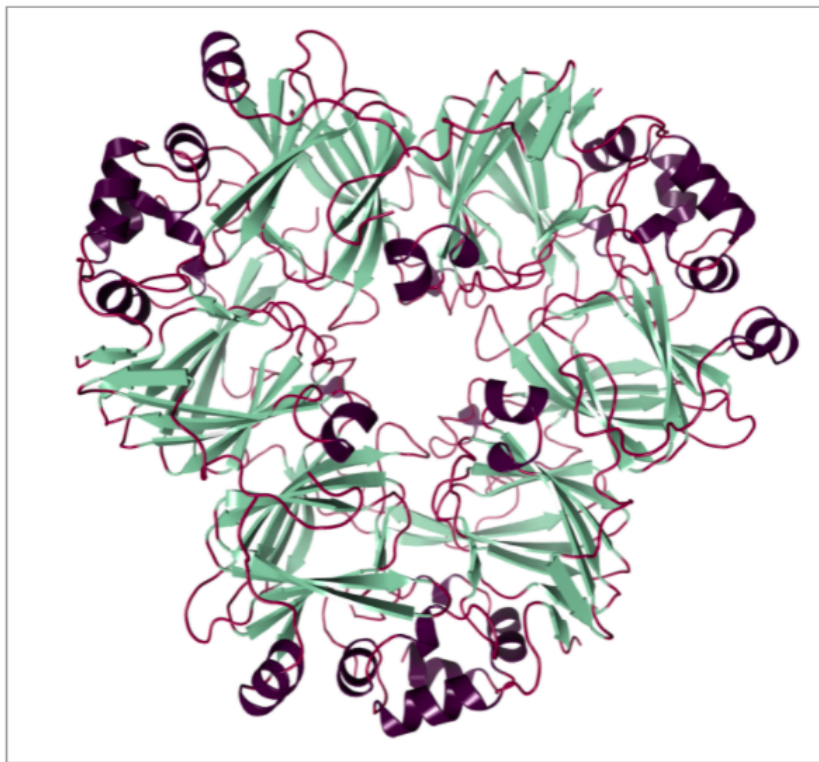


Figura 5.7: Estructura cristalina de *Adzuki Bean 7S Globulin-1* (PDB 2EA7).

Tabla 5.3: Resultado búsqueda de templado en *Blastp* para secuencias alfa y beta.

Secuencia	Nombre científico	Puntaje máximo	Puntaje total	Query Cover	Valor E	Porcentaje identidad	Largo secuencia
Alfa	<i>Vigna angularis</i>	440	440	91 %	1e-152	57.98 %	434
Beta	<i>Vigna angularis</i>	417	417	91 %	9e-144	58,21 %	434

La estructura del templado escogido para modelar ambos monómeros, corresponde a una globulina 7S de poroto de la especie *Vigna angularis* (Figura 5.7). Los resultados obtenidos de las secuencias alfa y beta con el templado luego de realizar el alineamiento local en *BLAST* se visualizan en la tabla 5.3. Para ambos casos el templado escogido posee un valor E que indica que la coincidencia ha sido significativa, además de un porcentaje de identidad mayor al 30 % y un query cover por sobre el 90 %.

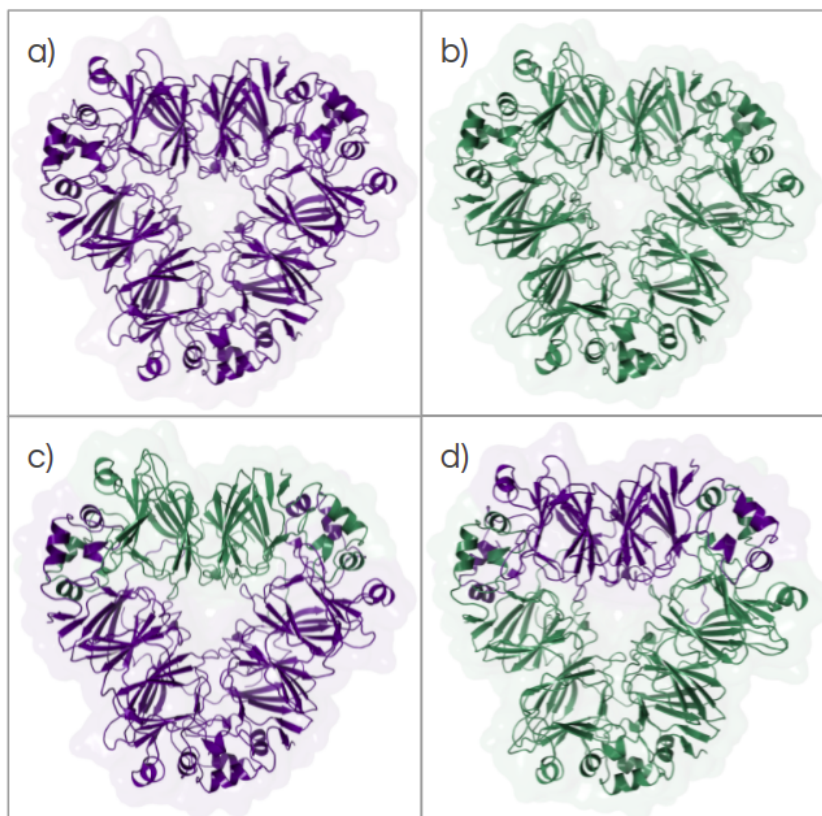


Figura 5.8: Trímeros ensamblados a partir de los monómeros alfa y beta modelados. En color morado se muestran las subunidades alfa, mientras que el color verde muestra las subunidades beta. A) Trímero conformado por tres subunidades alfa. B) Trímero conformado por tres subunidades beta. C) Trímero conformado por dos subunidades alfa y una subunidad beta. D) Trímero conformado por dos subunidades beta y una subunidad alfa.

Después de elegir el templado, se procedió a construir los modelos por homología de los monómeros alfa y beta de *Phaseolus vulgaris* L., donde una vez construidos fueron ensamblados para formar los trímeros en sus cuatro conformaciones (Figura 5.8).

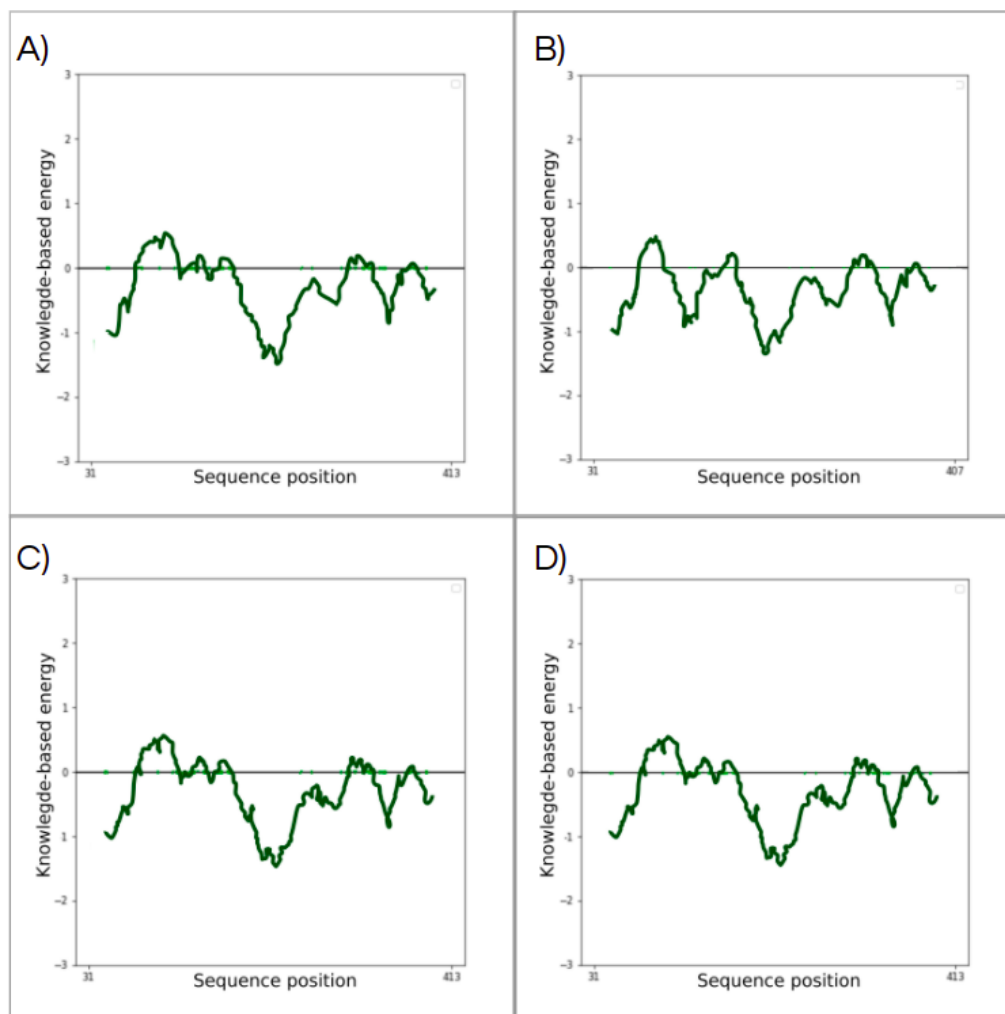


Figura 5.9: Evaluación en *ProSA* de los trímeros ensamblados antes de equilibrio. A) Trímero conformado por tres subunidades alfa. B) Trímero conformado por tres subunidades beta. C) Trímero conformado por dos subunidades alfa y una subunidad beta. D) Trímero conformado por dos subunidades beta y una subunidad alfa.

Posteriormente al ensamblado de los cuatro trímeros se realizó una evaluación de estos en dos diferentes servidores. La primera evaluación se hizo en *ProSA*, donde el resultado más significativo que entrega está dado por gráficos de energía. En la figura 5.9, se muestra la calidad de cada modelo local al representar las energías en función de la posición de la secuencia de aminoácidos. El color verde oscuro se toma como representativo. Por lo tanto, los valores positivos corresponden a partes problemáticas o erróneas de la

estructura de entrada. En este caso, los cuatro trímeros presentan valores positivos en algunos aminoácidos a lo largo de la secuencia.

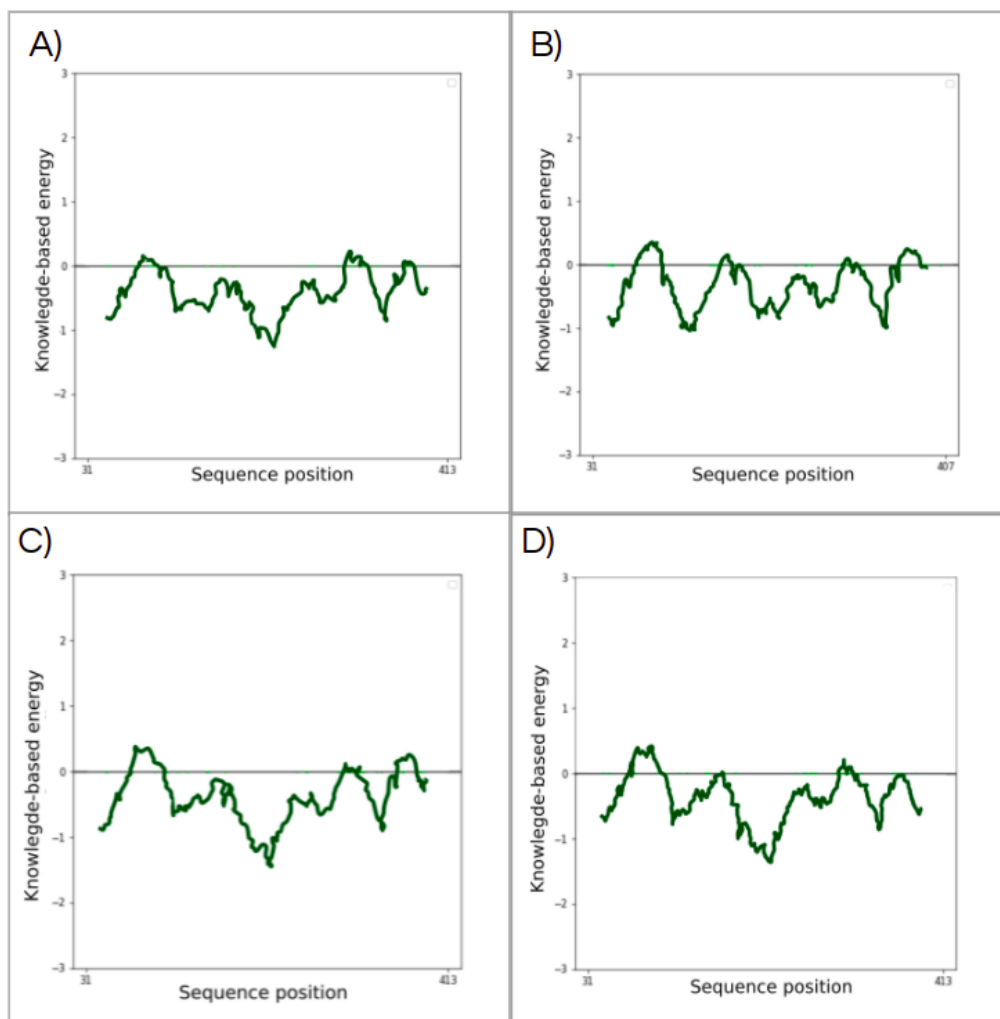


Figura 5.10: Evaluación en *ProSA* de los trímeros ensamblados después de equilibrio. A) Trímero conformado por tres subunidades alfa. B) Trímero conformado por tres subunidades beta. C) Trímero conformado por dos subunidades alfa y una subunidad beta. D) Trímero conformado por dos subunidades beta y una subunidad alfa.

