



UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO

FACULTAD DE CIENCIAS PECUARIAS Y BIOLÓGICAS

CARRERA INGENIERÍA AGROPECUARIA

Trabajo de Integración Curricular
previa la obtención del Grado
Académico de Ingeniera Agropecuaria

Proyecto de Investigación:

“ANÁLISIS BIOINFORMÁTICO DE RUTAS METABÓLICAS RELACIONADAS
CON LA RESISTENCIA POR CONTAMINACIÓN DE ALUMINIO Y CADMIO EN EL
CULTIVO DE CACAO (*Theobroma cacao* L.)”

Autora:

Yuliana Stefania Campuzano Herrera

Director de Proyecto de Investigación:

Ing. Agron. Ronald Oswaldo Villamar Torres PhD

Mocache – Los Ríos – Ecuador

2023



DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

Yo, Yuliana Stefania Campuzano Herrera, declaro que la investigación aquí descrita es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Universidad Técnica Estatal de Quevedo, puede hacer uso de los derechos correspondientes a este documento, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

.....
Yuliana Campuzano

Yuliana Stefania Campuzano Herrera

C.C.1207532068



CERTIFICADO DE CULMINACIÓN DEL PROYECTO DE INVESTIGACIÓN

El suscrito, **Ing. Agron. Ronald Oswaldo Villamar Torres PhD.**, docente de la Universidad Técnica Estatal de Quevedo, **CERTIFICA** que la estudiante Yuliana Stefania Campuzano Herrera, realizó el Proyecto de Investigación de grado titulado “**ANÁLISIS BIOINFORMÁTICO DE RUTAS METABÓLICAS RELACIONADAS CON LA RESISTENCIA POR CONTAMINACIÓN DE ALUMINIO Y CADMIO EN EL CULTIVO DE CACAO (*Theobroma cacao* L.)**”, previo a la obtención del título de Ingeniera Agropecuario, bajo mi dirección, habiendo cumplido con las disposiciones reglamentarias establecidas para el efecto.

Ing. Agron. Ronald Oswaldo Villamar Torres PhD.

Director del Proyecto de Investigación



CERTIFICADO DEL REPORTE DE LA HERRAMIENTA DE PREVENCIÓN DE COINCIDENCIA Y/O PLAGIO ACADÉMICO

El suscrito, Ing. Agron. Ronald Oswaldo Villamar Torres PhD., docente de la Universidad Técnica Estatal de Quevedo, en calidad de Director del Proyecto de Investigación titulado “ANÁLISIS BIOINFORMÁTICO DE RUTAS METABÓLICAS RELACIONADAS CON LA RESISTENCIA POR CONTAMINACIÓN DE ALUMINIO Y CADMIO EN EL CULTIVO DE CACAO (*Theobroma cacao* L.)”, perteneciente a la egresada Yuliana Stefania Campuzano Herrera, CERTIFICA: el cumplimiento de parámetros establecidos por el SENESCYT, y se evidencia el reporte de la herramienta de prevención de coincidencia y/o plagio académico (URKUND) con un porcentaje de coincidencia del 3%.



Document Information

Analyzed document	Yuliana Campuzano.pdf (D170202582)
Submitted	6/9/2023 7:47:00 PM
Submitted by	Gregorio Váscquez
Submitter email	gvasconez@uteq.edu.ec
Similarity	3%
Analysis address	gvasconez.uteq@analysis.urkund.com

Sources included in the report

SA M0.154_20221_PEC 2_18877380.txt
Document M0.154_20221_PEC 2_18877380.txt (D154563734)

88 1

Ing. Agron. Ronald Oswaldo Villamar Torres PhD.

DIRECTOR DE PROYECTO DE INVESTIGACIÓN



UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO
FACULTAD DE CIENCIAS PECUARIAS Y BIOLÓGICAS
CARRERA AGROPECUARIA
PROYECTO DE INVESTIGACIÓN

Título:

“ANÁLISIS BIOINFORMÁTICO DE RUTAS METABÓLICAS RELACIONADAS CON LA RESISTENCIA POR CONTAMINACIÓN DE ALUMINIO Y CADMIO EN EL CULTIVO DE CACAO (*Theobroma cacao* L.)”

Presentado al Consejo Directivo de Facultad como requisito previo a la obtención del título de Ingeniera Agropecuaria.

Aprobado por:

.....
PRESIDENTE DE TRIBUNAL
Ing. Agron. Luis Godoy Montiel PhD

.....
MIEMBRO DE TRIBUNAL
Ing. Agron. Roxanna Palma Leon PhD

.....
MIEMBRO DEL TRIBUNAL
Ing. Marco Heredia Rengifo Msc

MOCACHE – LOS RÍOS – ECUADOR

2023

AGRADECIMIENTO

Primero doy gracias a Dios, quien siempre me llena de bendiciones. Y agradecer de todo corazón a las personas involucradas en este camino o trayecto universitario.

Suele decirse que el mayor regalo que nos pueden dejar nuestros padres es la universidad. En lo personal, no es lo único que nos dejan. También incluyo su amor, valores y el respeto. Mi eterna gratitud a mi maravillosa familia. Mi madre, mi padre y mis hermanos, y a mi acompañante de vida mi esposo, quienes son las piedras angulares de mi finalización de la universidad.

En el transcurso del camino conocemos a personas que brillan con la más alta intensidad, llamados amigos de por vida. Siempre estarán ahí cuando te caigas, serán hermanos para levantarte y mantenerte brillando. Mis más sinceros agradecimientos a mis mejores amigas Jackeline Segura C. y Cecibel Carrión J. y a mis amigos Adnan Murillo O. y Diego Murillo O.

Mi meta de estudiar en una universidad de prestigio lo logre ya que la UTEQ me abrió las puertas a mi formación profesional y cada docente, con paciencia y esmero, integraron los conocimientos adquiridos para convertirme en una flamante INGENIERA AGROPECUARIA. Este título es lo que reconozco como herencia de mi formación. Los esfuerzos de los maestros, de mis padres y yo mismo.

Un buen maestro puede inspirar esperanza, inspirar la imaginación, nunca olvidaré su enseñanza y tiempo que me ofreció a cada instante, y el aprendizaje que durará toda la vida. Al Dr. Ronald Villamar Torres, buen maestro y buen amigo, por su dedicación en su valiosa labor en la enseñanza de otros estudiantes, y por ser mi mentor y gran tutor, siempre dispuesto para cualquier pregunta o enseñanza durante el periodo de tesis.

También agradezco al Dr. Seyed Mehdi Jazayeri de la Universidad de Marsella y el CIRAD de Francia, quién como un co-tutor me brindó su gran apoyo durante varias jornadas por videoconferencia, sin importar el cambio horario, ni la distancia, compartiendo sus conocimientos, para así poder culminar con éxito mi tesis.

Si el proceso es duro es porque cosas grandes vienen...

Yuliana Stefania Campuzano Herrera.

DEDICATORIA

Este trabajo se completó con éxitos antes de obtener el título de Ingeniero en Agropecuaria. Este trabajo refleja algunos de los conocimientos adquiridos en la prestigiosa UTEQ dedico a:

Dios por siempre estar conmigo y por su infinita misericordia, a mi familia, amigos/as y personas muy especiales y únicas en mi vida.

Quienes ayudaron a cumplir una meta más que anhelaba, es un simple gesto de inspiración que ha contribuido al desarrollo de mi formación académica.

Ahora es cuando empieza la verdadera aventura de la vida.

Yuliana Stefania Campuzano Herrera.

RESUMEN

El presente trabajo investigativo tuvo como finalidad hacer uso de la Bioinformática para estudiar las rutas metabólicas relacionadas con la resistencia por contaminación de Aluminio y Cadmio en el cultivo de cacao (*Theobroma cacao* L.). El estudio se llevó a cabo en el Campus Experimental “La María”, Facultad de Ciencias Pecuarias y Biológicas perteneciente a la Universidad Técnica Estatal de Quevedo, ubicada en el kilómetro 7.5 de la vía Quevedo - El Empalme, Recinto San Felipe, Cantón Mocache, Los Ríos. Para comparar los genomas secuenciados de cacao con otras especies de las Malvaceae y los metallophytes con el fin de revelar datos genómicos relacionados a factores genéticos en los procesos evolutivos de ambas especies, adicionalmente caracterizar y elucidar de forma in silico familia de genes, factores de transcripción y proteínas relacionados con efectos de metales pesados, para lo cual se utilizaron varios softwares disponibles de forma gratuita en la web. Los resultados fueron filtrados por Evalúe cutoff 10E-10. PlantTFcat es una herramienta para analizar la identificación y categorización de genes de plantas TF / TR / CR a escala genómica.

Palabras claves: Cacao, genoma, bioinformática, comparación in silico, metales pesados

ABSTRACT

The present research was aimed at using Bioinformatics to study metabolic pathways related to resistance by aluminum and cadmium pollution in cocoa (*Theobroma cacao* L.) cultivation. The study was carried out at the Campus Experimental “La María”, Facultad de Ciencias Pecuarias y Biológicas, belonging to the Universidad Técnica Estatal de Quevedo, located at kilometer 7.5 of the Quevedo - El Empalme road, San Felipe Campus, Canton Mocache, To compare the sequenced cacao genomes with other species of the Malvaceae and the metallophytes in order to reveal genomic data related to genetic factors in the evolutionary processes of both species, additionally to characterize and elucidate in silico family of genes, transcription factors and protein related heavy metal effects various softwares available for free on the web will be used. The results were filtered by Evaluate cutoff $10E-10$. PlantTFcat is a tool to analyze the identification and categorization of TF/TR/CR plant genes at the genomic scale.

Keywords: Cocoa, genome, bioinformatics, in silico comparation, heavy metals

TABLA DE CONTENIDO

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS	ii
CERTIFICADO DE CULMINACIÓN DEL PROYECTO DE INVESTIGACIÓN	iii
CERTIFICADO DEL REPORTE DE LA HERRAMIENTA DE PREVENCIÓN DE COINCIDENCIA Y/O PLAGIO ACADÉMICO	iv
DEDICATORIA	vii
RESUMEN	viii
ABSTRACT	ix
CÓDIGO DUBLÍN.....	xiv
Introducción	1
CAPÍTULO I 3	
CONTEXTUALIZACIÓN DE LA INVESTIGACIÓN	3
1.1. Problema de investigación.	4
1.1.1. Planteamiento del problema.	4
Diagnóstico.	4
Pronóstico.	5
1.1.2. Formulación del problema.	5
1.1.3. Sistematización del problema.....	5
1.2. Objetivos.	5
1.2.1. Objetivo general.	5
1.2.2. Objetivos específicos.	6
CAPÍTULO II.....	7
FUNDAMENTACIÓN TEÓRICA DE LA INVESTIGACIÓN.....	7
2.1. Marco conceptual.	8
2.1.1. Cacao.	8
2.1.2. Metales pesados.....	8
2.1.3. Bioinformática.....	8
2.1.4. Rutas metabólicas.	8
2.2. Marco referencial.....	8
2.2.1. Origen y generalidades del cacao.	8
2.2.1.1. Descripción taxonómica.....	9
2.2.1.2. Descripción botánica.....	9
2.2.2. Producción de cacao en Ecuador.	9
2.2.2.1. Variedades.....	11
2.2.3. Aspectos teóricos del Aluminio (Al) y el Cadmio (Cd).....	11
2.2.3.1. Aluminio.	11

2.2.3.2.	Cadmio.....	11
2.2.4.	Efectos de metales pesados en el cultivo de cacao.	12
2.2.5.	Estudios efectuados de metales pesados en cultivo de cacao.	12
CAPÍTULO III		14
METODOLOGÍA DE LA INVESTIGACIÓN		14
3.1.	Localización.	15
3.2.	Tipo de investigación.....	15
3.3.	Métodos de investigación.....	15
3.8.	Recursos humanos y materiales.....	20
3.8.1.	Materiales.	20
3.8.1.1.	Materiales de oficina.....	20
CAPÍTULO IV		21
RESULTADOS Y DISCUSIÓN		21
4.1.	Comparación de clúster de proteínas de las 4 especies.	22
4.1.1.	Secuencias de proteínas.	24
4.1.2.	Análisis de secuencias de proteínas basada en OrthoVenn2	25
4.1.3.	Resultados de análisis de proteínas en Orthovenn2.	27
4.2.1.	Análisis de las proteínas agrupadas (Clúster) para Cadmio (Cd)	33
4.2.2.	Interpretación del Clúster133.	33
4.2.3.	Interpretación del Clúster500.	34
4.2.4.	Interpretación del Clúster676.	35
4.2.5.	Análisis de las proteínas agrupadas (Clúster) para Aluminio (Al).	35
4.2.6.	Interpretación del Clúster919	36
4.2.7.	Interpretación del Clúster958.	36
CAPÍTULO V		39
CONCLUSIONES Y RECOMENDACIONES		39
5.1.	Conclusiones	40
5.2.	Recomendaciones	40
CAPÍTULO VI		41
BIBLIOGRAFÍA		41
6.1.	Bibliografía	42
CAPÍTULO VII.....		49
ANEXOS		49
7.1.	Evidencia fotográfica.....	50

ÍNDICE DE TABLAS

Tabla 1 Condiciones meteorológicas aproximadas del Campus Experimental “La María”	15
Tabla 2 Ensablamiento completos de los genomas de Cacao criollo, <i>Herrania umbrática</i> , <i>Arabidopsis haliana</i> y <i>Arabidopsis halleri</i> , con sus respectivas referencias, tipo de anotación, tamaño y fecha de carga sobre las plataformas NCBI y Phytozome.	23
Tabla 3 Número de proteínas, grupos y datos únicos de cada conjunto de datos de proteínas de las 4 especies.	25
Tabla 4 Secuencias de proteínas relacionadas a la resistencia a Aluminio (Al) encontradas por similitud.	27
Tabla 5 Secuencias de Proteínas relacionadas a la resistencia a Cadmio (Cd) encontradas por similitud.	28

ÍNDICE DE FIGURAS

Figura 1	Participación de las provincias en la superficie plantada total de cacao	10
Figura 2	Esquema de búsqueda de secuencias de genómicas de proteínas en la plataforma NCBI(National Center for Biotechnology Information).....	17
Figura 3	Esquema de búsqueda de secuencias de genómicas de proteínas en la plataforma.	18
Figura 4	Resultado de la carga a la plataforma Orthovenn 2 de las secuencias de proteínas de las cuatro especies: Theobroma cacao (cacao), Herrania umbrática, Arabidopsis thaliana (thale cress) y Arabidopsis halleri, y criterios de análisis bioinformáticos para la comparación genómica.	24
Figura 5	Diagrama de Venn de comparación entre 4 especies. 11407 grupos fueron comunes entre las 4 especies. 552 grupos son exclusivamente de T. cacao, 231 grupos para H. umbratica, 240 grupos para C. canephora y C. eugenioides y 111 para C. arabica, C. eugenioides y T. cacao.....	26
Figura 6	Compatibilidad de los clústers de las cuatro especies: Theobroma cacao (cacao), Herrania umbratica, Arabidopsis thaliana (thale cress) y Arabidopsis halleri.	33
Figura 7	Análisis de clúster 133 para Cadmio (Cd).....	33
Figura 8	Análisis de clúster 500 para Cadmio (Cd).....	34
Figura 9	Análisis de clúster 676 para Cadmio (Cd).....	35
Figura 10	Análisis de clúster 919 para Aluminio (Al).....	36
Figura 11	Análisis de clúster 958 para Aluminio (Al).....	36

CÓDIGO DUBLÍN

Título	“Análisis bioinformático de rutas metabólicas relacionadas con la resistencia por contaminación de aluminio y cadmio en el cultivo de cacao (<i>Theobroma cacao</i> L.)”,			
Autora:	Yuliana Stefania Campuzano Herrera			
Palabras claves:	Cultivo	Genoma	Metales pesado	Información
Fecha de publicación :	2023			
Editorial :				
Resumen:	El presente trabajo investigativo tuvo como finalidad hacer uso de la Bioinformática para estudiar las rutas metabólicas relacionadas con la resistencia por contaminación de Aluminio y cadmio en el cultivo de cacao (<i>Theobroma cacao</i> L.). el estudio se llevó a cabo en el Campus Experimental “La María”, Facultad de Ciencias Pecuarias y Biológicas perteneciente a la Universidad Técnica Estatal de Quevedo, ubicada en el kilómetro 7.5 de la vía Quevedo - El Empalme, Recinto San Felipe, Cantón Mocache, Los Ríos. Para comparar los genomas secuenciados de cacao con otras especies de las Malvaceae y los metallophytes con el fin de revelar datos genómicos relacionados a factores genéticos en los procesos evolutivos de ambas especies, adicionalmente caracterizar y elucidar de forma in silico familia de genes, factores de transcripción y proteínas relacionados con efectos de metales pesados varios softwares disponibles de forma gratuita en la web serán utilizados. Los resultados fueron filtrados por Evaluate cutoff 10E-10.			
Abstract:	The present research was aimed at using Bioinformatics to study metabolic pathways related to resistance by aluminum and cadmium pollution in cocoa (<i>Theobroma cacao</i> L.) cultivation. The study was carried out at the Campus Experimental “La María”, Facultad de Ciencias Pecuarias y Biológicas, belonging to the Universidad Técnica Estatal de Quevedo, located at kilometer 7.5 of the Quevedo - El Empalme road, San Felipe Campus, Canton Mocache, To compare the sequenced cacao genomes with other species of the Malvaceae and the metallophytes in order to reveal genomic data related to genetic factors in the evolutionary processes of both species, additionally to characterize and elucidate in silico family of genes, transcription factors and protein related heavy metal effects various softwares available for free on the web will be used. The results were filtered by Evaluate cutoff 10E-10.			
Descripción:	69 hojas: dimensiones, 29 x 21 cm + CD-ROM.			
URL:				

Introducción

Actualmente, la acumulación y presencia de metales pesados en cultivos de cacao (*Theobroma cacao* L.), se ha convertido en un problema de alto impacto, dada la importancia de este cultivo para el Ecuador, principalmente por la actividad de exportación en calidad de materia prima, así como de los productos elaborados (1) Además, es considerado diferente al resto de países, esto gracias a sus características organolépticas y su gran calidad (2); sin embargo, en su desarrollo, la planta de cacao puede absorber distintos metales pesados provenientes del suelo, concentrándose en sus semillas (3); y, por lo tanto, ser transferidos hacia el ser humano (4).

Diversos trabajos investigativos consideran este tipo de metales como elementos muy peligrosos para la alimentación humana situación que ha captado la atención de la comunidad científica orientada a su descripción y comportamiento en los sistemas biológicos a plantear alternativas de prevención, control y remediación (5) (6). En Ecuador se ha presentado evidencia científica de que la semilla de cacao contiene metales pesados. Donde la provincia de El Oro presenta los valores más altos de cadmio en cacao, seguido por Guayas y Manabí (7).

La tolerancia y/o resistencia de las plantas al estrés por metales puede estar asociada a uno o más mecanismos, tales como la excreción de compuestos quemantes que reducen la disponibilidad del metal en el suelo o en el agua; la exclusión del metal mediante absorción selectiva de elementos; retención de metales en las raíces, evitando su translocación a los brotes; secuestro de metales pesados por aglutinantes, biotransformación, compartimentación y mecanismos de reparación celular; el desarrollo de enzimas tolerantes a metales, entre otros (8).

Ante esto, considerando las tendencias de consumo, se necesitan cultivos que sean más sanos y de buena producción, para obtener mayores rendimientos en el cultivo de cacao. Por tal razón, la presente investigación pretende estudiar con el uso de la Bioinformática, las rutas metabólicas relacionadas con la resistencia por contaminación de Aluminio y Cadmio en el cultivo de cacao.

Esto permitirá identificar posibles secuencias relacionadas a la resistencia a metales pesados como aluminio y cadmio que podrían permitir ser la base inicial de un programa de fitomejoramiento en este cultivo de tanta importancia para el país. Estos tipos de datos ómicos son una enorme alternativa hoy en día para acelerar procesos que tomaban mucho tiempo antes de la era ómica, y en lo posterior hacer estudios de expresión génica y funcionabilidad de genes relacionados a la resistencia a metales pesados como Aluminio y Cadmio.

CAPÍTULO I

CONTEXTUALIZACIÓN DE LA INVESTIGACIÓN

1.1. Problema de investigación.

1.1.1. Planteamiento del problema.

El exceso de metales pesados es un problema grave que se ha planteado durante las últimas décadas debido al cambio climático y las actividades indeseables causadas por el hombre (como la industrialización, los fertilizantes sintéticos, etc.). Las plantas son sensibles y no pueden escapar si su entorno, la rizosfera y el aire se contaminan (9). La acumulación y presencia de metales pesados en los cultivos de cacao se encuentran de manera natural en la corteza terrestre, mismos que pueden ser absorbidos por las semillas, las plantas y sus frutos (10).

Las plantas adoptan su propio metabolismo y se adaptan a condiciones adversas. Por eso, han resistido desde la prehistoria hasta la actualidad. El cacao es uno de los cultivos con facturación de miles de millones de dólares a nivel mundial. Uno de los mayores problemas hoy en día en el cultivo de cacao es que este puede acumular metales pesados en sus granos, lo que se traduce en un grave riesgo para la salud humana y también para la producción de cacao.

Para comprender cómo las especies de cacao pueden enfrentar el problema de metales pesados y cómo las plantas de cacao pueden adaptarse y tolerar el estrés de metales pesados, es necesario elucidar y explorar algunas rutas metabólicas que permitan identificar secuencias génicas que lideren algunos procesos de resistencia o tolerancia a la presencia de metales pesados como Aluminio y Cadmio. Los genes son factores claves, pues, aquellos lideran, regulan y conducen ciertas reacciones biológicas en las plantas, entre ellas, las rutas y las redes involucradas en respuestas a metales pesados (9).

Diagnóstico.

La disponibilidad de bases de datos genómicas a través de sistemas abiertos demuestra la aplicabilidad de estas soluciones bioinformáticas para problemas actuales tanto en el reino animal como vegetal. Estas herramientas varían desde una colección de herramientas autónomas con un formato de datos común, y bajo una única interface autónoma o basada en la web, hasta sistemas integradores y extensibles para la gestión del flujo de trabajo bioinformático.

- Compilación de conjuntos de datos.
- Comparaciones (los programas y las herramientas).

Pronóstico.

Se muestran nuevas alternativas para acelerar los procesos de selección o fitomejoramiento en el cultivo de cacao, pues las secuencias caracterizadas permitirán iniciar investigaciones de expresión génica, utilizando un barrido de genes de resistencia o tolerancia por vía biocomputacional con relación a la presencia de metales pesados como Aluminio y Cadmio en el cultivo de cacao (*Theobroma cacao* L.).

1.1.2. Formulación del problema.

¿Cuáles serán las rutas biocomputacional metabólicas relacionadas con la resistencia por contaminación de metales pesados de plomo y Cadmio en el cultivo de cacao (*Theobroma cacao* L.)?

1.1.3. Sistematización del problema.

- ¿Cuáles son las secuencias genómicas relacionadas con la resistencia por contaminación de Aluminio y Cadmio en el cultivo de cacao (*Theobroma cacao*)?
- ¿Existen secuencias de proteínas compartidas del genoma de (*Theobroma cacao*) con otras especies relativas para elucidar algunas funciones relacionadas con la resistencia por contaminación de Aluminio y Cadmio?
- ¿Cuál es la filogenia de secuencias relacionadas con la resistencia por contaminación de Aluminio y Cadmio en (*Theobroma cacao*) que defina con mayor probabilidad su historia evolutiva?

1.2.Objetivos.

1.2.1. Objetivo general.

Estudiar con el uso de la Bioinformática las rutas metabólicas relacionadas con la resistencia por contaminación de Aluminio y Cadmio en el cultivo de cacao (*Theobroma cacao* L.).

1.2.2. Objetivos específicos.

- Caracterizar secuencias genómicas en repositorios bioinformáticos relacionadas con la resistencia por contaminación de Aluminio y Cadmio utilizando los algoritmos BLASTN, BLASTP y BLASTX.
- Comparar secuencias de proteínas del genoma de (*Theobroma cacao* L.) con otras especies relativas para observar la similitud y elucidar algunas funciones relacionadas con la resistencia por contaminación de Aluminio y Cadmio.
- Realizar un análisis filogenético utilizando las secuencias genómicas seleccionadas en (*Theobroma cacao* L.) para definir con mayor probabilidad la historia evolutiva de las moléculas comparadas.

1.3. Justificación.

Se han desarrollado interfaces para una amplia variedad de aplicaciones bioinformáticas, permitiendo que una aplicación, corriendo en un ordenador de cualquier parte del mundo, pueda usar algoritmos, datos y recursos de computación alojados en servidores en cualesquiera otras partes del planeta. Las principales ventajas radican en que el usuario final se despreocupa de actualizaciones y modificaciones en el software o en las bases de datos.

La presente investigación se enfocó en comparar las secuencias de genes/proteínas de cacao con los genes/proteínas de otras plantas cuya función está relacionada con la respuesta, actividad, acumulación y tolerancia de metales pesados (HM). Para realizar este tipo de análisis comparativos, hay algunos elementos que proporcionar:

Secuencias de genes y/o proteínas del cacao (datos genómicos), secuencias de genes y/o proteínas que están relacionadas con metales pesados de otras especies vegetales, especialmente plantas con capacidad potencial o probada de tolerancia a HM y herramientas bioinformáticas que permiten comparar secuencias de diferentes conjuntos de datos, analizar funciones, comprender cómo reaccionan las plantas a HM por vías y redes

Así, el presente trabajo permitió mostrar las comparaciones y obtener resultados más precisos, se pueden emplear varias herramientas bioinformáticas según los objetivos del estudio y los recursos computacionales disponibles.

CAPÍTULO II

FUNDAMENTACIÓN TEÓRICA DE LA INVESTIGACIÓN

2.1. Marco conceptual.

2.1.1. Cacao.

El cacao es una fruta originaria de América tropical en la región superior del Amazonas. Su historia se remonta al tercer milenio antes de Cristo (a.C.), es decir, hace más de 2.000 años, donde civilizaciones del sur de México y Centroamérica fueron los principales pueblos en reconocer y utilizar las apreciables cualidades de la almendra de cacao adaptándola a su alimentación habitual (11).