Después de terminar el equilibrio en cada sistema con sus respectivos trímeros, se volvieron a realizar las evaluaciones de estos modelos construidos. En la figura 5.10, se muestra la calidad de cada modelo local al representar las energías en función de la posición de la secuencia de aminoácidos. Al igual que en la figura 5.9, se toma la curva de color verde como representativa. En esta ocasión, se puede visualizar que en los cuatro casos han disminuido muy poco los valores positivos, esto quiere decir que han disminuido las partes problemáticas o erróneas de las estructuras consultadas.

Tabla 5.4: Evaluación de la calidad de los trímeros antes y después del equilibrio.

	Antes de equilibrio		Después de equilibrio	
	<i>LGscore</i>	<i>MaxSub</i>	<i>LGscore</i>	<i>MaxSub</i>
Trímero alfa alfa alfa	4.972	0.113	4.972	0.113
Trímero beta beta beta	4.972	0.113	4.972	0.113
Trímero alfa alfa beta	4.972	0.113	4.972	0.113
Trímero beta beta alfa	4.972	0.113	4.972	0.113

La evaluación de los modelos también se hizo en *ProQ Server*. Este servidor entrega los valores *LGscore* y *MaxSub* comparando la estructura consultada con estructuras de referencias. *LGscore* detecta el segmento no continuo más significativo del modelo utilizando el logaritmo negativo del valor P, mientras que *MaxSub* se calcula a partir del mayor número de residuos que se pueden encontrar cuando todas las distancias entre el modelo y la estructura correcta son inferiores a 3,5 Å (Cristobal et al., 2001).

Los diferentes rangos de calidad están determinados de la siguiente manera:

- *LGscore* > 1.5 modelo bastante bueno
 - *LGscore* > 2.5 muy buen modelo
- *LGscore* > 4 modelo extremadamente bueno
 - *MaxSub* > 0.1 modelo bastante bueno
 - *MaxSub* > 0.5 muy buen modelo
- *MaxSub* > 0.8 modelo extremadamente bueno

Como *LGscore* y *MaxSub* detectan segmentos en común entre los modelos y una estructura objetivo correcta, los valores en ambos casos no han cambiado antes y después del equilibrio (Tabla 5.4). Esto podría deberse a que en el equilibrio se restringieron los *backbone* de cada trímero. En cuanto a los resultados obtenidos, todos los modelos obtuvieron un valor de *LGscore* sobre 4, lo que indica que los cuatro trímeros modelados son extremadamente buenos, en cuanto a los valores de *MaxSub*, los cuatro modelos obtuvieron un valor por sobre 0.1, los que los hace posicionarse en la categoría de modelos bastantes buenos.

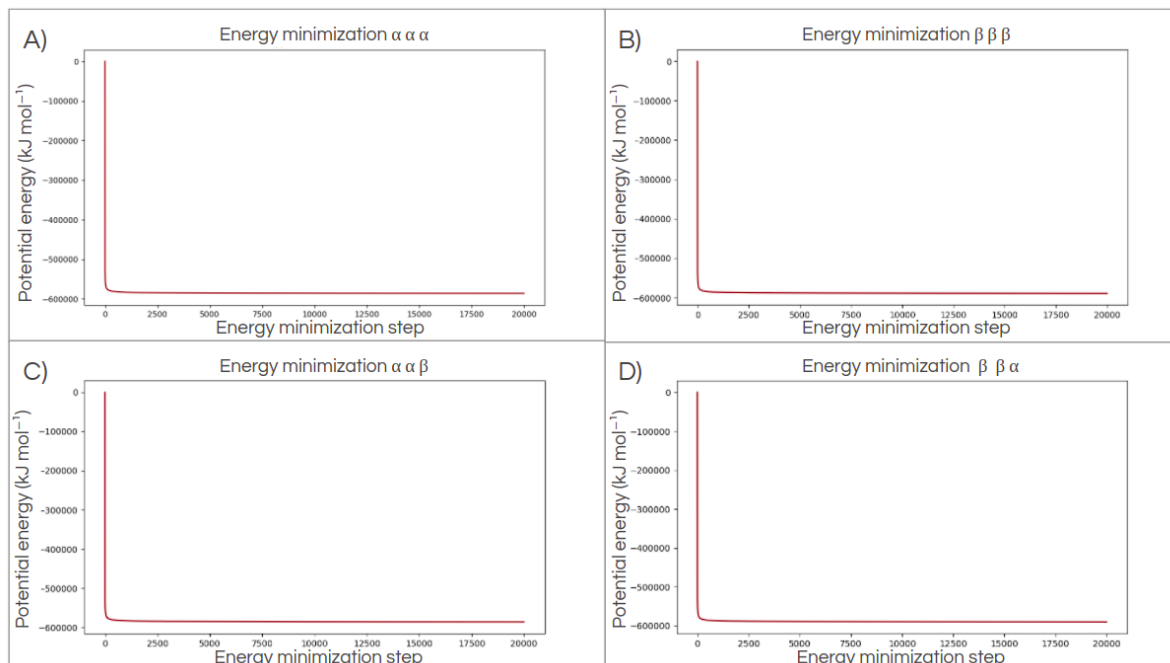


Figura 5.11: Evaluación de la energía potencial en minimización. A) Energía potencial para trímero conformado por tres subunidades alfa. B) Energía potencial para trímero conformado por tres subunidades beta. C) Energía potencial para trímero conformado por dos subunidades alfa y una subunidad beta. D) Energía potencial para trímero conformado por dos subunidades beta y una subunidad alfa.

Uno de los factores muy importantes a evaluar para determinar si la minimización de energía fue exitosa, es verificar si la energía potencial es negativa para la proteína minimizada. Es por esto, que se ha graficado la energía potencial de cada sistema luego de terminar la minimización de energía con el objetivo de verificar si esta es negativa llegando a un mínimo de energía. En la figura 5.11, es posible visualizar que los sistemas de cada trímero han logrado llegar a un mínimo de energía al obtener valores negativos de energía potencial durante una minimización de 20.000 pasos. Luego de verificar que los sistemas se encontraban en un mínimo de energía, se procedió a realizar la producción de dinámica molecular.

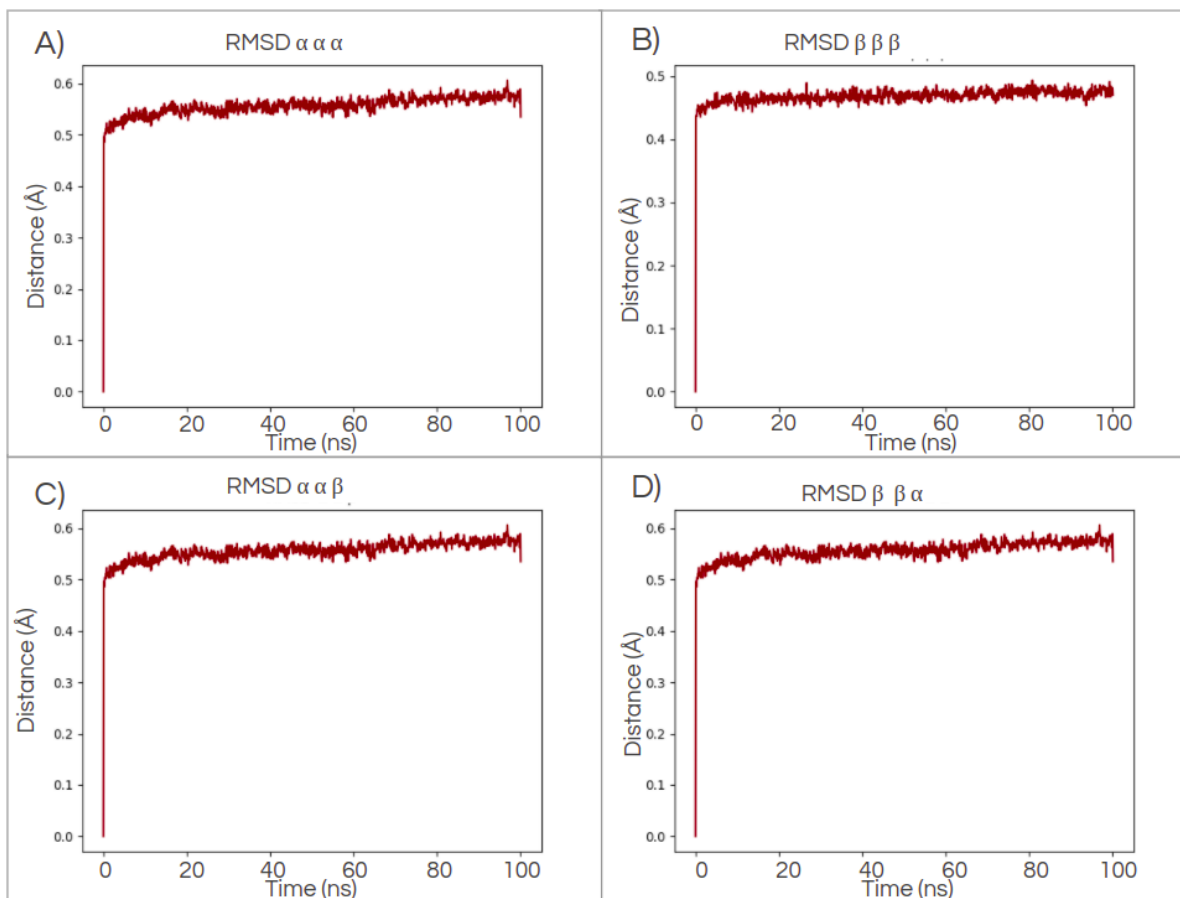


Figura 5.12: *RMSD* del backbone de las proteínas. A) *RMSD* para trímero conformado por tres subunidades alfa. B) *RMSD* para trímero conformado por tres subunidades beta. C) *RMSD* para trímero conformado por dos subunidades alfa y una subunidad beta. D) *RMSD* para trímero conformado por dos subunidades beta y una subunidad alfa.

Una vez realizada las simulaciones de dinámica molecular para cada sistema, se evaluó su estabilidad estructural mediante el cálculo de la raíz de la desviación cuadrática media (*RMSD*, del inglés *root-mean square deviation*) por medio de las herramientas de análisis del *software Maestro* (Schrödinger, 2021). Utilizando cálculos de *RMSD* sobre las simulaciones de dinámica molecular para los trímeros, se observó en la figura 5.12 que estos alcanzan una estabilización estructural luego de 20 ns de simulación molecular. El mayor intervalo de *RMSD*, se observa al inicio de las simulaciones, atribuyendo esto a la activación de los pistones de presión y temperatura de las dinámicas moleculares. Mientras que la variación promedio de *RMSD* durante las simulaciones de los cuatro trímeros es menor a 0.1 Å aproximadamente, determinando que las estructuras fueron estables a lo largo de la trayectoria.

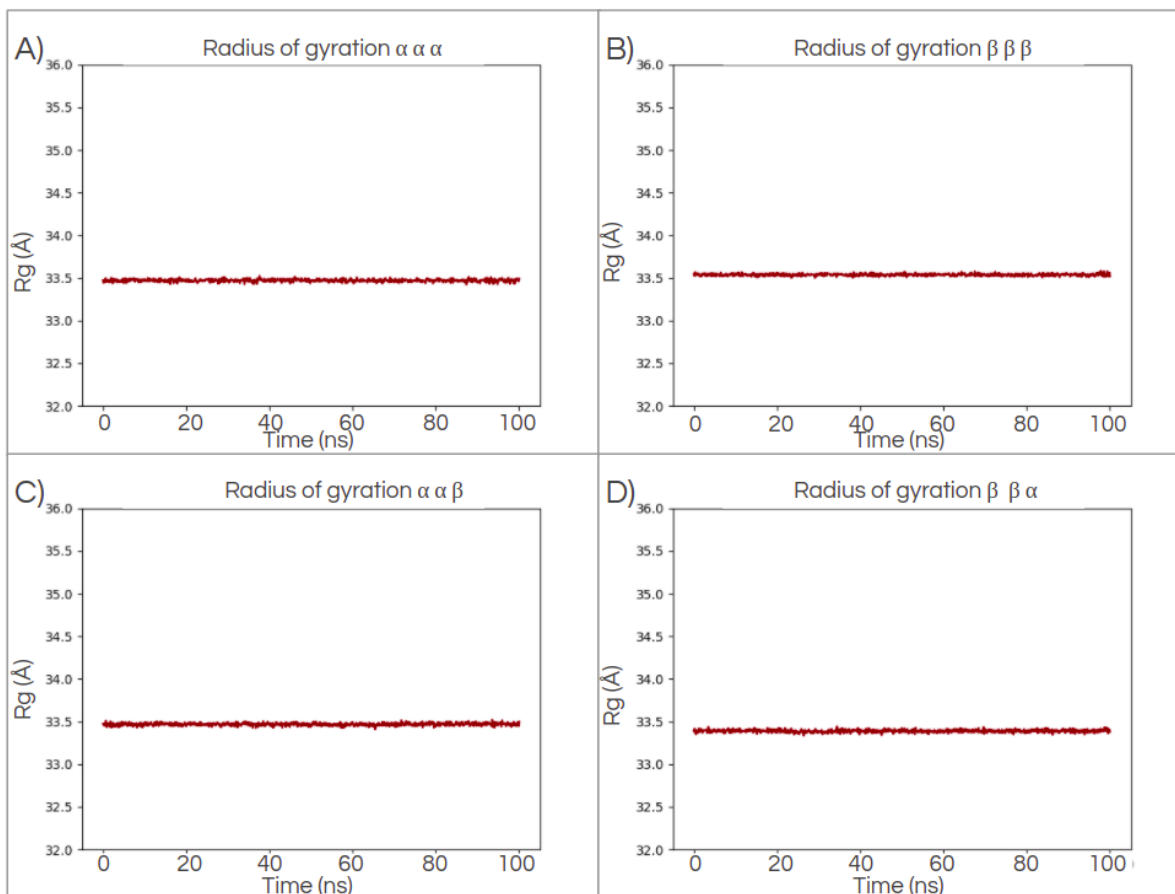


Figura 5.13: Radio de giro (Rg) del *backbone* de las proteínas. A) Rg para trímero conformado por tres subunidades alfa. B) Rg para trímero conformado por tres subunidades beta. C) Rg para trímero conformado por dos subunidades alfa y una subunidad beta. D) Rg para trímero conformado por dos subunidades beta y una subunidad alfa.

También se evaluó la estabilidad estructural de las simulaciones de dinámica molecular para cada sistema mediante el cálculo de radio de giro (Rg) por medio de las herramientas de análisis del *software Maestro* (Schrödinger, 2021). Utilizando cálculos de Rg sobre las simulaciones de dinámica molecular para los trímeros, se observó en la figura 5.13 que en todos los casos se mantiene un valor de Rg estable durante toda la simulación molecular, comprobando que las proteínas se mantienen plegadas. Por lo tanto, al realizar el análisis de *RMSD* y posterior Rg se determina que las estructuras fueron estables a lo largo de la trayectoria.

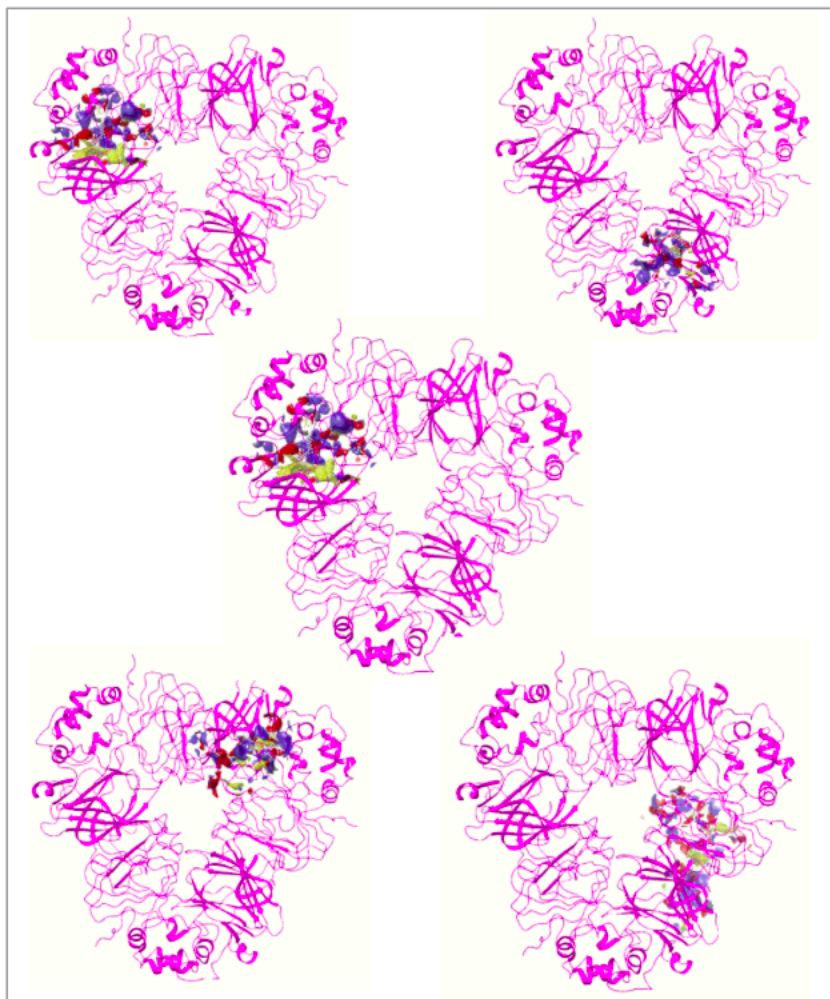


Figura 5.14: Análisis de *sitemap* para el trímero de tres subunidades alfa luego de la DM. En color amarillo se muestra la superficie hidrofóbica. La superficie de donantes de puentes de hidrógeno en color azul. La superficie de aceptores de puentes de hidrógeno en color rojo.

Posterior a la evaluación de la estabilidad estructural de las simulaciones de dinámica molecular de las proteínas, se seleccionaron los cuatro trímeros. Esto con la finalidad de realizar, con la herramienta *sitemap* de *Maestro* (Schrödinger, 2021) la búsqueda de sitios de unión probables para cada uno de ellos. La búsqueda comenzó con la localización de los sitios. Luego, se generaron los mapas de contorno. Finalizando con la entrega de un archivo que contiene varias propiedades para realizar una evaluación de los sitios sugeridos.

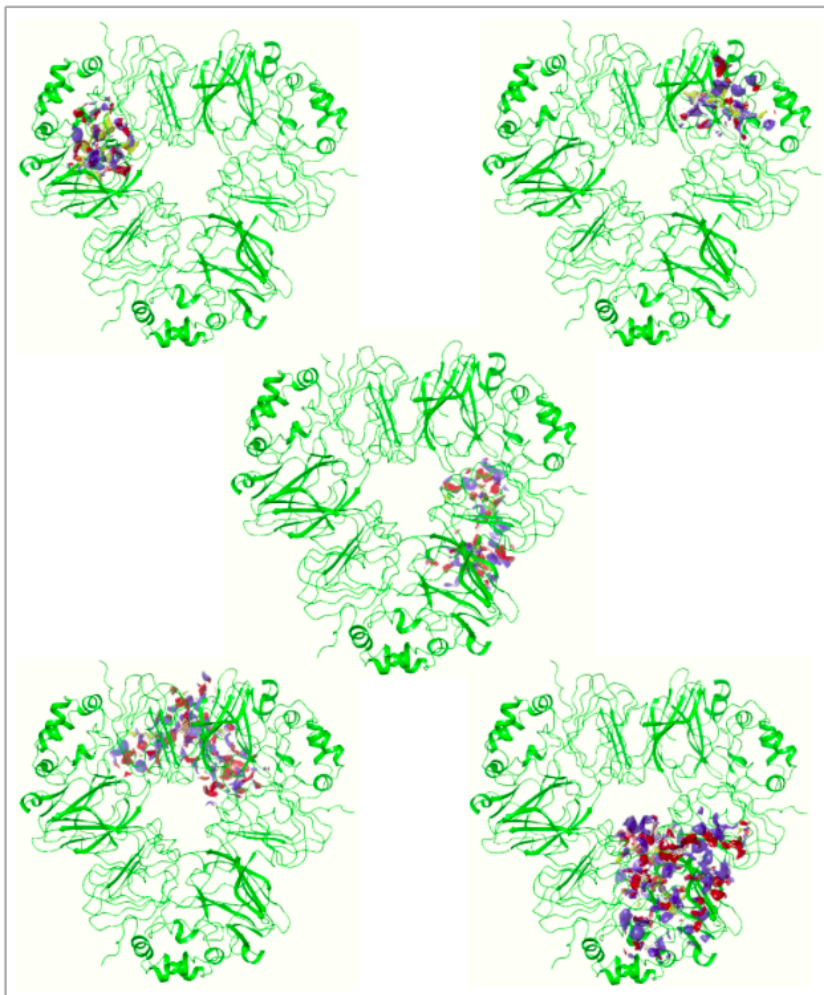


Figura 5.15: Análisis de *sitemap* para el trímero de tres subunidades beta luego de la DM. En color amarillo se muestra la superficie hidrofóbica. La superficie de donantes de puentes de hidrógeno en color azul. La superficie de aceptores de puentes de hidrógeno en color rojo.

Por lo tanto, se han obtenido cinco sitios de unión probables para cada uno de los trímeros (Figuras 5.14, 5.15, 5.16, 5.17). El *sitescore* indica que una puntuación superior a 1 sugiere un sitio particularmente prometedor. *Exposure* y *Enclosure* entregan una medida de qué tan abierto está el sitio al solvente. Para el valor de *Exposure*, cuanto menor sea la puntuación, será mejor, indicando un valor promedio de 0,49. En cuanto al valor de *Enclosure*, entre más alto mejor (Tabla 5.5). De acuerdo a esto, la mayoría de los sitios obtenidos poseen un buen puntaje.

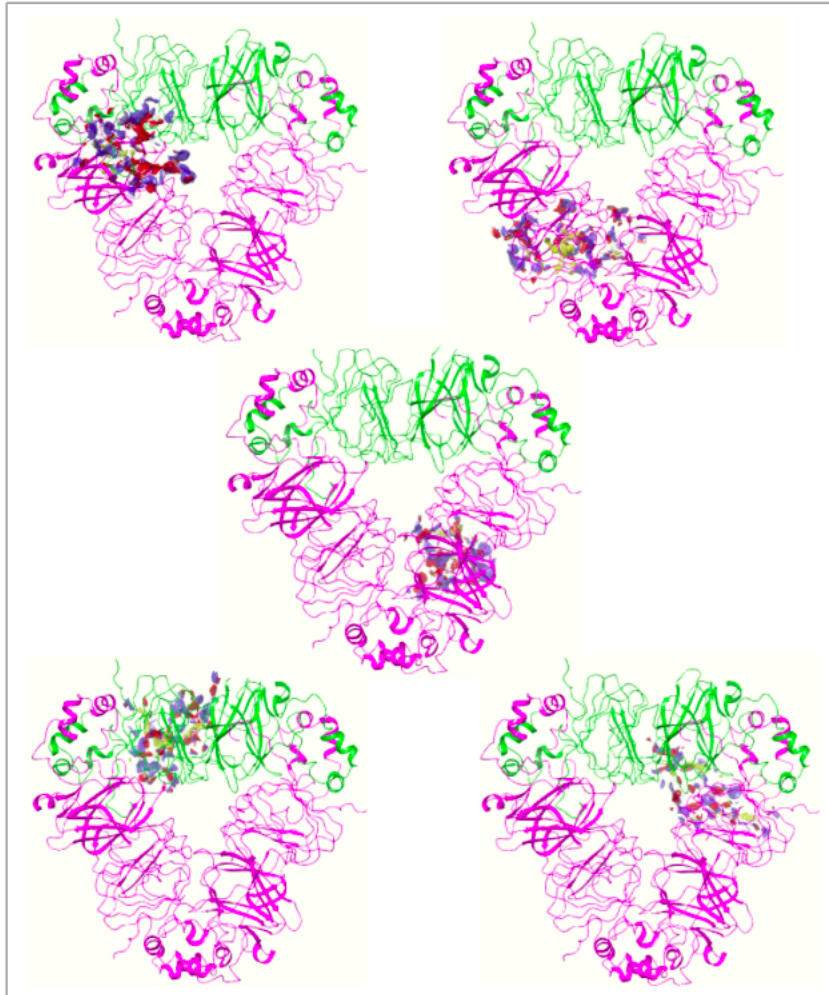


Figura 5.16: Análisis de *sitemap* para el trímero de dos subunidades alfa y una subunidad beta luego de la DM. En color amarillo se muestra la superficie hidrofóbica. La superficie de donantes de puentes de hidrógeno en color azul. La superficie de aceptores de puentes de hidrógeno en color rojo.