2.1.2. Metales pesados.

Los metales son componentes naturales de la corteza terrestre, que juegan un papel importante en los organismos al ser parte fundamental de sus funciones bioquímicas y fisiológicas (12).

2.1.3. Bioinformática.

La bioinformática se refiere a la disciplina científica que permite investigar, desdoblar y emplear herramientas computacionales para mejorar el manejo de datos biológicos (13).

2.1.4. Rutas metabólicas.

Las rutas o vías metabólicas se definen como la sucesión de reacciones químicas donde el sustrato inicial se convierte dando paso a productos finales, mediante metabolitos que actúan como intermediarios (14).

2.2. Marco referencial.

2.2.1. Origen y generalidades del cacao.

Theobroma cacao L., es un árbol tropical perteneciente a la familia de la *Esterculiaceae*, cuyo fruto es color púrpura y similar a las almendras, pero de sabor amargo. Su nombre “*Theobroma*” significa “alimento de los dioses” en idioma griego, “cacao” por su parte viene de la palabra maya “*kakaw*”, con la que esta civilización conocía a la planta (15). Los mayas establecieron las primeras plantaciones cacaoteras en Centroamérica.

Barazarte *et al.* (16) afirman que es muy probable que los Olmecas fueran los responsables de la domesticación del cacao hace tres mil años, pero la difusión de su uso se le atribuye a

los Mayas ya que formaban parte importante de sus actividades culturales como alimento, medicina e incluso parte de su estructura económica.

2.2.1.1. Descripción taxonómica.

Taxonómicamente, el cacao es una de las 22 especies del género *Theobroma*, su taxonomía es la siguiente (17):

- Reino: Plantae
- División: Magnoliophyta
- Clase: Magnoliopsida
- Subclase: Dilleniidae
- Orden: Malvales
- Familia: Malvaceae
- Género: Theobromeae

2.2.1.2. Descripción botánica.

A nivel botánico, el cacao es un árbol que crece silvestre, en los bosques de América Central, en la zona situada entre los 26°N y 26°S de Ecuador, los árboles cultivados son más pequeños los cuales facilitan su recolección y cultivo, no suelen sobrepasar los dos o tres metros de altura.

También se conoce que se cultiva en las zonas tropicales del oeste de África y Asia su tamaño mediano normalmente tiene una altura entre 6 a 8 metros de altura y alcanza hasta los 20m² crece libremente bajo sombra intensa la corona es de un diámetro de 7 y 9 metros el tronco es recto y se puede desarrollaren formas muy variadas según las condiciones (2).

2.2.2. Producción de cacao en Ecuador.

Chiriboga (18) indica que la producción de cacao en Ecuador tuvo su esplendor a fines del siglo XIX, lo que posicionó al país como el principal exportador del mundo en los años ochenta. Debido a este auge, el cacao es conocido en el país como “la pepa de oro”, al promover la mayor generación de divisas para el país a fines del siglo XIX. Por las condiciones climáticas, en Ecuador se produce cacao primordialmente en las provincias del, Guayas, Los Ríos, Manabí y Sucumbíos (19).

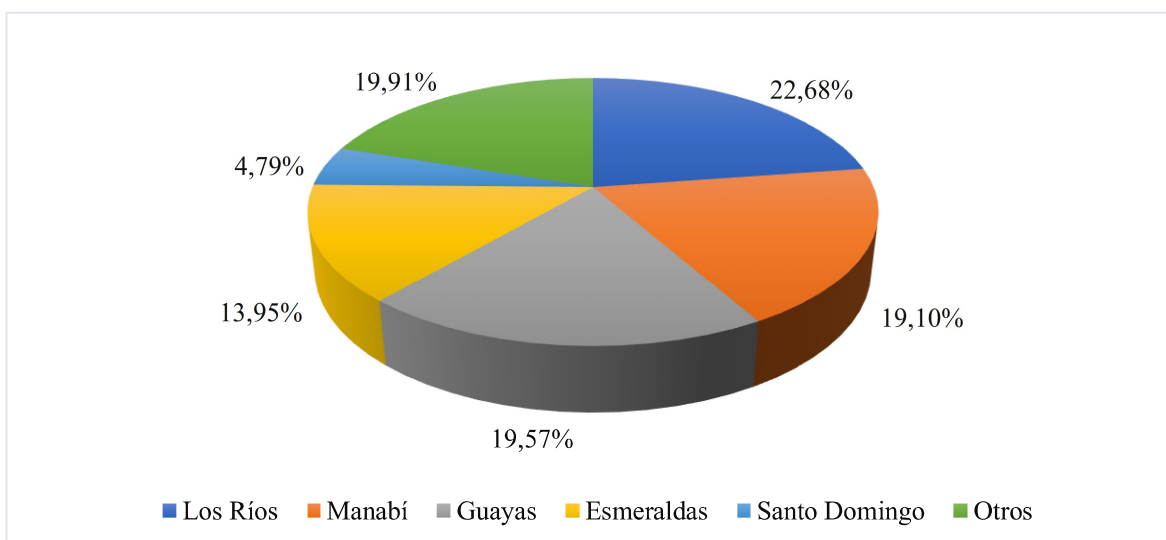
De acuerdo con la Organización de las Naciones Unidas para la Alimentación y la Agricultura, FAO (20), el cacao en el contexto económico ecuatoriano es un factor muy influyente, siendo una de las actividades más importantes en la actividad agrícola del país, considerado como el principal país productor de América y el cuarto a nivel mundial, permitiendo generar empleo e ingresos; su producción tiene una alta diferencia en relación otras producciones, es decir, se cultiva, cosecha y comercializa con mayor frecuencia, demostrando que este nicho de mercado es rentable en todos sus aspectos.

En Ecuador se produce cacao en 14 provincias. Aproximadamente el 80% de la producción se concentra en la cuenca del río Guayas, por la fertilidad de los suelos. Pero existen otras zonas del país que tienen un cacao diferente y que ofrece otras cualidades, por ejemplo, la provincia de Napo, que produce volúmenes menores y no figura como provincia productora de cacao, pero tiene un cacao diferente (E-7) (21).

Según los datos de la Encuesta de Superficie y Producción Agropecuaria Continua (ESPAC) (22), hasta el año 2020, la superficie plantada de cacao a nivel nacional fue de 590579 hectáreas (has); y la mayor producción se concentra en la provincia de Los Ríos con el 28.36% (167488.20 has). La participación en la superficie total de la plantación de cacao, se representa en la Figura 1.

Figura 1

Participación de las provincias en la superficie plantada total de cacao



Fuente: INEC, (1)

2.2.2.1. *Variedades.*

Sánchez-Ramírez (17) detalla que existen tres variedades principales de cacao las cuáles son:

- **Cacao forastero:** Su nombre se debe a que es originario de África, porque “venía de afuera”. Es el más cultivado en el mundo, abarcando un total de 80% sobre la producción mundial de cacao. Se caracteriza por su aroma ácido y sabor astringente.
- **Cacao trinitario:** Es una cepa híbrida que se originó al sembrar cacao forastero en suelos donde habían muerto árboles de cacao criollo, por esto tiene características de las dos variedades. Su nombre se debe a que se originó en la isla de Trinidad, y ocupa del 10% al 15% de la producción mundial.
- **Cacao criollo:** Siendo bautizado por los españoles como “criollo” por provenir del nuevo mundo. Posee un aroma y sabor superior a las otras variedades; sin embargo, es más difícil de cultivar que los otros cacaos y es más susceptible a las enfermedades y cambios de clima. Representa el 5% de la cosecha mundial aproximadamente.

2.2.3. Aspectos teóricos del Aluminio (Al) y el Cadmio (Cd).

2.2.3.1. *Aluminio.*

El Aluminio es el metal más abundante sobre la corteza terrestre, y el tercer elemento más abundante después del oxígeno y el silicio (23), considerado como un metal que ha contribuido al desarrollo del ser humano a través de la historia, siendo la evolución en las necesidades humanas y tecnológicas lo que permitió el desarrollo de los procesos industriales hacia condiciones inimaginables y sorprendentes (24), siendo conocido por primera vez en 1808, pero, no fue hasta 1827 que se le atribuyó el descubrimiento a Frederich Wöler (25).

El Aluminio se usa en forma pura, aleado con otros metales o en compuestos no metálicos. Su uso más popular es como papel Aluminio, el cual consiste en una lámina con un espesor tan delgado que resulta fácilmente maleable (26).

2.2.3.2. *Cadmio.*

El Cadmio es un metal pesado considerado como uno de los elementos más tóxicos. Se encuentra ampliamente distribuido en la naturaleza, asociado principalmente a los depósitos de Zinc (27). Este metal se encuentra presente en el suelo como resultado de la meteorización de las rocas y erupciones volcánicas, fosfatos marinos y por aportes de actividades humanas

(explotación minera, la fundición y refinación de metales no ferrosos, la quema de combustibles fósiles, la aplicación de fertilizantes fosfatados y la quema de residuos industriales y urbanos, considerándolo un tóxico y riesgos a la salud humana, especialmente a los niños (28).

2.2.4. Efectos de metales pesados en el cultivo de cacao.

Huaraca *et al.* (29), afirman que el problema ambiental sobre el Cd, al ser un metal altamente tóxico, con un alto poder de persistir en el suelo por periodos largos a causa de la mínima pérdida microbiana y química; respecto a lo mencionado, Chávez *et al.* (30) sostienen que el cacao puede absorber el Cadmio presente en el suelo y posteriormente acumularlo en sus almendras.

Las plantas expuestas a altos niveles de Cd les causan la reducción en la fotosíntesis, la absorción de agua y la absorción de nutrientes y, en consecuencia, se observa clorosis, inhibición del crecimiento, pardeamiento de las puntas de las raíces y, finalmente, la muerte (31). La gravedad de este metal depende de su toxicidad específica, bioacumulación, persistencia y no biodegradabilidad (32).

El Cadmio se absorbe en cultivos de cacao por algunos factores como: temperatura del suelo (a mayor temperatura existe más velocidad de metabolismo de reacción, ya que existe una mayor solubilidad), pocos minerales en la tierra (mientras más minerales, menor interacción de Cadmio), acidez (en suelos alcalinos no se ha encontrado mayor incidencia de Cadmio), y presencia de Cadmio de forma natural (33).

En China se hizo un estudio que comprobó que existen variedades de cacao que acumulan Cadmio en sus raíces o semillas y por esta razón si la variedad con la que se trabaja hiperacumula en la semilla debe ser usada para la fabricación de biodiesel y en caso contrario, que la acumulación ocurra en las raíces ésta debe ser usada en la fabricación de aceites vegetales comestibles (34).

2.2.5. Estudios efectuados de metales pesados en cultivo de cacao.

El desarrollo de la presente investigación partió de artículos relacionados a los metales pesados y su relación con el cacao. La búsqueda bibliográfica fue realizada en diferentes bases electrónicas a partir de trabajos realizados en los últimos 5 años.

Charrupi y Martínez (35), describen que el Cadmio en las plantas de cacao en proporciones ideales libera una serie de reacciones metabólicas positivas, como inducir la a diferentes tipos de expresión génica e incrementa la actividad de enzimas antioxidantes como las peroxidasas (III) y la superóxido dismutasa (SOD) las que ayudan a hacer frente al estrés oxidativo ocasionado por los radicales libres, ayudando a prolongar su vida.

Jaramillo *et al.* (36), muestrearon y analizaron suelos de terrenos agrícolas de seis organizaciones que conforman la asociación Unión de Organizaciones Campesinas Cacaoteras (UNOCACE), observando que las concentraciones más altas de Cadmio estaban en las muestras en rango de profundidades más superficiales (0-20 y 20-40 cm), lo que hace presumir que el metal se encuentra asociado a la materia orgánica.

Por su parte, Chancay *et al.* (37) determinaron que, a nivel global y local se identifica la presencia del Cd en plantas de cacao que supone un problema no sólo medioambiental (en relación agua-suelo), sino que, también presume un factor más en la degradación de los suelos y, por tanto, una amenaza para la producción agrícola y la seguridad alimentaria. La presencia del Cd es un problema potencial que limita su comercialización y posibilidades de exportación. América Latina, en especial Ecuador, se ve particularmente afectada por los preocupantes niveles de Cadmio en los granos de cacao.

Arévalo *et al.* (38), determinó que las plantaciones de cacao con diferentes clones presentaron diferencias en la acumulación de Cd tanto en hojas como en granos de cacao; por lo tanto, seleccionar genotipos de cacao con bajo acumulador de Cd para una producción segura de cacao en suelos contaminados de leve a moderadamente con Cd. Además, la sinergia entre el Zn y el Cd presentes tanto en la planta como en el suelo sugiere que el Zn tiene un efecto directo sobre la acumulación de Cd en el cacao.

CAPÍTULO III

METODOLOGÍA DE LA INVESTIGACIÓN

3.1. Localización.

La presente investigación se efectuó en el Campus Experimental “La María”, Facultad de Ciencias Pecuarias y Biológicas perteneciente a la Universidad Técnica Estatal de Quevedo, ubicada en el kilómetro 7.5 de la vía Quevedo - El Empalme, Recinto San Felipe, Cantón Mocache, Los Ríos.

3.1.1. Condiciones agroclimáticas

Las condiciones meteorológicas aproximadas del lugar de investigación se describen en la Tabla 1.

Tabla 1

Condiciones meteorológicas aproximadas del Campus Experimental “La María”

Características	Valores aproximados
Temperatura	Entre 18 a 24 °C
Precipitaciones	2583 mm/año
Altitud	74 m.s.n.m.
Humedad relativa	73 a 74%

Fuente INIAP (39)

3.2. Tipo de investigación.

El presente trabajo de investigación de carácter bioinformático para elucidar los genes involucrados en acumulación y detoxificación de los metales pesados en el cultivo de cacao.

3.3. Métodos de investigación.

El presente estudio es de carácter científico casi experimental *in silico*, no posee un diseño experimental con factores, repeticiones o tratamientos, más bien se basa en el uso de algoritmos fuertemente creados desde el punto de vista matemático para análisis bioinformáticos. En ese sentido tenemos al algoritmo BLAST (Basic Local Alignment Search Tool), en este caso insertado como una función de Orthovenn2, el mismo que es un algoritmo y un programa informático de alineamiento de secuencias de tipo local, ya sea de ADN o de proteínas, empleado en bioinformática.

3.4.Fuentes de recopilación de la información.

Para la presente investigación se utilizó como fuente primaria la información obtenida a través de la recolección de datos provenientes del trabajo de campo; mientras que las fuentes secundarias fueron libros, fuentes de internet, revistas científicas, tesis e informes de instituciones oficiales del estado.

3.5.Diseño de la investigación.

Al ser un trabajo de carácter bioinformático, no existió un diseño experimental como tal, es decir no se realizó manipulación de variables.

3.6.Manejo del experimento.

Se compararon los genomas secuenciados del cultivo de (*Theobroma cacao*) con otras especies de las *Malvaceae* y los *metallophytes* con el fin de revelar datos genómicos relacionados a factores genéticos en los procesos evolutivos de ambas especies, adicionalmente, caracterizar y elucidar de forma *in silico* proteínas relacionadas con características de resistencia por contamina

ción de metales pesados como Cadmio y Aluminio, utilizando el software Orthovenn2. Para obtener los genomas completos con sus respectivas anotaciones, se procedió a descargar todo el conjunto de datos (genomas completos, nucleótidos, proteínas y transcritos), desde la plataforma NCBI o Phytozome bajo las accesiones de ensamblaje correspondientes.

El programa es capaz de comparar una secuencia problema contra una gran cantidad de secuencias que se encuentren en una base de datos y encontrar las que tienen mayor parecido, así como la significancia de cada resultado, el nivel de significancia escogido para encontrar la similitud de las secuencias de proteínas fue de 1×10^{-5} .

El plan de búsqueda de secuencias genómicas de proteínas diseñado para la presente investigación se detalla a continuación:

Figura 2

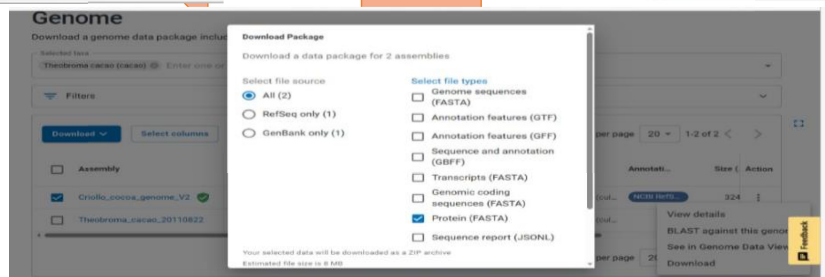
Esquema de búsqueda de secuencias de genómicas de proteínas en la plataforma NCBI(National Center for E

1.- Introducir nombre científico de la especie en estudio



Esquema de obtención de Secuencias de Proteínas

4.- Resultado: descarga de secuencias de proteínas del Genoma de *Theobroma cacao* para los análisis bioinformáticos



Fuente: Autora

Figura 3

Esquema de búsqueda de secuencias de genómicas de proteínas en la plataforma.

1.- Búsqueda sobre Phytozome el genoma de *A. halleri*

2.- Genoma de *A. halleri*

Esquema de obtención de Secuencias de Proteínas

5.- Archivo de proteínas de *A. halleri* en formato FASTA

The screenshot shows the Phytozome 13 website. The search bar at the top contains 'Arabidopsis halleri'. The search results show 'Arabidopsis halleri v1.1' as the selected genome. The 'Download' button is highlighted. A large orange cross is overlaid on the page, indicating the search process. The text 'Esquema de obtención de Secuencias de Proteínas' is written across the cross. The text '1.- Búsqueda sobre Phytozome el genoma de A. halleri' is in a box on the left. The text '2.- Genoma de A. halleri' is in a box on the right. The text '5.- Archivo de proteínas de A. halleri en formato FASTA' is in a box at the bottom left. The text 'Fueron: Autora' is at the bottom left of the page.

Fueron: Autora

Para comparar los genomas secuenciados de cacao con otras especies de las *Malvaceae* y los *metallophytes* con el fin de revelar datos genómicos relacionados a factores genéticos en los procesos evolutivos de ambas especies, adicionalmente caracterizar y elucidar de forma in silico familia de genes, factores de transcripción y proteínas relacionados con efectos de metales pesados varios softwares disponibles de forma gratuita en la web.

Para tener el genoma de la especie bajo estudio (*Theobroma cacao* L.) se procedió a descargar todo el conjunto de datos (genomas completos, nucleótidos, proteínas y transcritos) de cacao, desde la plataforma NCBI o Phytozome bajo las accesiones de ensamblaje GCF_000208745.1 (Criollo_cocoa_genome_V2) y GCF_002168275.1 (ASM216827v2), respectivamente.

Para comparar especies entre sí, los transcriptos y los conjuntos de datos de proteínas se eliminaron teniendo en cuenta el umbral de corte de Evalue $10E-10$ como criterio para filtrar los resultados. Tres búsquedas de BLAST; es decir, blastn, blastp y blastx se realizaron para evaluar las diferencias genómicas y cómo varían sus resultados.

El software OrthoVenn2 fue utilizado para comparar conjuntos de datos de proteínas y generar mapas para la comparación de identidad entre especies bajo estudio. OrthoVenn2 es una plataforma web en línea que compara conjuntos de datos de proteínas, los agrupa y genera anotaciones genéticas, mapas y diagramas de Venn basados en similitudes entre las especies de entrada.

Para comparar los factores de transcripción (TF) a nivel de especie, el conjunto de datos de Factores de Transcripción (FT) de *Theobroma cacao* se descendieron de la plataforma PlantTFDB. Los transcritos y el conjunto de datos de proteínas de *Herrania umbratica* se blastearon contra el conjunto de datos FT de cacao.

Los resultados serán filtrados por Evalue cutoff $10E-10$. PlantTFcat es una herramienta para analizar la identificación y categorización de genes de plantas TF / TR / CR a escala genómica. Los microsatélites se utilizan como marcadores moleculares en fitomejoramiento, genética de poblaciones y estudios filogenéticos. Son repeticiones en tándem de motivos de nucleótidos cortos en secuencias genómicas. FullSSR es una nueva herramienta bioinformática para la detección de loci de microsatélites (SSR) y el diseño de cebadores que utilizan datos genómicos del ensayo NGS. FullSSR permitió crear marcadores moleculares, los cuales son un buen medio para usarse en fitomejoramiento y estudios filogenéticos.

3.7.Tratamientos de los datos de secuencias de datos.

Los datos de las secuencias recolectadas fueron sistematizados y ordenados en tablas con la ayuda del programa Microsoft Excel.

3.8. Recursos humanos y materiales.

El recurso humano que intervienen en el desarrollo del presente proyecto de investigación fue, el Dr. Ronald Oswaldo Villamar Torres, quien prestó su contingente como director del Proyecto de Investigación, mientras que como autora del proyecto de investigación se presenta la Srta. Egresada. Yuliana Stefania Campuzano Herrera. Como un aporte externo se contó con la participación de un Investigador internacional, especialista en Bioinformática, el Dr. Seyed Mehdi Jazayeri, de la Universidad de Marsella y el CIRAD de Francia.

3.8.1. Materiales.

3.8.1.1. Materiales de oficina.

- Computadora
- Impresora
- Calculadora
- Lápiz
- Cuaderno
- Servicios de internet
- Hojas A4

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1. Comparación de clúster de proteínas de las 4 especies.

La Tabla 2 muestra que la versión del genoma *Criollo_cocoa_genome_V2* es la que presenta un mayor tamaño por pares de bases con 324,7 Mb, y el genoma con menor tamaño fue el de *Arabidopsis.thaliana* (TAIR10.1) con 119,1 Mb. Con respecto a la anotación, se obtuvo que 3 de los 4 genomas presentan anotaciones procedentes de NCBI RefSeq, y tan solo uno que no se muestra disponible en esta plataforma.

Por lo tanto fue obtenido a partir de la plataforma Phytozome, con su anotación Phytozome v1.1 RefSeq. De acuerdo a los resultados mostrados las 4 últimas versiones de los genomas bajo estudio fueron actualizados entre agosto 2016 – julio 2017.

A continuación, las secuencias fueron descargadas en formato FASTA, las mismas que posteriormente fueron guardadas en un servidor local de uso personal. El análisis de la comparación de proteínas entre las cuatro especies bajo estudio fue realizado en la plataforma OrthoVenn2 disponible en el sitio web <https://orthovenn2.bioinfotoolkits.net/home>.

Este software es una herramienta gráfica eficiente e interactiva que proporciona una vista de diagrama de Venn para comparar secuencias de proteínas de múltiples especies. Lo único que debe hacer el usuario es elegir especies o cargar secuencias de proteínas.

Teniendo en cuenta eso, preestablecimos una limitación en la versión web para las comparaciones dentro de los clados. Sin embargo, tales limitaciones podrían eludirse mediante el uso de datos de clúster definidos por el usuario o mediante el uso de la versión independiente.

OrthoVenn es una plataforma web para la comparación y anotación de grupos de genes ortólogos entre múltiples especies. Las comparaciones se realizaron para comprender mejor cómo se agrupan las proteínas entre las 4 especies considerando la ortología. OrthoVenn2 fue utilizado para comparar los 4 conjuntos de datos de las proteínas genómicas de cada especie en estudio.

Todos los parámetros predeterminados fueron considerados para el análisis, con excepción del valor-E que se fijó en 1E-05 como valor único para los análisis Figura 4.

Tabla 2

Ensamblamientos completos de los genomas de Cacao criollo, Herrania umbrática, Arabidopsis haliana y Ara referencias, tipo de anotación, tamaño y fecha de carga sobre las plataformas NCBI y Phytozome.

	Ensamblamiento	GenBank	RefSeq	Nombre científico	Anotación
1	Criollo_cocoa_genome_V2	GCA_000208745.2	GCF_000208745.1	<i>Theobroma cacao</i> (cacao)	NCBI RefS
2	ASM216827v2	GCA_002168275.2	GCF_002168275.1	<i>Herrania umbratica</i>	NCBI RefS
3	TAIR10.1	GCA_000001735.2	GCF_000001735.4	<i>Arabidopsis thaliana</i> (thale cress)	NCBI RefS
4	Ahalleri_264_v1.1. protein	GCA_900078215.1	JGIAP: 1355409	<i>Arabidopsis halleri</i>	Phytozome RefSeq

Fuente: Autora

4.1.1. Secuencias de proteínas.

Figura 4

Resultado de la carga a la plataforma Orthovenn 2 de las secuencias de proteínas de las cuatro especies: *Theobroma cacao* (cacao), *Herrania umbrática*, *Arabidopsis thaliana* (thale cress) y *Arabidopsis halleri*, y criterios de análisis bioinformáticos para la comparación genómica.

The screenshot displays the Orthovenn 2 web interface. The top section lists four uploaded protein sequence files, each with a file icon, a name, a FASTA path, and a close button (X):

- 1. Theobroma_cacao_L
fasta: GCF_000208745.1_Criollo_cocoa_genome_V2_protein.fa
- 2. Herrania_umbratica
fasta: GCF_002168275.1_ASM216827v2_protein.fa
- 3. Arabidopsis_halleri
fasta: Ahalleri_264_v1.1.protein.fa
- 4. Arabidopsis_thaliana
fasta: Athaliana_167_TAIR10.protein.fa

Below the list are buttons for "Upload protein sequences", "Load Example", and "Download Example file". The bottom section contains analysis parameters:

- E-value: 1e-5 (dropdown menu)
- Inflation value: 1.5 (text input)
- Annotation: ☒ Enable Annotation
- Protein similarity: ☒ Protein similarity network

Fuente: Autora

4.1.2. Análisis de secuencias de proteínas basada en OrthoVenn2

La Tabla 3 muestra los resultados de la comparación de las secuencias de proteínas por método de agrupación (clustering) con base en la herramienta OrthoVen. En total 120899 proteínas se obtuvieron como resultado entre las cuatro especies estudiadas.