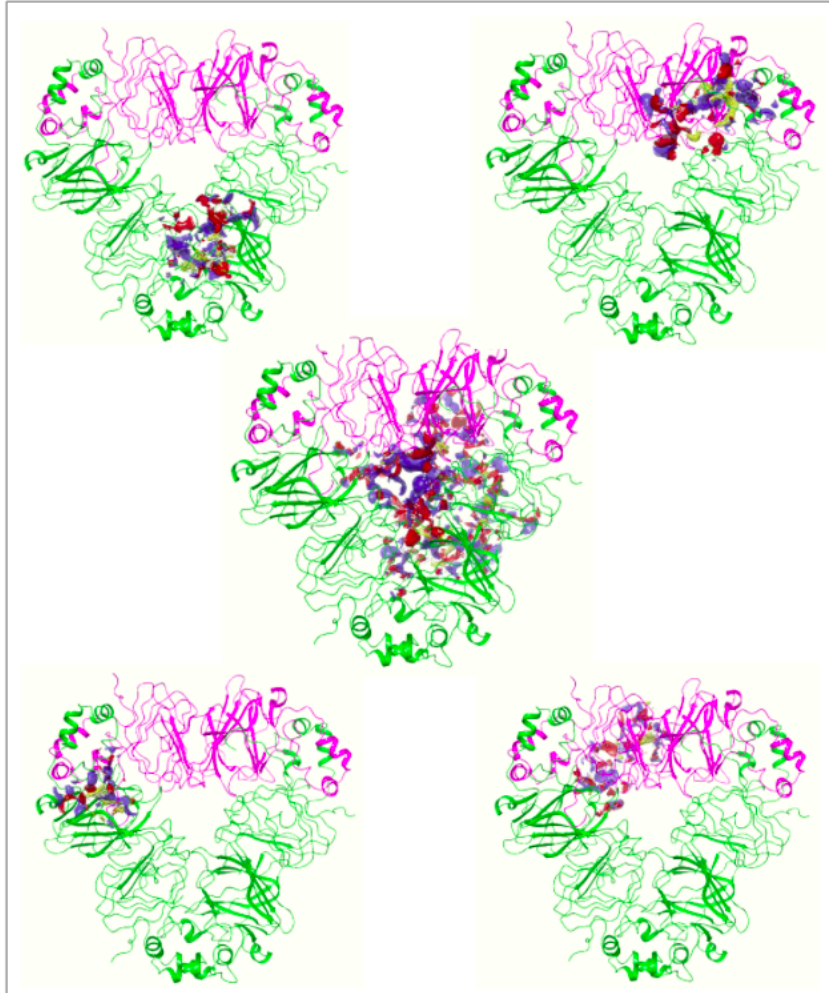


Figura 5.17: Análisis de *sitemap* para el trímero de dos subunidades beta y una subunidad alfa luego de la DM. En color amarillo se muestra la superficie hidrofóbica. La superficie de donantes de puentes de hidrógeno en color azul. La superficie de aceptores de puentes de hidrógeno en color rojo.

La herramienta utilizada ha permitido encontrar, visualizar y caracterizar sitios de unión a proteínas para cada uno de los trímeros analizados. Los sitios corresponden a un conjunto de puntos de sitio que se ubican en la superficie de los trímeros o cerca de ellos pudiendo abarcar la funcionalidad de un posible ligando.

Este análisis crea una primera aproximación a lo que podrían ser investigaciones futuras, como por ejemplo, mejorar la digestibilidad de las faseolinas del poroto presente en las variedades locales.

Tabla 5.5: Valores de evaluación obtenidos para los sitios de cada trímero.

Trímeros	Evaluaciones			
	Número sitio	Sitescore	Exposure	Enclosure
Trímero alfa alfa alfa	1	1.081	0.413	0.819
	2	1.023	0.545	0.733
	3	1.046	0.500	0.767
	4	1.009	0.509	0.711
	5	1.065	0.413	0.796
Trímero beta beta beta	1	0.993	0.585	0.687
	2	1.005	0.555	0.705
	3	1.123	0.419	0.883
	4	1.085	0.411	0.825
	5	1.026	0.565	0.565
Trímero alfa alfa beta	1	1.036	0.479	0.752
	2	1.026	0.568	0.730
	3	1.018	0.503	0.725
	4	1.022	0.514	0.731
	5	1.005	0.684	0.656
Trímero beta beta alfa	1	1.013	0.599	0.717
	2	1.074	0.404	0.808
	3	1.067	0.465	0.799
	4	0.994	0.582	0.689
	5	1.064	0.419	0.793

Un último análisis a las simulaciones de dinámica molecular fue realizado. Este corresponde a la superficie de potencial electrostático, el cual fue realizado en la herramienta *Electrostatic Potential Surfaces* de *Maestro* (Schrödinger, 2021). Los cuatro trímeros fueron representados como una superficie coloreada de acuerdo con el potencial electrostático (Figura 5.18, 5.19, 5.20, 5.21). El color rojo (potencial negativo) surge de un exceso de cargas negativas cerca de la superficie y el color azul (potencial positivo) ocurre cuando la superficie está cargada positivamente. Las regiones blancas corresponden a potenciales generalmente neutros.

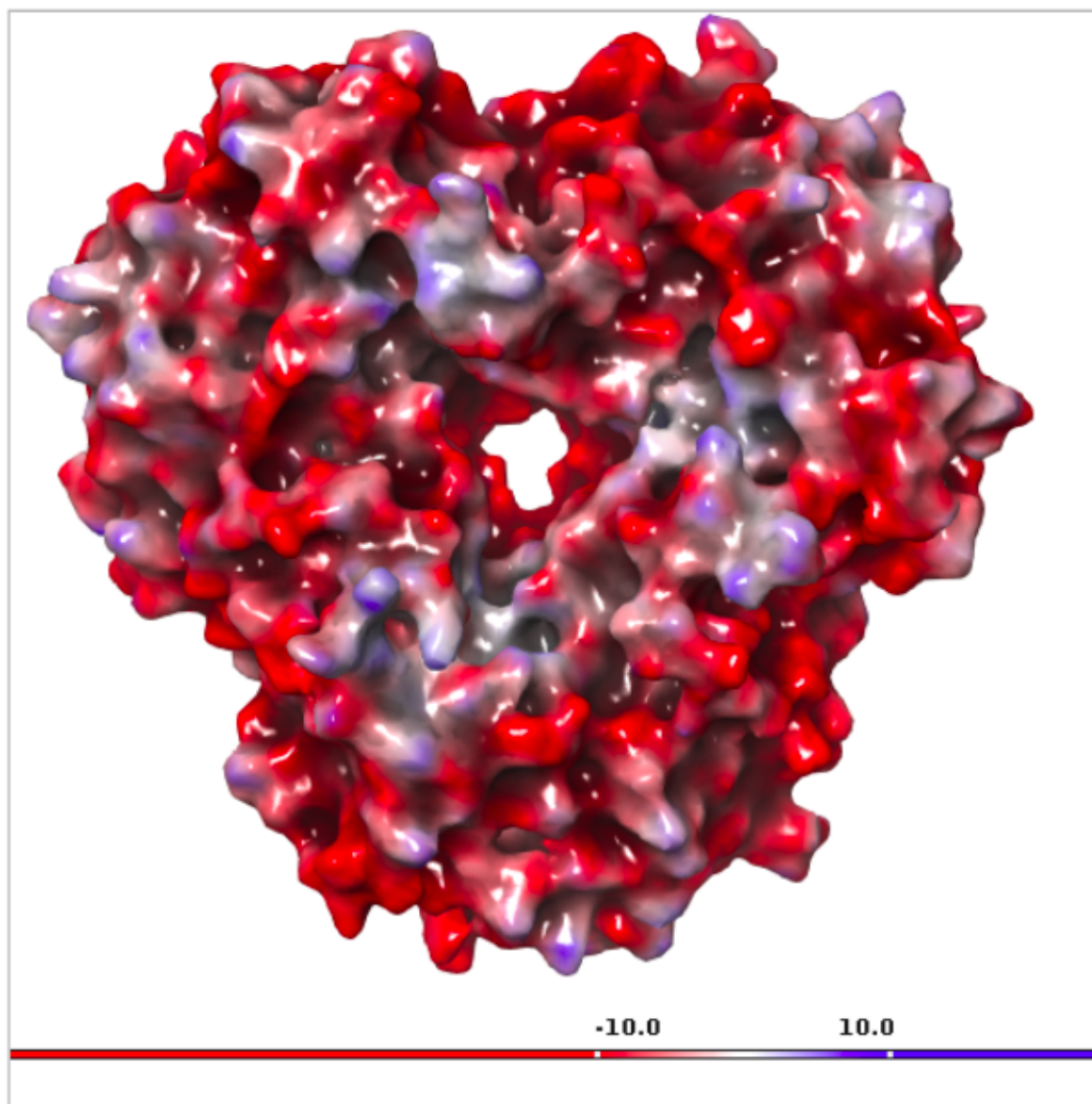


Figura 5.18: Análisis de potencial electrostático para el trímero compuesto por tres subunidades alfa. Color rojo indica potencial negativo, el color azul potencial positivo, color blanco potencial neutro.

Las secuencias aminoacídicas alfa y beta se diferencian en ocho aminoácidos, siendo interesante esta diferencia en dos posiciones. Al realizar un alineamiento de secuencias entre ambas (Figura anexa 8.5), la primera diferencia se obtiene el residuo 225 de beta correspondiente a lisina (aminoácido cargado positivamente), mientras que en alfa corresponde a glutamato (aminoácido con carga negativa) en la posición 231. La segunda diferencia interesante ocurre en el aminoácido glutamato 271 de beta y el aminoácido lisina 277 de alfa, siendo lo contrario a la primera observación. A partir de estas observaciones,

en todos ellos se puede apreciar que poseen mayoritariamente un potencial electrostático negativo.

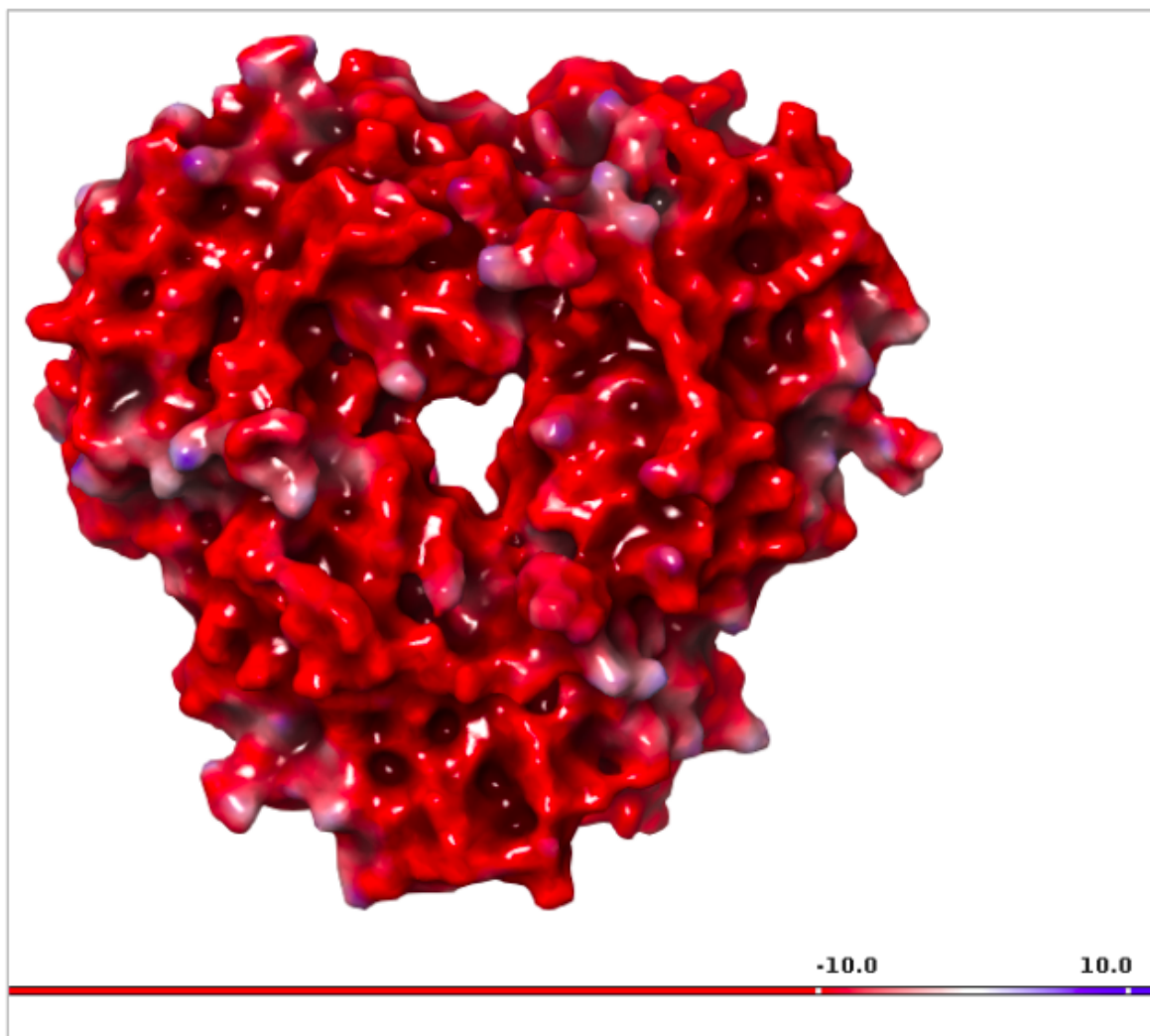


Figura 5.19: Análisis de potencial electrostático para el trímero compuesto por tres subunidades beta. Color rojo indica potencial negativo, el color azul potencial positivo, color blanco potencial neutro.