Tabla 3

Número de proteínas, grupos y datos únicos de cada conjunto de datos de proteínas de las 4 especies.

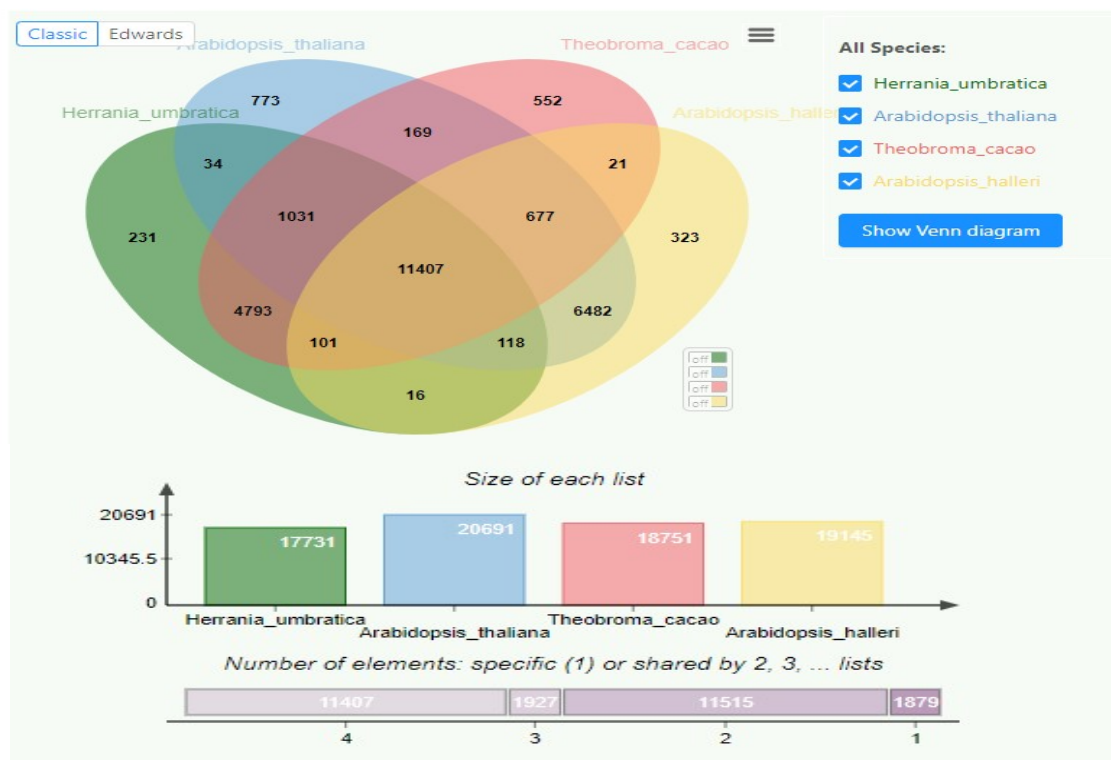
Species	Proteins	Clústers	Singletons
<i>Herrania umbrática</i>	27748	17731	2107
<i>Theobroma cacao</i>	30854	18751	2619
<i>Arabidopsis thaliana</i>	35386	20691	5457
<i>Arabidopsis halleri</i>	26911	19145	4759
Total	120899	76318	14942

Fuente: Autora

En cuanto a clústers 76318 fueron formados, mientras que 14942 Singletons de igual manera para las cuatro especies. Un singleton es una lectura con una secuencia que está presente exactamente una vez, es decir, es única entre las lecturas. La mayor cantidad de proteínas y singletons caracterizadas en el análisis se reportaron para *Arabidopsis thaliana*, así mismo la especie *Theobroma cacao* presentó la mayor cantidad de clúster.

Figura 5

Diagrama de Venn de comparación entre 4 especies. 11407 grupos fueron comunes entre las 4 especies. 552 grupos son exclusivamente de *T. cacao*, 231 grupos para *H. umbratica*, 240 grupos para *C. canephora* y *C. eugenioides* y 111 para *C. arabica*, *C. eugenioides* y *T. cacao*.



Fuente: Autora

La Figura 5 muestra que las proteínas entre las 4 especies se agrupan en 11407 clústers de proteínas, incluyendo 63993 de proteínas que se distribuyen entre 4 especies como 16881 de *Arabidopsis thaliana*, 13266 de *A. halleri*, 16911 de *Theobroma cacao* y 16935 de *Herrania umbrática*. 1927 fuera de 26728 clusters se comparten entre tres especies, mientras que 11515 clusters son comunes entre dos especies.

finalmente 1879 clúster son las proteínas específicas de especie que son 231 proteínas para *Herrania umbrática*, 552 para *T. cacao*, 773 para *Arabidopsis thaliana* y 323 para *Arabidopsis halleri*. específicamente entre las dos especies de cacao, 4793 clústers incluyendo 13273 proteínas son compartidos en esta manera que 6465 proteínas perteneces a *Herrania umbrática* y 6808 proteínas a *Theobroma cacao*. Sumando las

proteínas compartidas entre dos especies de cacao que también se encuentran en una especie de *Arabidopsis* 1031 y 101 clústers de proteínas se agrupan.

Esto quiere decir que estas proteínas son comunes entre tres especies, dos especies de cacao y una de *Arabidopsis*. Considerando las proteínas entre dos cacaos y *Arabidopsis halleri* 154 proteínas de *Herrania umbrática*, 278 proteínas de *T. cacao* y 124 proteínas de *Arabidopsis halleri* son compartidas en 101 clústers.

De otro lado 1426 proteínas de *Theobroma cacao*, 1307 proteínas de *Herrania umbrática* y 1458 proteínas de *Arabidopsis thaliana* se comparten en 1031 clúster. Por lo tanto, en total entre dos especies de cacao 17332 clústers son compartidos.

4.1.3. Resultados de análisis de proteínas en Orthovenn2.

Tabla 4

Secuencias de proteínas relacionadas a la resistencia a Aluminio (Al) encontradas por similitud.

Número de Secuencia	Id_Secuencia	Secuencias
1	>Arabidopsis_thaliana AT5G62090.1	MASSTSGIFFQGDEESQSFINSHLTSSYGNSNSAPGCGG PTGGYHNLSM
2	>Arabidopsis_thaliana AT5G62090.2	MASSTSGIFFQGDEESQSFINSHLTSSYGNSNSAPGCGG PTGGYHNLSM
3	>Arabidopsis_halleri Araha.4168s0002.1.p	MASSTSGIFFQGDEESQSFINSHLTSSYGNSNSVPGCGP TGGYHNLSMV
4	>Herrania_umbrática XP_021298120.1	MAPSRVAGGLTQSSSSSGIFFQGDGQSQAVVNSRLSSPYE NSSNSIPGTG
5	>Herrania_umbrática XP_021298050.1	MAPSRVAGGLTQSSSSSGIFFQGDGQSQAVVNSRLSSPYE NSSNSIPGTG
6	>Theobroma_cacao XP_017972890.1	MAPSRVAGGLTQSSSSSGIFFQGDGQSQAVVNSRLSSPYE NSSNSIPGTG
7	>Theobroma_cacao XP_017972889.1	MAPSRVAGGLTQSSSSSGIFFQGDGQSQAVVNSRLSSPYE NSSNSIPGTG
8	>Herrania_umbrática XP_021298284.1	MRLPASPMSFSSNNISMSGSSVVDGSSIGQQGSHQDPSVQ QMQQSQQLQQ
9	>Herrania_umbrática XP_021298201.1	MRLPASPMSFSSNNISMSGSSVVDGSSIGQQGSHQDPSVQ QMQQSQQLQQ
10	>Arabidopsis_thaliana AT1G51340.2	MMSEEDGYNTDFPRNPLYIFFSDFRSVLKFDELGLEIARIA LPAALALTAD
11	>Arabidopsis_thaliana AT1G51340.1	MATTQIFQETLYTFSLVISVLKFDELGLEIARIALPAALA LTADPIASLV
12	>Arabidopsis_halleri Araha.13904s0002.1.p	MMSENVYNTDFPRNPLYIFFSDFRSVLKFDELGLEIARIA YPAALALTAD
13	>Herrania_umbrática XP_021278147.1	MAEEDDPYLSRVKMRLPIFFKDVNRNVFKLDDLGEIAQ IALPAALALT
14	>Theobroma_cacao XP_017983433.1	MMAEEDDPYLSRVKMRLPIFFKDVNRNVFKLDDLGEIA QIALPAALAL
15	>Theobroma_cacao XP_017983432.1	MMAEEDDPYLSRVKMRLPIFFKDVNRNVFKLDDLGEIA QIALPAALAL
16	>Theobroma_cacao XP_017983431.1	MMAEEDDPYLSRVKMRLPIFFKDVNRNVFKLDDLGEIA QIALPAALAL
17	>Theobroma_cacao XP_007015048.2	MMAEEDDPYLSRVKMRLPIFFKDVNRNVFKLDDLGEIA QIALPAALAL
18	>Theobroma_cacao XP_017983434.1	MMAEEDDPYLSRVKMRLPIFFKDVNRNVFKLDDLGEIA QIALPAALAL
19	>Arabidopsis_thaliana AT1G34370.1	METEDDLCNTNWGSSSSKSPREPSSDCGNSTFAGFTSQK WEDASILDYE

20	>Arabidopsis_thaliana AT1G34370.2	METEDDLCNTNWGSSSSKSPREPSSDCGNSTFAGFTSQQK WEDASILDYE
21	>Arabidopsis_halleri Araha.5271s0005.1.p	METEDDLCNNWSSSSKSPREPSSDGGNQTFAFTSQQK WEDASILDYE
22	>Herrania_umbratica XP_021300128.1	MDLKERLSGELQKKVPEHPSFPNFNTQHQQKWEDPSILDY STRIEPPFHV
23	>Theobroma_cacao XP_007020278.1	MDLKERLSGELQKKVPEHPSFTNFNTQHQQKWEDPSILDY STRIEPPFHV
24	>Arabidopsis_thaliana AT5G39040.1	MGNKKLLTGGSSKTHGSGSSYRDPLLQNQEDKPKANGSEN GLNDLEHGVV
25	>Arabidopsis_halleri Araha.1571s0002.1.p	MHQEIAFYDVTKTGELLRLSEDTQIIKNAATTNLSEALR NVTALIGVG
26	>Theobroma_cacao XP_007032789.2	MGKKQHLDTWSKSISSSGSNRVPLLYKEKSKQANDHQSPN GTVSDLEQGD
27	>Herrania_umbratica XP_021297554.1	MLILQPKYGGMIIDIVSRDIRTPEQQSDALKAVKDTVIYI VLIVVVGSLC
28	>Arabidopsis_thaliana AT5G39040.1	MGNKKLLTGGSSKTHGSGSSYRDPLLQNQEDKPKANGSEN GLNDLEHGVV
29	>Arabidopsis_halleri Araha.1571s0002.1.p	MHQEIAFYDVTKTGELLRLSEDTQIIKNAATTNLSEALR NVTALIGVG
30	>Theobroma_cacao XP_007032789.2	MGKKQHLDTWSKSISSSGSNRVPLLYKEKSKQANDHQSPN GTVSDLEQGD
31	>Herrania_umbratica XP_021297554.1	MLILQPKYGGMIIDIVSRDIRTPEQQSDALKAVKDTVIYI VLIVVVGSLC
32	>Arabidopsis_thaliana AT2G37330.1	MDLKWDDFNDYEWLIVFLKGMVKPAAALVVLAVILSY SQNLSLEGEM
33	>Arabidopsis_halleri Araha.14169s0003.1.p	DDDKTKVAIDREMDLYWEDFKDYEWLIVFLKGMVKPAAA LVVLLAVVL
34	>Herrania_umbratica XP_021290008.1	MDWDWLEFVKGMKPAALSVVLMVLLSFMQKLGLEKE MIYSILRAFL
35	>Theobroma_cacao XP_007027163.1	MDWDWLEFELKGMKPAALSVVLMVLLSFMQKLGLEKE MIYSILRAFL

Fuente: Autora

Tabla 5

Secuencias de Proteínas relacionadas a la resistencia a Cadmio (Cd) encontradas por similitud.

Numero de Secuencia	Id_Secuencia	Secuencias
1	>Arabidopsis_thaliana AT1G15990.1	MMQRNCFGNLKNRGGEKKKASKSFREGVKIRSEGLITI GKSVTRAVF
2	>Arabidopsis_thaliana AT1G19780.1	MYKSQYISGHREKFVRLDDTDSRVSMSSNATGMKKRSCFGL FNVTSRGGG
3	>Arabidopsis_thaliana AT4G30560.2	MDSRYQTSQDGLNRCTNLQGPTRGGAQGNVSSGSFVK GFRKSGKGL
4	>Arabidopsis_thaliana AT2G23980.1	MFDTGCGPKGVKSQVISGQRENFVRLDSMDSRYQSSETGLN KCTLNIQGG
5	>Arabidopsis_thaliana AT4G30560.1	MLDCGKKAVKSQVISGRLEKFVRLDSMDSRYQTSQDGLNR CTLNLQGP
6	>Arabidopsis_halleri Araha.1835s0004.1.p	MYKSQYIGGQREKFVRLDDTDSRVSMSSNATGMKKRSCFGL FNLSSRGGG
7	>Arabidopsis_halleri Araha.6923s0001.1.p	MFDTGCGPKGVKSQVISGQREKFVRLDSMDSRYQSSETGLN KCTLNIQGP
8	>Arabidopsis_halleri Araha.2161s0004.2.p	MDSRYQTSQDAGLNRCNLQGPTRGGAQGNVSSGSFQKG FRKGSRLGW
9	>Arabidopsis_halleri Araha.2161s0004.1.p	MLDCAKKAVKSQVISGHLEKFVRLDSMDSRYQTSQDAGLN CTLNLQGP
10	>Herrania_umbratica XP_021277680.1	MFDCGYKAQYMSGQREKFVRLDDLSRLSSPSDAGLKNCGF NIEGLNRSG
11	>Theobroma_cacao XP_007011917.2	MFDCGYKAQYMSGQREKFVRLDDLSRLSSPSDAGLKNCGF NIEGFNRSG
12	>Herrania_umbratica XP_021298116.1	MFDCAYKSQLIGGQKEKFVRLDDLSMTSMQSGTGMKKCR FNIEGLTFG
13	>Theobroma_cacao XP_007023447.2	MFDCAYKSQLIGGQKEKFVRLDDLSMTSMQSGTGMKKCR FNIEGLTFG
14	>Arabidopsis_thaliana AT5G57940.2	MAGKRENFVRVDDLSRLPSSSVAFQQNYASNSGQLHPIH ASNETSRSF

15	>Arabidopsis_thaliana AT5G57940.1	MAGKRENFVRVDDLDSRLPSSSVAFQQNYASNFSGQLHPIH ASNETSRSF
16	>Arabidopsis_thaliana AT5G57940.3	MAGKRENFVRLPSSSVAFQQNYASNFSGQLHPIHASNETSR SFKKGIQKG
17	>Arabidopsis_thaliana AT4G14880.2	MASRIAKDVTELGNTPLVYLNVAEGCVGRVAAKLEMMEP CSSVKDRIG
18	>Arabidopsis_thaliana AT4G14880.1	MASRIAKDVTELGNTPLVYLNVAEGCVGRVAAKLEMMEP CSSVKDRIG
19	>Arabidopsis_thaliana AT4G14880.3	MASRIAKDVTELGNTPLVYLNVAEGCVGRVAAKLEMMEP CSSVKDRIG
20	>Arabidopsis_thaliana AT4G14880.4	MASRIAKDVTELGNTPLVYLNVAEGCVGRVAAKLEMMEP CSSVKDRIG
21	>Arabidopsis_halleri Araha.9544s0006.1.p	MASGTSIAKDVTELGNTPLVYLNKAKDCVGHVAAKLEMM EPCSSVKDR
22	>Herrania_umbratica XP_021282092.1	MAEKSLIAKDVTELVGKTPLVYLNNVVDGCVARIAAKLEAM EPCSSVKDR
23	>Herrania_umbratica XP_021282091.1	MAEKSLIAKDVTELVGKTPLVYLNNVVDGCVARIAAKLEAM EPCSSVKDR
24	>Theobroma_cacao XP_017974671.1	MAHESVGIKDVTELGKTPLVYLNNVVDGCVARIAAKLEI MEPCSSVKD
25	>Theobroma_cacao XP_007034764.1	MAHESVGIKDVTELGKTPLVYLNNVVDGCVARIAAKLEI MEPCSSVKD
26	>Theobroma_cacao XP_007030649.2	MEEKSLIAKDVTELVGKTPLVYLNNVVDGCVARIAAKLEAM EPCSSVKDR
27	>Theobroma_cacao XP_007030650.2	MEEKSLIAKDVTELVGKTPLVYLNNVVDGCVARIAAKLEAM EPCSSVKDR
28	>Arabidopsis_thaliana AT5G28030.1	MEDRVLIKNDVTELGNTPMVYLNKIVDGCVARIAAKLEMM EPCSSIKDR
29	>Arabidopsis_thaliana AT5G28030.2	MEDRVLIKNDVTELGNTPMVYLNKIVDGCVARIAAKLEMM EPCSSIKDR
30	>Arabidopsis_halleri Araha.3433s0002.1.p	MEERISLIKNDVTELGNTPMVYLNKIVDGCVAHIAAKLEM MEPCSSIKD
31	>Arabidopsis_thaliana AT3G04940.1	MEEDRCSIKDDATQLIGNTPMVYLNINVDGCVARIAAKLEM MEPCSSVKE
32	>Arabidopsis_halleri Araha.6236s0008.1.p	MEEDRCSIKDDATQLIGNTPMVYLNINVDGCVARIAAKLEM MEPCSSVKE
33	>Herrania_umbratica XP_021281719.1	MEEKCEIKKDVTELGNTPMVYLNINVDGCVARIAAKLELL EPCSSVKDR
34	>Herrania_umbratica XP_021281718.1	MEEKCEIKKDVTELGNTPMVYLNINVDGCVARIAAKLELL EPCSSVKDR
35	>Herrania_umbratica XP_021281716.1	MEEKCEIKKDVTELGNTPMVYLNINVDGCVARIAAKLELL EPCSSVKDR
36	>Herrania_umbratica XP_021281715.1	MEEKCEIKKDVTELGNTPMVYLNINVDGCVARIAAKLELL EPCSSVKDR
37	>Theobroma_cacao XP_017975971.1	MEEKCAIKKDVSELIGNTPMVYLNINVDGCVARIAAKLELL EPCSSVKDR
38	>Arabidopsis_thaliana AT2G39920.3	MSAYTHPMERELSGLSRRGNSELGSRYIESGCMYTSLAAS IFIASLVTF
39	>Arabidopsis_thaliana AT2G39920.1	MSAYTHPMERELSGLSRRGNSELGSRYIESGCMYTSLAAS IFIASLVTF
40	>Arabidopsis_thaliana AT2G39920.2	MSAYTHPMERELSGLSRRGNSELGSRYIESGCMYTSLAAS IFIASLVTF
41	>Arabidopsis_halleri Araha.25214s0001.1.p	MSAYAHQMEREFSGLSRRGNSELGSRYMESGCMYTSLAAS IFIASLVTF
42	>Theobroma_cacao XP_017985146.1	MSAYGHQMERQFSAQSLLSRGDTGDMGSRYLVESGFYMTS FAATIFIAS
43	>Theobroma_cacao XP_017985144.1	MSAYGHQMERQFSAQSLLSRGDTGDMGSRYLVESGFYMTS FAATIFIAS
44	>Herrania_umbratica XP_021295353.1	MSAYGHQMEREFSAQSLLSRGDTGDMGSRHVESGFYMTSFA GTIFIAGLA
45	>Theobroma_cacao XP_017985147.1	MSAYGHQMERQFSAQSLLSRGDTGDMGSRYLVESGFYMTSFA ATIFIASLA
46	>Herrania_umbratica XP_021295352.1	MSAYGHQMEREFSAQSLLSRGDTGDMGSRHVESGFYMTS FAGTIFIAAG
47	>Herrania_umbratica XP_021295351.1	MSAYGHQMEREFSAQSLLSRGDTGDMGSRHVESGFYMTS FAGTIFIAAG
48	>Arabidopsis_thaliana AT1G72270.1	MVAGKRKSDRNPTNVSDKLGKDLLRRLSSAAVTNSFLQ ANDLFLSPS
49	>Arabidopsis_thaliana AT4G27010.1	MASEIAKKFDFKGFALAEYNTQGTEKVKKHSTRKAFVGFA ISFLEVGP

50	>Arabidopsis_thaliana AT4G27010.2	MASEIAKKFDFKGFALAEYNTQGTEKVKKHSTRKAFVGFASISFLEVGKP
51	>Arabidopsis_halleri Araha.8775s0007.1.p	DLVSSVRISSEDFLNPDPFPGGSEVHTIMKCICPRPFSRLLIVRGMQH
52	>Arabidopsis_halleri Araha.21176s0005.1.p	MVVDGGNRDVEVENIPVMAFRPSHEAKLRELLHKICLYEIKLSSDAAKEF
53	>Arabidopsis_halleri Araha.14919s0001.1.p	MVAKSLTRKGTAMAKCSKGYEPIETYQEALAEHCNPETLKLNEAERP
54	>Herrania_umbratica XP_021287334.1	MMECDSDTESEFEGEEEEGVPKAIVFESSRAAKLKELLHKINSIEIKLFS
55	>Theobroma_cacao XP_017980622.1	MEECIFASDNQVMEGDNDTESEFEGEEEEGVPKTIVFESSRSKALKDLLH
56	>Theobroma_cacao XP_017980621.1	MEECIFASDNQVMEGDNDTESEFEGEEEEGVPKTIVFESSRSKALKDLLH
57	>Arabidopsis_thaliana AT1G53850.1	MFLTRTEYDRGVNTFSPEGRLFQVEYAIEAIKLGSTAGVKTKEGVVLAV
58	>Arabidopsis_thaliana AT1G53850.2	MFLTRTEYDRGVNTFSPEGRLFQVEYAIEAIKLGSTAGVKTKEGVVLAV
59	>Arabidopsis_halleri Araha.13408s0001.1.p	MFLTRTEYDRGVNTFSPEGRLFQVEYAIEAIKLGSTAGVKTKEGVVLAV
60	>Herrania_umbratica XP_021294557.1	MFLTRTEYDRGVNTFSPEGRLFQVEYAIEAIKLGSTAGIKTKDGVVLAV
61	>Arabidopsis_thaliana AT3G14290.1	MFLTRTEYDRGVNTFSPEGRLFQVEYAIEAIKLGSTAGVKTKEGVVLAV
62	>Theobroma_cacao XP_017978938.1	MFLTRTEYDRGVNTFSPEGRLFQVEYAIEAIKLGSTAGIKTKDGVVLAV
63	>Herrania_umbratica XP_021294556.1	MRIFQSRGGDFIVYTLSGESGIEENQLITIKRDMPAQAKRSRFSLACF
64	>Arabidopsis_halleri Araha.1940s0001.1.p	MFLTRTEYDRGVNTFSPEGRLFQVEYAIEAIKLGSTAGVKTKEGVVLAV
65	>Arabidopsis_thaliana AT1G11840.6	MNEIASASMLRLCQCFISICNVHFVSMRAESSFLLSRNMAEASDLEWP
66	>Arabidopsis_thaliana AT1G11840.1	MAEASDLEWPKKDNRRFLHVYRVGDLDRITIEFYTEVFGMKLLRKRDIP
67	>Arabidopsis_thaliana AT1G11840.2	MAEASDLEWPKKDNRRFLHVYRVGDLDRITIEFYTEVFGMKLLRKRDIP
68	>Arabidopsis_thaliana AT1G11840.3	MAEASDLEWPKKDNRRFLHVYRVGDLDRITIEFYTEVFGMKLLRKRDIP
69	>Arabidopsis_thaliana AT1G11840.4	MAEASDLEWPKKDNRRFLHVYRVGDLDRITIEFYTEVFGMKLLRKRDIP
70	>Arabidopsis_halleri Araha.2699s0017.1.p	MAEASAVPNADLLEWPKKDNRRFLHVYRVGDLDRITIAFYTECFGMKLL
71	>Herrania_umbratica XP_021294051.1	MAEGSAVPSTELLEWPKKDKRRFLHAVYRVGDLDRITIKFYTECFGMKLL
72	>Theobroma_cacao XP_007026103.2	MAEGLAVPSTELLEWPKKDKRRFLHAVYRVGDLDRITIKFYTECFGMKLL
73	>Arabidopsis_thaliana AT1G11840.6	MNEIASASMLRLCQCFISICNVHFVSMRAESSFLLSRNMAEASDLEWP
74	>Arabidopsis_thaliana AT1G11840.1	MAEASDLEWPKKDNRRFLHVYRVGDLDRITIEFYTEVFGMKLLRKRDIP
75	>Arabidopsis_thaliana AT1G11840.2	MAEASDLEWPKKDNRRFLHVYRVGDLDRITIEFYTEVFGMKLLRKRDIP
76	>Arabidopsis_thaliana AT1G11840.3	MAEASDLEWPKKDNRRFLHVYRVGDLDRITIEFYTEVFGMKLLRKRDIP
77	>Arabidopsis_thaliana AT1G11840.4	MAEASDLEWPKKDNRRFLHVYRVGDLDRITIEFYTEVFGMKLLRKRDIP
78	>Arabidopsis_halleri Araha.2699s0017.1.p	MAEASAVPNADLLEWPKKDNRRFLHVYRVGDLDRITIAFYTECFGMKLL
79	>Herrania_umbratica XP_021294051.1	MAEGSAVPSTELLEWPKKDKRRFLHAVYRVGDLDRITIKFYTECFGMKLL
80	>Theobroma_cacao XP_007026103.2	MAEGLAVPSTELLEWPKKDKRRFLHAVYRVGDLDRITIKFYTECFGMKLL
81	>Arabidopsis_thaliana AT4G33010.2	MERARRLAYRGIVKRLVNDTKRHRNAETPHLVPHAPARYVSSLSPFISTP
82	>Arabidopsis_thaliana AT2G26080.1	MERARRLAYRGIVKRLVNETKRHRNGESSLLPTTIVTPSRYVSSVSSFLH
83	>Arabidopsis_thaliana AT4G33010.1	MERARRLAYRGIVKRLVNDTKRHRNAETPHLVPHAPARYVSSLSPFISTP
84	>Arabidopsis_halleri Araha.1964s0001.1.p	MERARRLAYRGIVRRLVNDTKRHRNAETPHLVPHAPARYVSSLSPFISTH