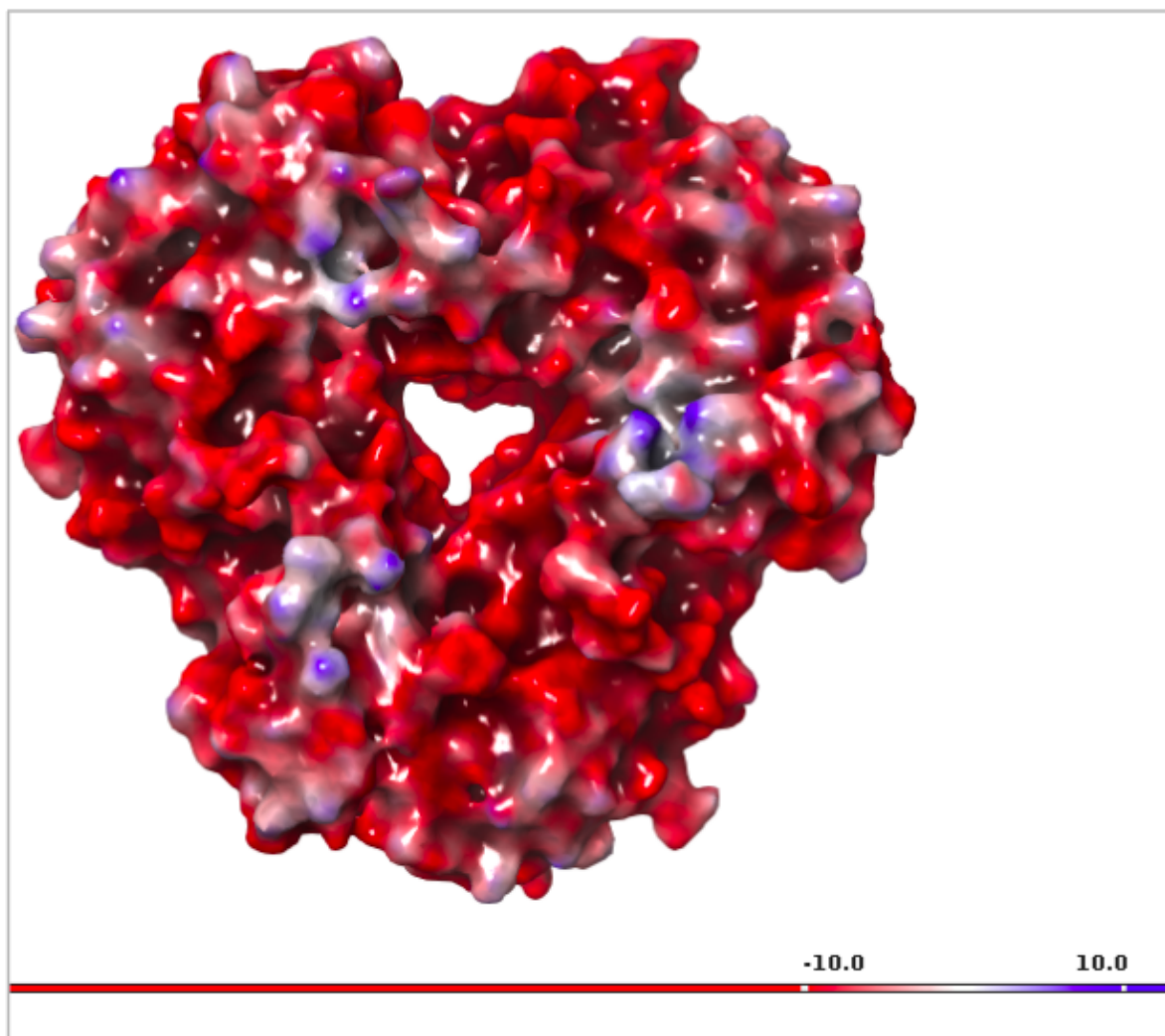


Figura 5.20: Análisis de potencial electrostático para el trímero compuesto por dos subunidades alfa y una subunidad beta. Color rojo indica potencial negativo, el color azul potencial positivo, color blanco potencial neutro.

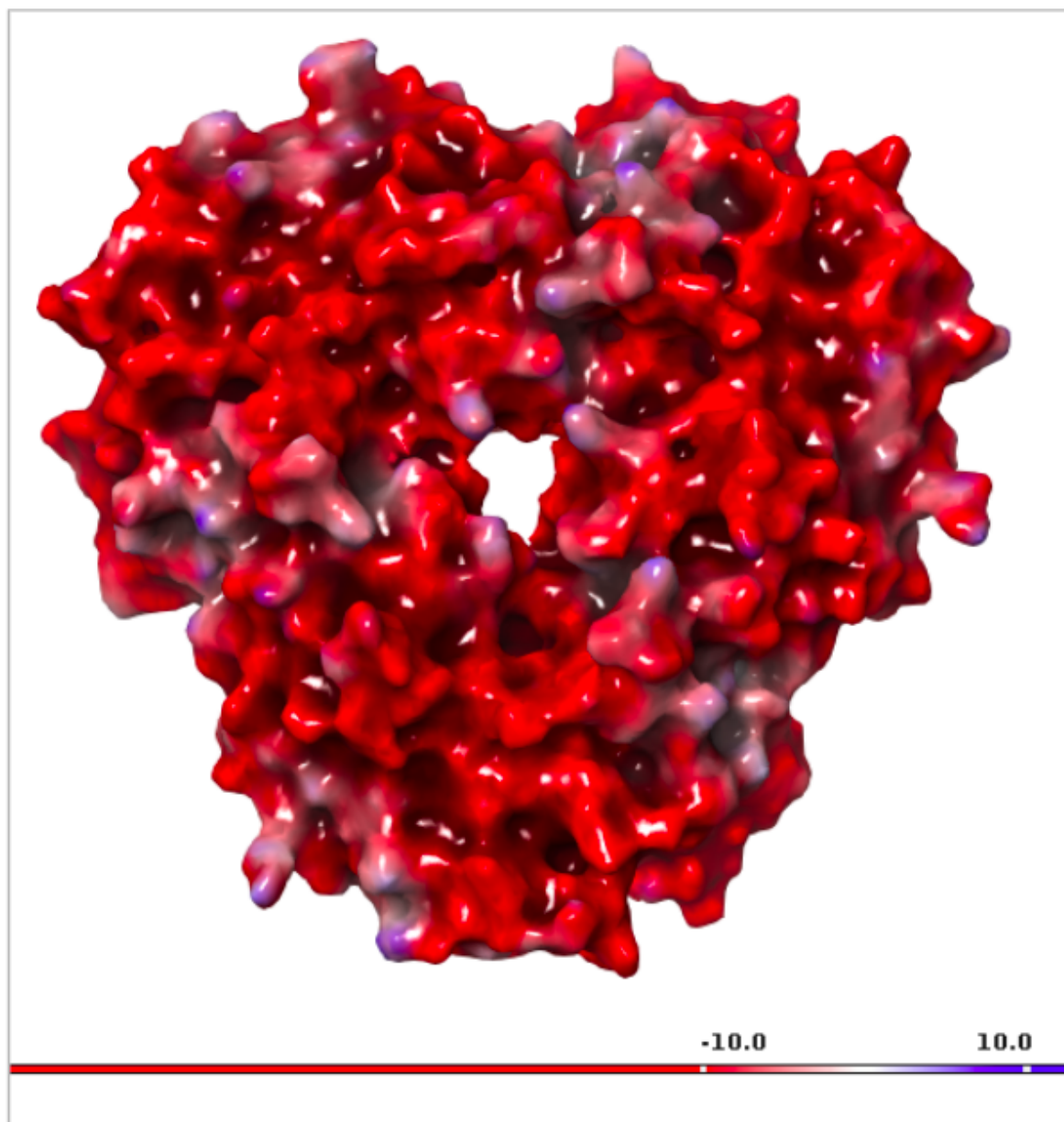


Figura 5.21: Análisis de potencial electrostático para el trímero compuesto por dos subunidades beta y una subunidad alfa. Color rojo indica potencial negativo, el color azul potencial positivo, color blanco potencial neutro.

Determinar que los trímeros poseen un potencial mayoritariamente negativo es de gran importante, siendo un complemento al análisis de *sitemap*. Pudiendo ser útil en la búsqueda de ligandos positivos para mejorar las propiedades de los poros de las variedades locales y extranjeras de esta investigación.

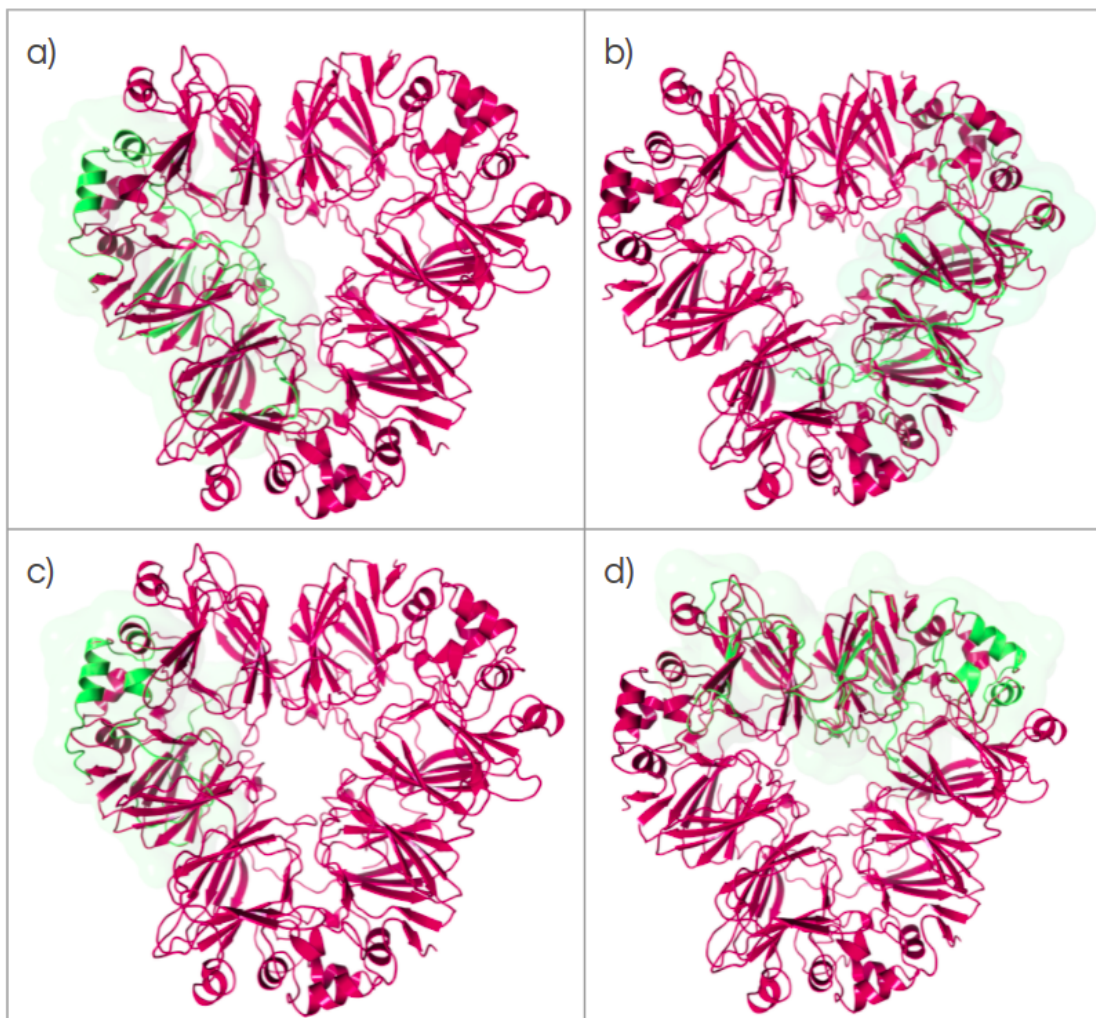


Figura 5.22: Alineamiento estructural de los modelos por homología de faseolina (magenta) y los modelos por homología de una fracción de faseolina a partir de muestras propias (verde). A) PV_24_R5_2 coscorrón B) PV48_R5_2 tórtola. C) PV132_R7.1 negro argel. D) PV217_R5_2 enriqueta.