85	>Herrania_umbratica XP_021276935.1	MERARRVANRAILKRLVNEAKQSRNGEISSRSPVLTPSRY VSSLSPFGS
86	>Theobroma_cacao XP_007012280.2	MERARRVANRAILKRLVNAAKQSRNGEISSRSPVLTPSRY VSSLSPFGS
87	>Arabidopsis_halleri Araha.23181s0001.1.p	MTAMYAVYHGPEGLKSLAQRVHGLAGVFTLGLKKLGTAQVQ DLPFFDVTVK
88	>Arabidopsis_thaliana AT1G11860.3	MRGGSWLQLGQSITRRLAQSDKKVSSRRYFASEADLKKTAL YDFHVAHGG
89	>Arabidopsis_thaliana AT1G11860.1	MRGGSWLQLGQSITRRLAQSDKKVSSRRYFASEADLKKTAL YDFHVAHGG
90	>Arabidopsis_thaliana AT1G11860.2	MRGGSWLQLGQSITRRLAQSDKKVSSRRYFASEADLKKTAL YDFHVAHGG
91	>Arabidopsis_halleri Araha.2699s0018.1.p	MRGGSWLQLGQSITRRLAQSDKKGVSSRRYFASEADLKKTAL YDFHVAHGG
92	>Herrania_umbratica XP_021294083.1	MRGSLWQLGQSITRRLAQADKKAVTRRYFASEADLKKTAL DFHVAHGGK
93	>Theobroma_cacao XP_017978820.1	MRGSLWQLGQSITRRLAQADKKAITRRYFASEADLKKTAL DFHVAHGGK
94	>Theobroma_cacao XP_007026120.1	MRGSLWQLGQSITRRLAQADKKAITRRYFASEADLKKTAL DFHVAHGGK
95	>Arabidopsis_thaliana AT4G37870.1	MSAGNGNATNGDGGFSFPKGPVMPKITTTGAARKSGVCHDD SGPTVNATT
96	>Arabidopsis_thaliana AT5G65690.1	MAGNGNESTGGDFSFSAARAALPRITTEKGGKSPGPADV CQDDIAPRV
97	>Arabidopsis_halleri Araha.4965s0007.1.p	MSAGDGNATNGDGGFSFPKGPVMPKITTTGAARKNGICHDD SGPTVNATT
98	>Herrania_umbratica XP_021285266.1	MANCGIGNGEFSFGSSGVQNGLAKIQTQRKEADVCHDDSA APVKAQTID
99	>Theobroma_cacao XP_007018343.2	MANRGIGNGEFSFGSSGVQNGLAKIQTQRKEADVCHDDSA APVKAQTID
100	>Herrania_umbratica XP_021273837.1	MAANANGVPTPRSPRLARIQTQKKQNGICHDDSGKPVKAQTI DELHSLQKK
101	>Theobroma_cacao XP_007047680.2	MAANANGVPTPRSPRLARIQTQKKQNGICHDDSGKPVKAQTI DELHSLQKK
102	>Arabidopsis_thaliana AT3G13235.1	MRITVMTAGEQIITLDVDSQETVENVKALLEVESNVPIQQQ QLLYNGNEM
103	>Arabidopsis_thaliana AT3G13235.2	MRITVMTAGEQIITLDVDSQETVENVKALLEVESNVPIQQQ QLLYNGNEM
104	>Arabidopsis_thaliana AT3G13235.3	MRITVMTAGEQIITLDVDSQETVENVKALLEVESNVPIQQQ QLLYNGNEM
105	>Arabidopsis_halleri Araha.10371s0002.1.p	MRITVMTAGEQIITLDVDSQETVENVKALLEVESNVPIQQQ QLLYNGNEM
106	>Herrania_umbratica XP_021288019.1	MKITVMTADEQILSLDVPDHPHETVENLKALLEVETTVPLQQQ QLLYNGREM
107	>Theobroma_cacao XP_007020374.2	MKITVMTADEQILSLDVPDHPHETVENLKALLEVETTVPLQQQ QLLYNGREM
108	>Theobroma_cacao XP_017980634.1	MKITVMTADEQILSLDVPDHPHETVENLKALLEVESVFGVNGQ TTVPLQQQQ
109	>Arabidopsis_thaliana AT5G12170.2	MATTSRRFTTGLFASITSVKSHSANRPQSISLIRRNHIDHR LPLIVPSSR
110	>Arabidopsis_halleri Araha.11630s0011.1.p	MATTSRRFTTGLTASITSVESHANRPQSISLIRRNHIDP RLPPILRSS
111	>Arabidopsis_thaliana AT5G19380.1	MATTSSDRLIAGLTASIGSIESRYANPAQSVSLICRNQING APPVLRSS
112	>Arabidopsis_thaliana AT5G19380.2	MATTSSDRLIAGLTASIGSIESRYANPAQSVSLICRNQING APPVLRSS
113	>Arabidopsis_halleri Araha.38345s0001.1.p	IVIAAATTAALGVGNRVLYKLALIPKQYPFFLAQLSTFGY VAVYFSILY
114	>Herrania_umbratica XP_021279049.1	MTSCYHRLTAGQTLDPDIPVQPPKIKPTASNIPLIPPRPCIN QRPVRIVLR
115	>Theobroma_cacao XP_017983858.1	MTSCYHRLTAGQTLNPVQPPKIKPTASNIPLIPPRPCMN QRPVRIVLR
116	>Arabidopsis_thaliana AT4G25200.1	MASALALKRLSSSIAPRSRSLRPAVSSRLFNTNAVRSYD DDGENGDGV
117	>Arabidopsis_halleri Araha.63261s0001.1.p	NAVRSYDDGGENGHGVDFDRRSVPRRGDFFSDFDPTRSV SQVLNMDQ
118	>Herrania_umbratica XP_021286539.1	MASFALALKRLVSSNIVPSSMCVIRPIATAPYTSRLFNTNSM REFDGDGDE
119	>Theobroma_cacao XP_007038384.2	MASSLALRRLVSSNILPSSMRVIRPIATSPYTSRLFNTNAM REFDDHGDE

120	>Arabidopsis_thaliana AT5G51440.1	MASSSALALRRLSSSTVAVPRALRAVRPVAASSRLFNTNA ARNYEDGVD
121	>Arabidopsis_halleri Araha.1614s0002.1.p	RNHNSNRHVTRRGDFFSDIFDPFTPTRLSQMLNFMQVS EIPLVAAATR
122	>Theobroma_cacao XP_017974095.1	MASSLALRRLVSSNILPSSMRVIRPIATSPYTSRLFNTNAM REFDDHGDE
123	>Arabidopsis_thaliana AT3G12810.1	MASKGGKSKPDIVMASKSGKSKPDNESRAKRQKTEAPKEP RRPKTHWDH
124	>Arabidopsis_halleri Araha.3837s0015.1.p	MASKGGKSKPDIVMASKGGKSKPDNSRAKRHKTEAPKEP RRPKTHWDH
125	>Herrania_umbratica XP_021281896.1	MASKGPRSKLEHETRARRQKALEAPREPQRPKTHWDHVLEE MVWLSKDFE
126	>Theobroma_cacao XP_017976873.1	MASKGPRSKLEHETRARRQKALEAPREPQRPKTHWDHVLEE MVWLSKDFE
127	>Theobroma_cacao XP_017976872.1	MASKGPRSKLEHETRARRQKALEAPREPQRPKTHWDHVLEE MVWLSKDFE
128	>Theobroma_cacao XP_017976871.1	MASKGPRSKLEHETRARRQKALEAPREPQRPKTHWDHVLEE MVWLSKDFE
129	>Theobroma_cacao XP_007029182.2	MASKGPRSKLEHETRARRQKALEAPREPQRPKTHWDHVLEE MVWLSKDFE
130	>Arabidopsis_thaliana AT1G02130.1	MNPEYDYLFKLLLIIGDSGVGKSCLLLRFSDDSYVESYISTI GVDFKIRTV
131	>Arabidopsis_halleri Araha.3702s0007.1.p	MNPEYDYLFKLLLIIGDSGVGKSCLLLRFSDDSYVESYISTI GVDFKIRTV
132	>Herrania_umbratica XP_021279966.1	MNPEYDYLFKLLLIIGDSGVGKSCLLLRFADDSYIESYISTI GVDFKIRTV
133	>Theobroma_cacao XP_017978788.1	MNPEYDYLFKLLLIIGDSGVGKSCLLLRFADDSYIESYISTI GVDFKIRTV
134	>Arabidopsis_halleri Araha.24635s0014.1.p	MNPEYDYLFKLLLIIGDSGVGKSCLLLRFADDSYIESYISTI GVDFKIRTV
135	>Theobroma_cacao XP_007039720.1	MNPEYDYLFKLLLIIGDSGVGKSCLLLRFADDSYIESYISTI GVDFKIRTV
136	>Herrania_umbratica XP_021298952.1	MNPEYDYLFKLLLIIGDSGVGKSCLLLRFADDSYIESYISTI GVDFKIRTV

Fuente: Autora

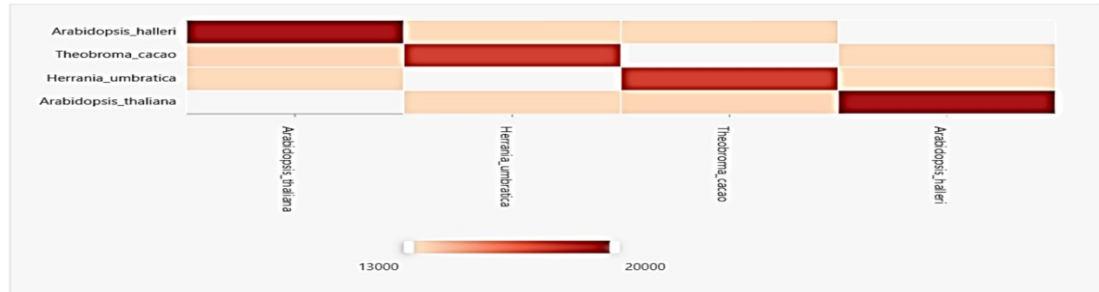
Las Tablas 4 y 5 muestran los resultados de la búsqueda de proteínas por similitud usando el programa Orthovenn 2. Para la Tabla 4 donde se presentan los datos de Aluminio (Al), la suma de las proteínas pertenecientes a las dos especies de *Arabidopsis* suman un 42,85 % y la de los cacaos (*Herrania umbratica* y *Theobroma cacao*) un 57,13 %, siendo estos últimos los de mayor proporción relacionados a resistencia a Aluminio.

Por otro lado, con respecto a la tabla 5, donde se muestran las proteínas relacionadas a la resistencia a Cadmio (Cd) se encontró que para las *Arabidopsis* se obtuvo un 58,81% de las proteínas totales, mientras que, para los cacaos, es decir para *Herrania umbratica* y *Theobroma cacao* se obtuvo un 41,9%.

A continuación, en la Figura 6 se pone de manifiesto los resultados del HeatMap realizado con el total de las secuencias de proteínas. Entre mayor coloración rojiza entre las especies se presente, más alta es la relación con respecto al contenido de proteínas totales.

Figura 6

Compatibilidad de los clústers de las cuatro especies: Theobroma cacao (cacao), Herrania umbratica, Arabidopsis thaliana (thale cress) y Arabidopsis halleri.



Fuente: Autora

4.2. Interpretación de resultados.

Para la interpretación de los resultados de las redes de proteínas formadas de acuerdo a la similitud generada por Orthovenn 2, se tomó como referencia las primeras tres agrupaciones (clústers) como salida del análisis bioinformático de las secuencias de proteínas. Las interpretaciones por clúster son detalladas a continuación:

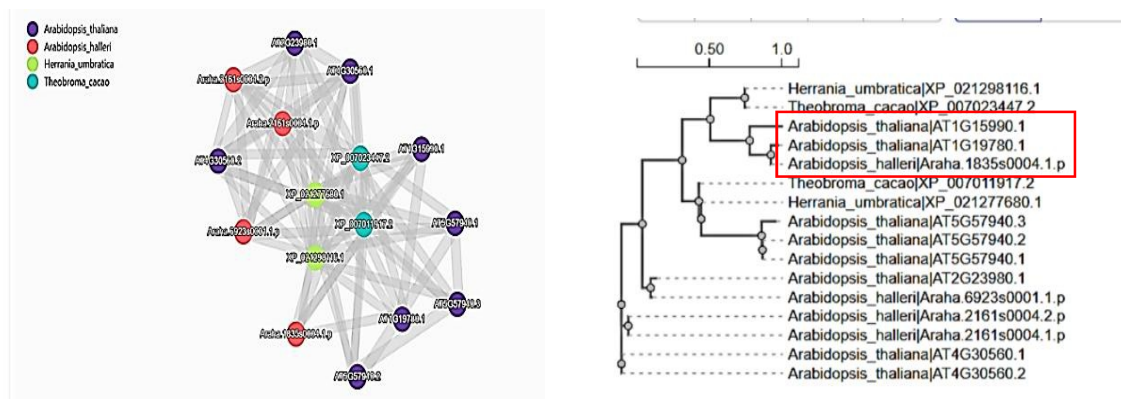
4.2.1. Análisis de las proteínas agrupadas (Clúster) para Cadmio (Cd)

Para el caso del metal pesado Cadmio (Cd) fueron considerados para la interpretación, las salidas (output) de los resultados del clúster 133, clúster 500 y clúster 676.