Se construyeron modelos por homología de faseolinas de las muestras propias correspondientes a las variedades coscorrón, tórtola, negro argel y enriqueta. Los resultados de los modelos corresponden a una fracción proteica de faseolina, ya que las secuencias aminoacídicas de entrada provienen de *RNAseq* completos y no de *RNAseq* de la proteína de faseolina aislada. Para analizar las fracciones modeladas versus el modelo de proteína completa, se realizó un alineamiento estructural (Figura 5.20) con sus respectivos templates de los modelos por homología de faseolinas con secuencias ya reportadas.

En este caso, el template escogido para todas las muestras propias fue el trímero alfa alfa alfa, ya que al realizar el alineamiento con los monómeros alfa y beta, se presentaron

valores similares, pero alfa presenta mayor largo de secuencia y obtiene un puntaje máximo mayor con respecto al monómero beta (Tabla 5.5). Además, en los casos para PV24_R5_2, PV48_R5_2 y PV217_R5_2 obtiene mejor porcentaje de identidad.

Tabla 5.6: Alineamiento en *Blastp* de muestras propias con monómeros alfa y beta.

Nombre muestras	Monómero	Puntaje máximo	Puntaje total	Query Cover	Valor E	Porcentaje identidad	Largo secuencia
PV24_R5_2	<i>alfa</i>	165	189	55 %	4e-54	68,99 %	383
PV24_R5_2	<i>beta</i>	161	184	55 %	2e-52	66,67 %	377
PV48_R5_2	alfa	35,8	35,8	8 %	3e-7	100 %	383
PV48_R5_2	beta	35,8	35,8	8 %	3e-7	100 %	377
PV132_R7_1	alfa	22,7	22,7	29 %	0,002	53,85 %	383
PV132_R7_1	beta	22,7	22,7	29 %	0,001	53,85 %	377
PV217_R5_2	alfa	67	100	20 %	1e-17	90,62 %	383
PV217_R5_2	beta	63,5	97,8	20 %	1e-17	64,58 %	377

6. CONCLUSIONES

Contar con herramientas bioinformáticas rápidas, precisas y flexibles como *HISAT2*, *StringTie* y *Ballgown* demostraron estar al nivel en el momento que redujeron los datos de lectura de *RNAseq* sin procesar, obteniendo resultados comprensibles en poco tiempo. Estas herramientas permitieron alinear las lecturas con un genoma, ensamblar transcripciones incluyendo nuevas variantes de empalme, calcular la abundancia de estas transcripciones en cada muestra, compararlas para identificar genes y expresarlos diferencialmente. *Ballgown* fue útil al entregar una disponibilidad directa de tablas de expresión característica por muestra, lo que facilitó la aplicación de pruebas de expresión diferencial entre todas las réplicas biológicas de porotos analizadas, obteniendo una expresión diferencial concreta y visualmente atractiva en los tres estados de maduración de los granos de porotos. En tres casos (coscorrón, enriqueta y tórtola) se determinó que la mayor expresión se daba en el estadio de grano R5 utilizando la réplica biológica 2. Mientras que para la variedad de poroto negro argel, la mayor expresión se presentó en el estado de maduración del grano R1 utilizando la réplica biológica 1, a pesar de que en todas las replicas el nivel de expresión se mantiene prácticamente constante.

Con respecto al estudio evolutivo, se calculó y presentó un árbol filogenético, en el que se incorporaron secuencias clasificadas en diferentes tipos de proteínas de la especie *Phaseolus vulgaris* L. El método escogido permitió un agrupamiento iterativo basándose en el criterio de mínima evolución, prefiriendo la topología con la menor longitud total de rama en cada paso. Además, el *bootstrap* escogido de 5.000 replicaciones, obtuvo una proporción de filogenias replicadas que recuperaron un clado particular de la filogenia original con un alto nivel de confianza. El clado superior (Figura 5.6), se agrupó confiablemente, mientras que el clado inferior obtuvo un valor menos confiable debiendo ser por el bajo nivel de cobertura de las muestras propias con la secuencia de faseolina de referencia incorporada en el análisis evolutivo.

Por otro lado, la construcción de los modelos por homología de las faseolinas de las variedades locales de la raza Chile, y su caracterización estructural mediante herramientas bioinformáticas, no se llevó a cabo completamente. Esto se debe a que solo se contaba con *RNAseq* completos de las variedades de porotos estudiados y no *RNAseq* aislados de la proteína de interés. Para solucionar esta problemática, se decidió modelar los trímeros de faseolinas a partir de secuencias alfa y beta que ya habían sido reportadas. De esa manera, se consiguió ensamblar los trímeros en sus cuatro conformaciones, obteniendo buenos resultados. Los modelos fueron evaluados satisfactoriamente a excepción de algunos residuos que presentaban una energía positiva o errores en su conformación.

En cuanto a los cálculos de minimización de energía para los cuatro sistemas, la energía potencial fue negativa en todos los casos, procediendo a realizar el equilibrio y producción de dinámica molecular de manera exitosa. Las dinámicas moleculares de los cuatro sistemas fueron analizados, determinando que los trímeros poseen una estabilidad estructural a lo largo de las trayectorias de acuerdo al *RMSD* y *Rg*. También se descubrieron posibles sitios de unión en cada trímero, siendo un hecho relevante para las futuras investigaciones. El análisis de la superficie de potencial electrostático concluye que todos los trímeros poseen un potencial electrostático altamente negativo.

Finalmente, se construyeron los modelos por homología de las faseolinas de las variedades locales de la raza Chile y extranjeras de esta investigación, dando como resultado modelos que pertenecen a sólo una fracción de faseolina al contar solo con el *RNAseq* completo de las variedades estudiadas. A pesar de esto, con los resultados obtenidos ya se puede hacer una prospección de lo que serán las investigaciones a realizar posteriormente.

El conocimiento generado con respecto al análisis de expresión diferencial, relación evolutiva, construcción de modelos por homología y producción de dinámica molecular permiten obtener una mirada profunda a la gran diversidad de lo que es la faseolina de *Phaseolus vulgaris* L. en el endosperma. Dentro de las proyecciones de esta investigación podría incluirse el análisis de electroforesis para determinar los tipos de faseolinas presentes en el endosperma de las variedades locales de la raza Chile y extranjeras, posteriormente aislar las proteínas y secuenciarlas obteniendo un ARN de proteína aislada. De esta manera, realizar los modelos completos de las faseolinas y los análisis realizados en esta investigación. Ya que la faseolina ha sido involucrada en diversas actividades biológicas de gran importancia mencionadas con anterioridad.

7. BIBLIOGRAFÍA

1. Aier, I., Varadwaj, P. K., y Raj, U. (2016). Structural insights into conformational stability of both wild-type and mutant EZH2 receptor. *Scientific Reports*, 6(1). <https://doi.org/10.1038/srep34984>
2. Allen, D. J., y Lenné, J. M. (1998). *The pathology of food and pasture legumes*. CAB International.
3. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., y Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403-410. [https://doi.org/10.1016/s0022-2836\(05\)80360-2](https://doi.org/10.1016/s0022-2836(05)80360-2)
4. Baldi, G., y Salamini, F. (1973). Variability of essential amino acid content in seeds of 22 *Phaseolus* species. *Theoretical and Applied Genetics*, 43(2), 75-78. [doi:10.1007/bf00274960](https://doi.org/10.1007/bf00274960)
5. Bascur B., G., y Tay U., J. (2005). Colecta, Caracterización y Utilización de la Variabilidad Genética en Germoplasma Chileno de Poroto (*Phaseolus vulgaris* L.). *Agricultura Técnica*, 65(2). <https://doi.org/10.4067/s0365-28072005000200003>
6. Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., da Silva, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., Castro, L. G., . . . Teodoro, D. (2020). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49(D1), D480-D489. <https://doi.org/10.1093/nar/gkaa1100>
7. Berman, H. M. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1), 235-242. <https://doi.org/10.1093/nar/28.1.235>

8. Bishop, T. O. A. (2008). Protein homology modelling and its use in South Africa. *SciELO*. http://www.scielo.org.za/scielo.php?script=sci_arttext&pid=S0038-23532008000100001

9. Bitocchi, E., Rau, D., Bellucci, E., Rodriguez, M., Murgia, M. L., Gioia, T., . . . Papa, R. (2017). Beans (*Phaseolus* spp.) as a Model for Understanding Crop Evolution. *Frontiers in Plant Science*, 8. doi:10.3389/fpls.2017.00722

10. Carrasco-Castilla, J., Hernández-Álvarez, A. J., Jiménez-Martínez, C., Jacinto-Hernández, C., Alaiz, M., Girón-Calle, J., . . . Dávila-Ortiz, G. (2012b). Antioxidant and metal chelating activities of *Phaseolus vulgaris* L. var. Jamapa protein isolates, phaseolin and lectin hydrolysates. *Food Chemistry*, 131(4), 1157-1164. doi:10.1016/j.foodchem.2011.09.084

11. Carrasco-Castilla, J., Hernández-Álvarez, A. J., Jiménez-Martínez, C., Jacinto-Hernández, C., Alaiz, M., Girón-Calle, J., . . . Dávila-Ortiz, G. (2012a). Antioxidant and metal chelating activities of peptide fractions from phaseolin and bean protein hydrolysates. *Food Chemistry*, 135(3), 1789-1795. doi:10.1016/j.foodchem.2012.06.016

12. Chatzou, M., Magis, C., Chang, J. M., Kemena, C., Bussotti, G., Erb, I., y Notredame, C. (2015). Multiple sequence alignment modeling: methods and applications. *Briefings in Bioinformatics*, 17(6), 1009-1023. <https://doi.org/10.1093/bib/bbv099>

13. Choudhuri, S. (2014). Phylogenetic Analysis. *Bioinformatics for Beginners*, 209-218. <https://doi.org/10.1016/b978-0-12-410471-6.00009-8>

14. Cristobal, S., Zemla, A., Fischer, D., Rychlewski, L., y Elofsson, A. (2001). A study of quality measures for protein threading models. *BMC Bioinformatics*, 2(1), 5. <https://doi.org/10.1186/1471-2105-2-5>

15. Delfini, J., Moda-Cirino, V., dos Santos Neto, J., Ruas, P. M., SantÁna, G. C., Gepts, P., y Gonalves, L. S. A. (2021). Population structure, genetic diversity and genomic selection signatures among a Brazilian common bean germplasm. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-82437-4>

16. Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced

time and space complexity. *BMC Bioinformatics*, 5(1), 113. <https://doi.org/10.1186/1471-2105-5-113>

17. FAOSTAT. (2021). Estadísticas del poroto. <http://www.fao.org/faostat/en/#data/QC>
18. Frazee, A. C., Perteza, G., Jaffe, A. E., Langmead, B., Salzberg, S. L., y Leek, J. T. (2015). Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nature biotechnology*, 33(3), 243-246. <https://doi.org/10.1038/nbt.3172>
19. Gao, M., Zhong, A., Patel, N., Alur, C., y Vyas, D. (2017). High throughput RNA sequencing utility for diagnosis and prognosis in colon diseases. *World Journal of Gastroenterology*, 23(16), 2819. <https://doi.org/10.3748/wjg.v23.i16.2819>
20. García-Cordero, J., Martínez-Palma, N., Madrigal-Bujaidar, E., Jiménez-Martínez, C., Madrigal-Santillán, E., Morales-González, J., Paniagua-Pérez, R., y Álvarez-González, I. (2021). Phaseolin, a Protein from the Seed of *Phaseolus vulgaris*, Has Antioxidant, Antigenotoxic, and Chemopreventive Properties. *Nutrients*, 13(6), 1750. <https://doi.org/10.3390/nu13061750>
21. Ganesan, K., y Xu, B. (2017). Polyphenol-Rich Dry Common Beans (*Phaseolus vulgaris* L.) and Their Health Benefits. *International Journal of Molecular Sciences*, 18(11), 2331. doi:10.3390/ijms18112331
22. Gelpi, J., Hospital, A., Goni, R., y Orozco, M. (2015). Molecular dynamics simulations: advances and applications. *Advances and Applications in Bioinformatics and Chemistry*, 37. <https://doi.org/10.2147/aabc.s70333>
23. Gepts, P. (1988). Phaseolin as an Evolutionary Marker. *Genetic Resources of Phaseolus Beans Current Plant Science and Biotechnology in Agriculture*, 215-241. doi:10.1007/978-94-009-2786-5_11
24. Guevara-González, R. G., y Torres-Pacheco, I. (2006). Advances in agricultural and food biotechnology. *Research Signpost*.
25. Halgren, T. A. (2009). Identifying and Characterizing Binding Sites and Assessing Druggability. *Journal of Chemical Information and Modeling*, 49(2), 377-389.

<https://doi.org/10.1021/ci800324m>

26. Hall, B. G., Pikiš, A., y Thompson, J. (2009). Evolution and Biochemistry of Family 4 Glycosidases: Implications for Assigning Enzyme Function in Sequence Annotations. *Molecular Biology and Evolution*, 26(11), 2487-2497. <https://doi.org/10.1093/molbev/msp162>

27. Hall, B. G. (2013). Building Phylogenetic Trees from Molecular Data with MEGA. *Molecular Biology and Evolution*, 30(5), 1229-1235. <https://doi.org/10.1093/molbev/mst012>

28. Hall, T. C., Mcleester, R. C., y Bliss, F. A. (1977). Equal Expression of the Maternal and Paternal Alleles for the Polypeptide Subunits of the Major Storage Protein of the Bean *Phaseolus vulgaris* L. *Plant Physiology*, 59(6), 1122-1124. doi:10.1104/pp.59.6.1122

29. Herrera G, G., y Bascur B, G. (1986). Bases del programa de mejoramiento genético del poroto en el INIA. Retrieved from <https://biblioteca.inia.cl/handle/123456789/31230>

30. Humann JL, Jung S, Cheng C-H, Lee T, Zheng P, Frank M, McGaughey D, Scott K, Buble K, Yu J, Hough H, Sanad M, Coyne C, McGee R, Main D (2019) Cool Season Food Legume Genome Database: A resource for pea, lentil, faba bean and chickpea genetics, genomics and breeding. Proceedings of the International Plant and Animal Genome Conference: January 2019, San Diego, CA, USA.

31. Humphrey, W., Dalke, A., y Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1), 33-38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)

32. INIA. (2017). Productores de Ñuble y Linares conocieron nuevo poroto INIA que saldría al mercado este año. INDAP. [http://www.indap.gob.cl/noticias/detalle/2017/02/16/productores-de- %C3 %B1uble-y-linares-conocieron -nuevo-poroto-inia-que-saldr %C3 %A1-al-mercado- este-a %C3 %B1o# : %7E:text=El %20poroto %20 en %20Chile,m %C3 %A1s %20demandado %20por %20los %20consumidores](http://www.indap.gob.cl/noticias/detalle/2017/02/16/productores-de-%C3%B1uble-y-linares-conocieron-nuevo-poroto-inia-que-saldr%C3%A1-al-mercado-este-a%C3%B1o#:~:text=El%20poroto%20en%20Chile,m%C3%A1s%20demandado%20por%20los%20consumidores).

33. Ira, L., Blitz, Luong, M., William, T., Chiu, y Cho, K. W. (2015). Applications of Deep Sequencing to Developmental Systems. *Principles of Developmental Genetics*, 37-48. <https://doi.org/10.1016/b978-0-12-405945-0.00002-8>

34. Kim, D., Paggi, J. M., Park, C., Bennett, C., y Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, 37(8), 907-915. <https://doi.org/10.1038/s41587-019-0201-4>
35. Krupa, U. (2008). Main nutritional and antinutritional compounds of bean seeds - a review - Polish Journal of Food and Nutrition Sciences - Tom 58, Numer 2 (2008) - Biblioteka Nauki - Yadda. Ceon. <http://yadda.icm.edu.pl/yadda/element/bwmeta1.element.agro-article-399f8e1c-3012-4785-9afb-f101eefa98e0>
36. Kumar, S., Stecher, G., Li, M., Knyaz, C., y Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, 35(6), 1547-1549. <https://doi.org/10.1093/molbev/msy096>
37. Langmead, B., Trapnell, C., Pop, M., y Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3), R25. <https://doi.org/10.1186/gb-2009-10-3-r25>
38. Li, W., y Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13), 1658-1659. <https://doi.org/10.1093/bioinformatics/btl158>
39. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., y Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079. <https://doi.org/10.1093/bioinformatics/btp352>
40. Lu, C., Wu, C., Ghoreishi, D., Chen, W., Wang, L., Damm, W., Ross, G. A., Dahlgren, M. K., Russell, E., von Bargen, C. D., Abel, R., Friesner, R. A., y Harder, E. D. (2021). OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *Journal of Chemical Theory and Computation*, 17(7), 4291-4300. <https://doi.org/10.1021/acs.jctc.1c00302>
41. Martino, H. S., Bigonha, S. M., Cardoso, L. D., Rosa, C. D., Costa, N. M., Lucia De Los Angeles Ramirez Cardenas, y Ribeiro, S. M. (2012). Nutritional and Bioactive Compounds of Bean: Benefits to Human Health. *ACS Symposium Series Hispanic Foods: Chemistry and Bioactive Compounds*, 233-258. doi:10.1021/bk-2012-1109.ch015

42. Martyna, G. J., Tobias, D. J., y Klein, M. L. (1994). Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics*, 101(5), 4177-4189. <https://doi.org/10.1063/1.467468>
43. McCammon, J. A., Gelin, B. R., y Karplus, M. (1977). Dynamics of folded proteins. *Nature*, 267(5612), 585-590. <https://doi.org/10.1038/267585a0>
44. Montoya, C. A., Leterme, P., Beebe, S., Souffrant, W. B., Molle, D., y Lalles, J. (2008). Phaseolin type and heat treatment influence the biochemistry of protein digestion in the rat intestine. *British Journal of Nutrition*, 99(3), 531-539. [doi:10.1017/s0007114507819179](https://doi.org/10.1017/s0007114507819179)
45. Morin, R. D., Bainbridge, M., Fejes, A., Hirst, M., Krzywinski, M., Pugh, T. J., McDonald, H., Varhol, R., Jones, S. J., y Marra, M. A. (2008). Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *BioTechniques*, 45(1), 81-94. <https://doi.org/10.2144/000112900>
46. Nose, S. (1984). A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics*, 81(1), 511-519. <https://doi.org/10.1063/1.447334>
47. Oliveira, A. P., Nomura, C. S., Silva, S. G., y Naozuka, J. (2018). Iron and selenium speciation in enriched adzuki bean sprouts after fractionation procedures by graphite furnace atomic absorption spectrometry. *Journal of Analytical y Pharmaceutical Research*, 7(2). [doi:10.15406/japlr.2018.07.00209](https://doi.org/10.15406/japlr.2018.07.00209)
48. Orry, A. J. W., y Abagyan, R. (2016). *Homology Modeling: Methods and Protocols: 857 (Softcover Reprint of the Original 1st 2012 ed.)*. Humana. 49. Osborne, T. (1924). The Vegetable Proteins. *Nature*, 114(2875), 822. <https://doi.org/10.1038/114822c0>
50. Paredes C, M., Becerra V, V., y Tay U, J. (2009). Inorganic Nutritional Composition of Common Bean (*Phaseolus vulgaris* L.) Genotypes Race Chile. *Chilean journal of agricultural research*, 69(4). <https://doi.org/10.4067/s0718-58392009000400002>
51. Perteau, M., Perteau, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., y Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, 33(3), 290-295. <https://doi.org/10.1038/nbt.3122>

52. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., y Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, 11(9), 1650-1667. <https://doi.org/10.1038/nprot.2016.095>
53. Prasad, B. N., y Mathew, L. (2008). *Recent advances in biotechnology*. Excel India.
54. Pullman, A. (Ed.). (2013). *Intermolecular Forces*. Springer Publishing.
55. Rates, S. (2001). Plants as source of drugs. *Toxicon*, 39(5), 603-613. doi:10.1016/s0041-0101(00)00154-9
56. Roos, K., Wu, C., Damm, W., Reboul, M., Stevenson, J. M., Lu, C., Dahlgren, M. K., Mondal, S., Chen, W., Wang, L., Abel, R., Friesner, R. A., y Harder, E. D. (2019). OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *Journal of Chemical Theory and Computation*, 15(3), 1863-1874. <https://doi.org/10.1021/acs.jctc.8b01026>
57. Rui, X., Boye, J. I., Ribereau, S., Simpson, B. K., y Prasher, S. O. (2011). Comparative study of the composition and thermal properties of protein isolates prepared from nine *Phaseolus vulgaris* legume varieties. *Food Research International*, 44(8), 2497-2504. doi:10.1016/j.foodres.2011.01.008
58. Sathe, S. (2016). Beans, Overview. Reference Module in Food Science. doi:10.1016/b978-0-08-100596-5.00033-0
59. Schrödinger Release 2021-3: SiteMap, Schrödinger, LLC, New York, NY, 2021.
60. Shewry, P. R., Napier, J. A., y Tatham, A. S. (1995). Seed storage proteins: Structures and biosynthesis. *The Plant Cell*, 7(7), 945-956. doi:10.1105/tpc.7.7.945
61. Silva, D. A. D., Tsai, S. M., Chiorato, A. F., da Silva Andrade, S. C., Esteves, J. A. D. F., Recchia, G. H., y Morais Carbonell, S. A. (2019). Analysis of the common bean (*Phaseolus vulgaris* L.) transcriptome regarding efficiency of phosphorus use. *PLOS ONE*, 14(1), e0210428. <https://doi.org/10.1371/journal.pone.0210428>

62. Singh, R. J., y Jauhar, P. P. (2006). *Cereals*. Taylor y Francis.
63. Slightom, J. L., Drong, R. F., Klassy, C., y Hoffman, L. M. (1985). Nucleotide sequences from pbaseolln cDNA clones: The major storage proteins from *Phaseolus vulgaris* are encoded by two unique gene families. *Nucleic Acids Research*, 13(18), 6483-6498. doi:10.1093/nar/13.18.6483
64. Stanke, M., Steinkamp, R., Waack, S., y Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Research*, 32(Web Server), W309-W312. <https://doi.org/10.1093/nar/gkh379>
65. Suarez-Martinez, S. E., Ferriz-Martinez, R. A., Campos-Vega, R., Elton-Puente, J. E., Carbot, K. D., y Garcia-Gasca, T. (2015). Bean seeds: Leading nutraceutical source for human health. *CyTA - Journal of Food*, 14(1), 131-137. doi:10.1080/19476337.2015.1063548
66. Suttie, J. M. (1999). *Temperate Forage Legumes*. By J. Frame, J. F. L. Charlton and A. S. Laidlaw. Wallingford, UK: CAB INTERNATIONAL (1998), pp. 327,. ISBN 0-85199-2145. *Experimental Agriculture*, 35(1), 111-114. doi:10.1017/s0014479799241075
67. Tang, C., y Sun, X. (2011). Structure-physicochemical function relationships of 7S globulins (vicilins) from red bean (*Phaseolus angularis*) with different polypeptide constituents. *Food Hydrocolloids*, 25(3), 536-544. doi:10.1016/j.foodhyd.2010.08.009
68. Tang, C. (2008). Thermal denaturation and gelation of vicilin-rich protein isolates from three *Phaseolus* legumes: A comparative study. *LWT - Food Science and Technology*, 41(8), 1380-1388. doi:10.1016/j.lwt.2007.08.025
69. Thompson, J. D., Higgins, D. G., y Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research*, 22(22), 4673-4680. <https://doi.org/10.1093/nar/22.22.4673>
70. Wallner, B., y Elofsson, A. (2003). Can correct protein models be identified? *Protein Science*, 12(5), 1073-1086. <https://doi.org/10.1110/ps.0236803>

71. Weissig, H., y Bourne, P. E. (2003). *Structural Bioinformatics* (1.a ed.). Wiley-Liss.
72. Wiederstein, M., y Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, 35(Web Server), W407-W410. <https://doi.org/10.1093/nar/gkm290>
73. Xiong, J. (2019). Multiple Sequence Alignment. *Essential Bioinformatics*, 63-74. <https://doi.org/10.1017/cbo9780511806087.006>

8. ANEXOS

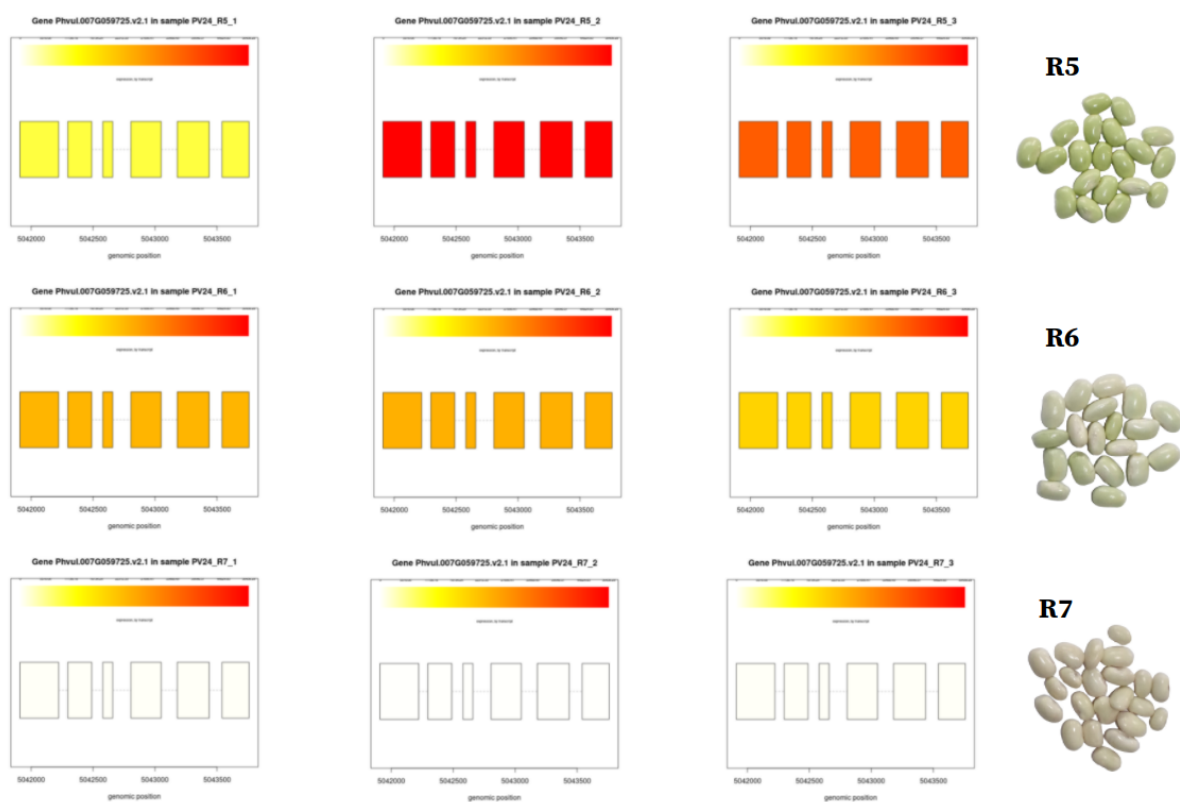


Figura 8.1: Análisis de expresión diferencial con respecto al gen Phvul.007G059725.v2.1 que codifica a faseolina en las muestras de poroto coscorrón.

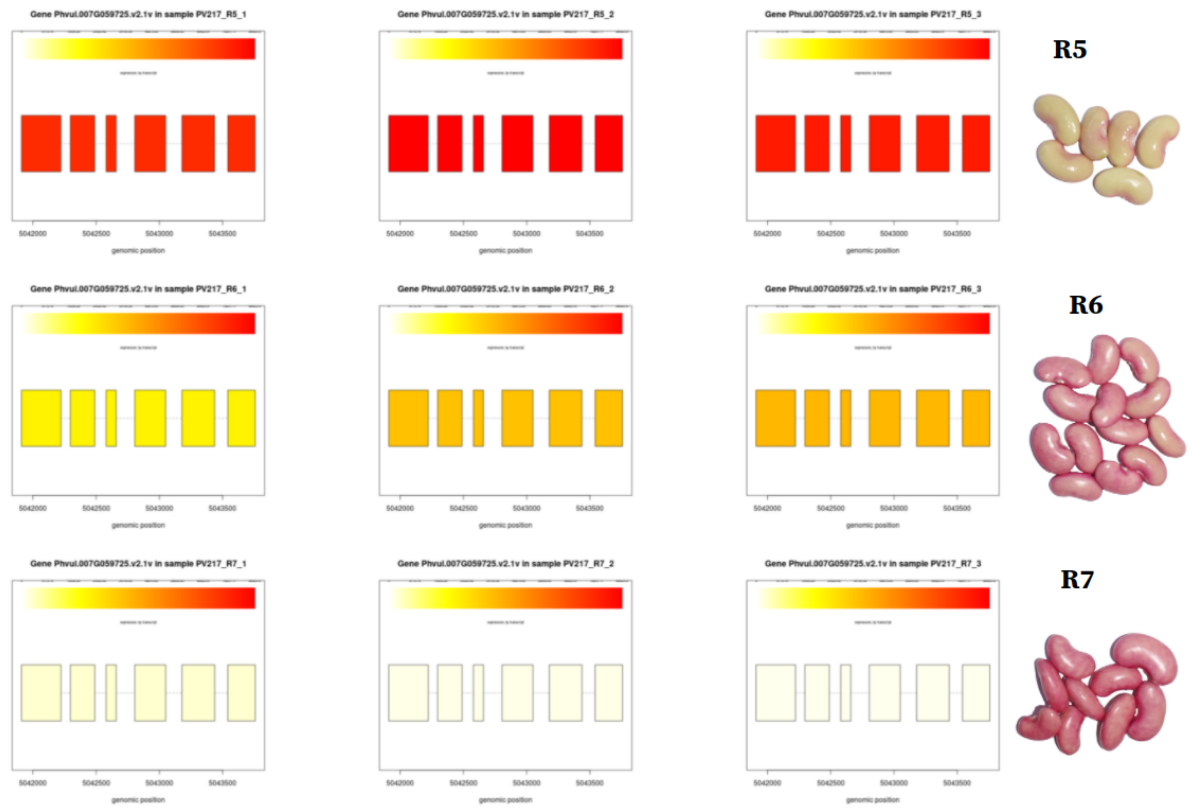


Figura 8.2: Análisis de expresión diferencial con respecto al gen *Phvul.007G059725.v2.1* que codifica a faseolina en las muestras de proto enriquetá.

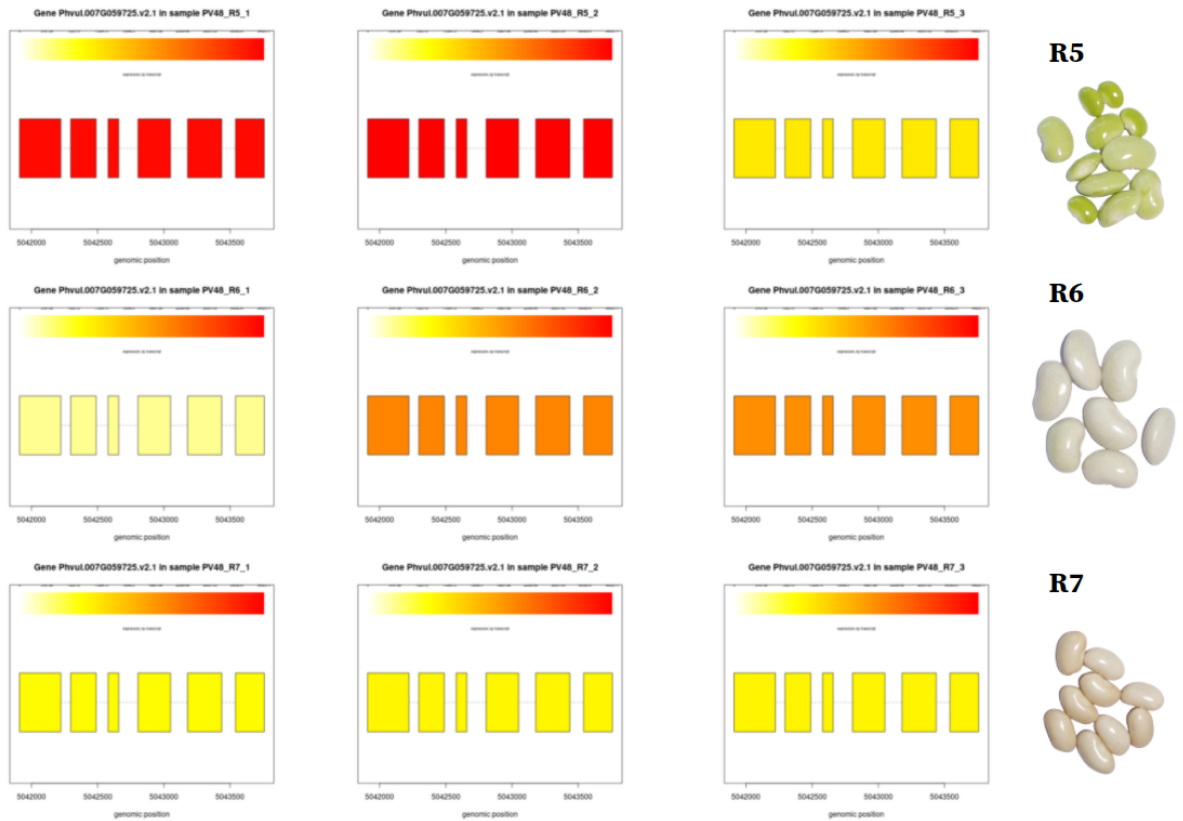


Figura 8.3: Análisis de expresión diferencial con respecto al gen *Phvul.007G059725.v2.1* que codifica a faseolina en las muestras de poroto tórtola.

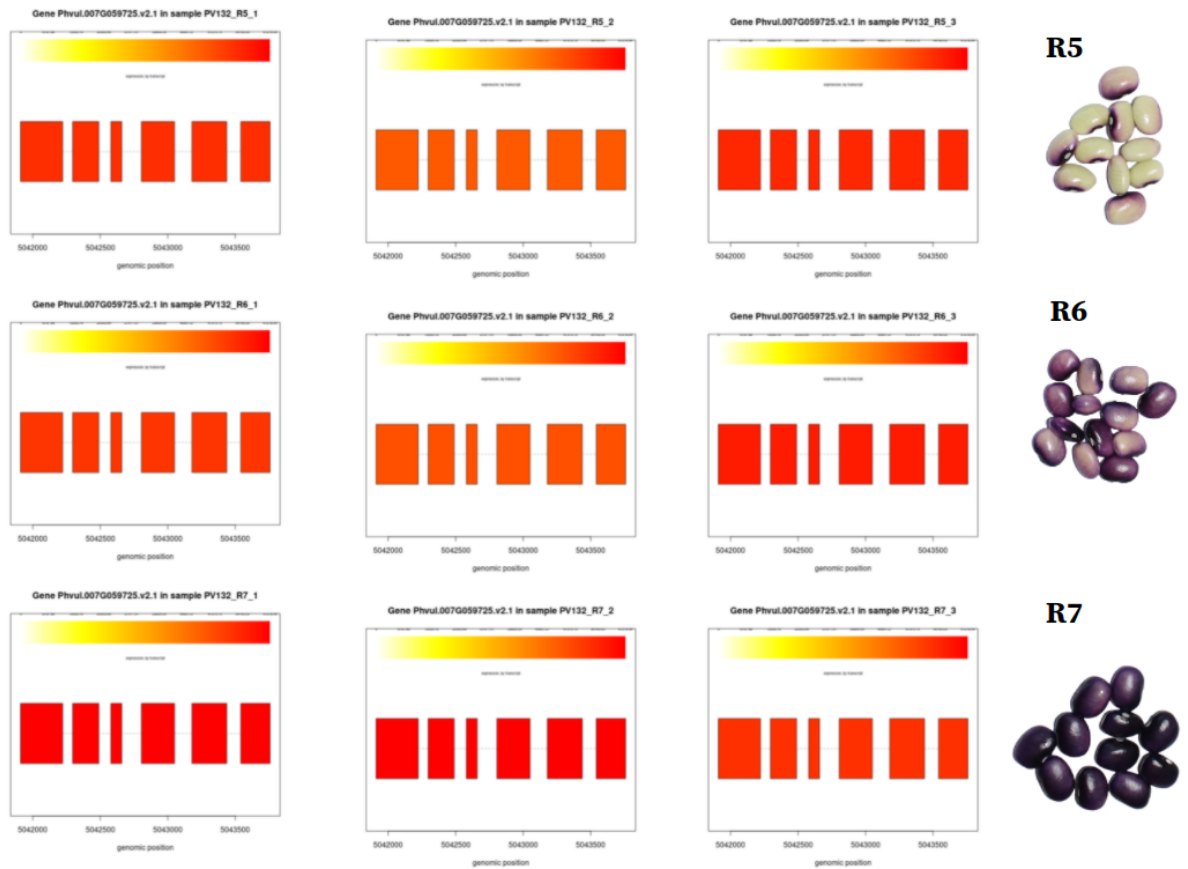


Figura 8.4: Análisis de expresión diferencial con respecto al gen *Phvul.007G059725.v2.1* que codifica a faseolina en las muestras de poroto negro argel.



Figura 8.5: Alineamiento de secuencias de faseolinas alfa y beta.