4.2.2. Interpretación del Clúster133.

Figura 7

Análisis de clúster 133 para Cadmio (Cd).



Fuente: Autora

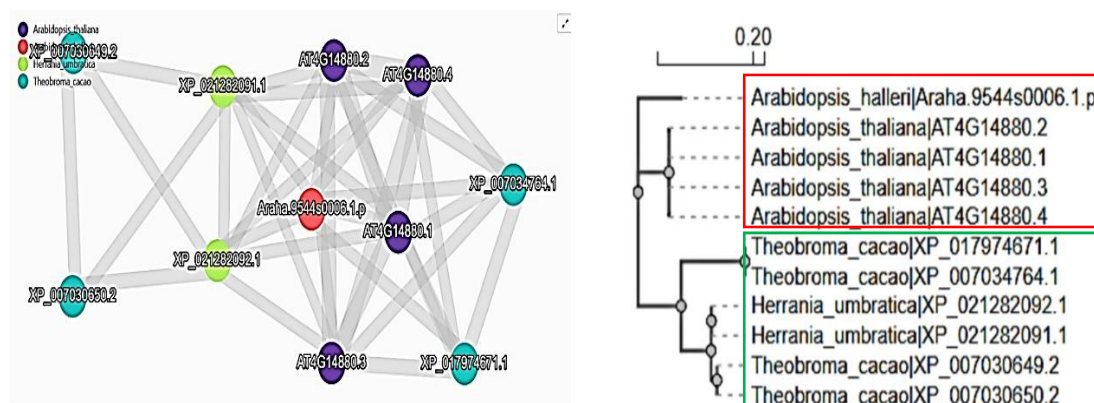
La Figura 7 muestra la combinación de las redes de proteínas realizado por Orthovenn 2 y el árbol filogenético de las proteínas en estudio. Para el primer caso se puede observar que el 50% de las proteínas agrupadas en este clúster pertenecen a *Arabidopsis. thaliana*, siendo el grupo con mayor representación, seguido de *Arabidopsis. halleri* con un 25 %, y el restante de proteínas se distribuyen entre *Herrania. umbrática* y *Theobroma. cacao*.

Para el caso del árbol filogenético (Figura 7) los resultados confirman una agrupación más fuerte entre las proteínas de las dos *Arabidopsis*, sin embargo, las proteínas AT1G15990.1, AT1G19780.1 de *Arabidopsis. thaliana* y Araha.1835s0004.1. p de *Arabidopsis.halleri* se muestran más cercanas a las dos especies de cacao. Estos resultados son originados por las cercanías de la especiación entre las especies en estudio, además por la relación genética existente.

4.2.3. Interpretación del Clúster500.

Figura 8

Análisis de clúster 500 para Cadmio (Cd).



Fuente: Autora

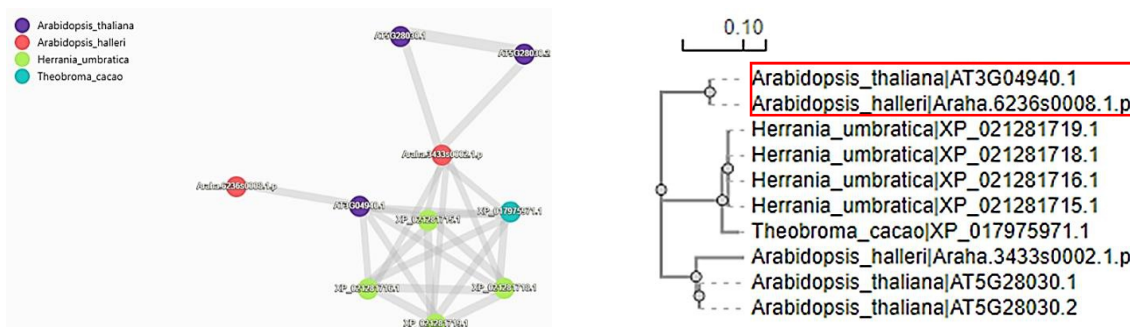
En la Figura 8 se muestran los resultados del clúster 500, los mismos que ponen de manifiesto las redes de proteínas con su respectivo árbol filogenético. Para el caso de las redes se puede apreciar que un 36% de las proteínas pertenecen a *Arabidopsis. thaliana* y *Theobroma. cacao*, siendo las de mayor aglomeración, continuando el de *Herranea. umbrática* con el 18%, y las demás proteínas reducidas a *Arabidopsis. halleri*.

En cuanto al árbol filogenético se pudo observar que se muestran dos grupos bien definidos, el grupo de *Arabidopsis. thaliana* y el grupo de los cacaos representados por *Theobroma. cacao* y *Herrania. umbrática*.

4.2.4. Interpretación del Clúster676.

Figura 9

Análisis de clúster 676 para Cadmio (Cd).



Fuente: Autora

En la Figura 9 se muestra el análisis del clúster 676, donde se exponen las redes de proteínas con su pertinente árbol filogenético. En esta ocasión en las redes se puede observar un 40% de las proteínas que pertenece a *Herranea. umbrática*, las cuales presentan una mayor agrupación, mientras que *Arabidopsis. thaliana* mostró un 30%.

Las demás proteínas están conformadas por *Arabidopsis.halleri* y *Theobroma. cacao*. En cuanto al árbol filogenético se pudo observar que dos proteínas: *Arabidopsis.thaliana* AT3G04940.1 y *Arabidopsis.halleri* Araha.6236s0008.1. p expresan una relación cercana con el grupo de *Herrania. umbrática*.

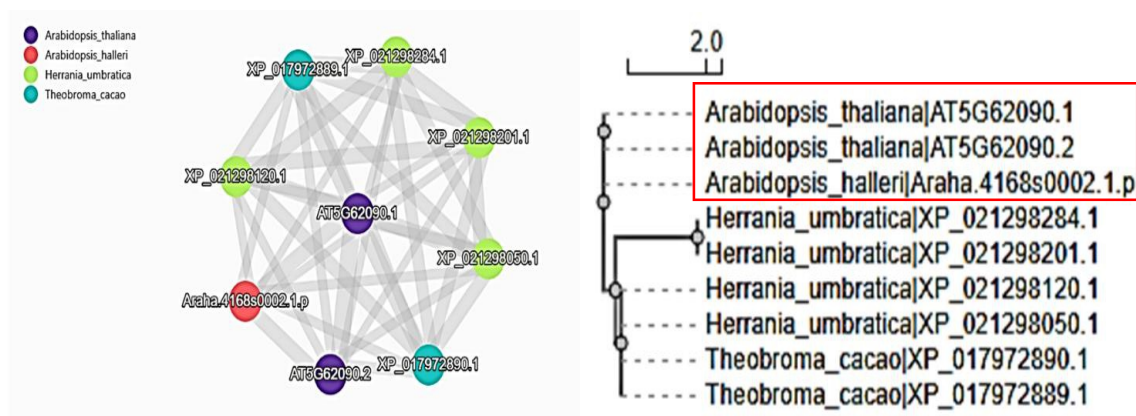
4.2.5. Análisis de las proteínas agrupadas (Clúster) para Aluminio (Al).

En cuanto al metal pesado Aluminio (Al) para la interpretación fueron considerados las salidas (output) del clúster 919 y clúster 958, porque fueron los que presentaron una mayor agrupación de proteínas de acuerdo a los resultados que arrojó Orthovenn2. Estos hallazgos se detallan de la siguiente forma:

4.2.6. Interpretación del Clúster919

Figura 10

Análisis de clúster 919 para Aluminio (Al).



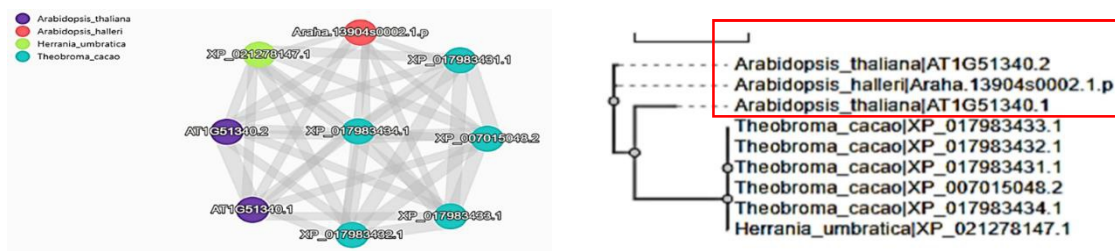
Fuente: Autora

Los resultados del clúster 919 muestran las redes de proteínas con su relativo árbol filogenético. Por lo tanto, en las redes se observa que un 44,44 % de las proteínas le corresponden a *Herrania. umbrática*, debido a que fue la especie que mostró una mayor agrupación de proteínas, respecto a *Arabidopsis. thaliana* y *Theobroma. cacao*, los cuales mostraron un 22% para cada una. En cuanto al análisis filogenético se observa que existe un solo grupo bien establecido para los cacaos (*Theobroma. cacao* y *Herrania. umbrática*) y otro para las *Arabidopsis* conformadas por las proteínas: AT5G62090.1, AT5G62090.2, Araha.4168s0002.1.p.

4.2.7. Interpretación del Clúster958.

Figura 11

Análisis de clúster 958 para Aluminio (Al).



Fuente: Autora

La Figura 11 muestra los resultados del clúster 958, donde se analizó las redes de proteínas con su respectivo árbol filogenético. Por lo consiguiente en las redes se observa que el 55% de las proteínas agrupadas fueron para *Theobroma. cacao*, continuando *Arabidopsis thaliana* con un 22%. En cuanto al análisis filogenético se observa que existe un solo grupo bien establecido para los cacaos (*Theobroma. cacao* y *Herrania. umbrática*) y otro para las *Arabidopsis* conformadas por las proteínas: AT1G51340.2, Araha.13904s0002.1. p, AT1G51340.1

4.3.Discusión.

Como respecto al número de Mb se obtuvo que la especie con mayor longitud en todo su ensamblamiento fue *Theobroma cacao*, este genoma fue reportado en su versión mejorada por Argout, *et al* (40) utilizando un enfoque basado en NGS para mejorar significativamente el ensamblaje del genoma del criollo beliceño B97-61/B2. Con respecto a la conceptualización de una megabase, en genética es una medida utilizada para designar la longitud del ADN. El equivalente de una una megabase es igual a 1 millón de bases (41).

Herrania umbrática, una especie originaria de la Amazonía colombiana, (42) fue la que en relación a número de Mb estuvo más cercana de la especie *Theobroma cacao*, esto debido a que se consideran especies relativas desde el punto de vista genético. Esta especie toma importancia últimamente debido a su potencial de adaptación a diversos factores ambientales, los mismos a los que se encuentra sometidas las plantas de esta especie en su ámbito natural, lo que podría sugerir una importancia en cuanto a riqueza para planes de recursos genéticos y mejoramiento genético (43).

Con respecto a los grupos de proteínas entre las cinco especies bajo estudio, los resultados sugieren que *Herrania umbratica* es una especie candidata no cultivable para estudios posteriores y puede ser un modelo para el cacao.

En este estudio en particular el número de proteínas está muy relacionada con las otras cuatro especies, es decir las dos *Arabidopsis* y Cacao, estos resultados indican que probablemente durante la especiación de *Theobroma Cacao* ocurrieron eventos evolutivos que contribuyeron a la consecución de la especie en particular, donde *Herrania* probablemente tuvo su participación con la aportación de genes y proteínas que siguen siendo conservados hasta la actualidad, cumpliendo un rol importante que podría revelar su influencia en las variedades de Cacao futuras.

A pesar de esto, sería importante también realizar estudios en campo considerando diferentes factores (Interacción Genotipo Ambiente), ya que este tipo de estudios permiten corroborar o pueden llegar a ser responsables de crear variaciones entre especies.

Los efectos de estudios de Interacción GxA pueden producir deforestación, la fragmentación y la selección natural en las poblaciones, por otro lado, tenemos la deriva/delección de genes y otros efectos genéticos particulares (44). La fuerza de la deriva genética durante los procesos de especiación afecta las mutaciones y sus interacciones entre las poblaciones y determina la aptitud parental e híbrida. Tanto la fragmentación como la degradación de los bosques pueden provocar un aumento de la deriva genética, una reducción del flujo de genes y una alteración de los patrones de apareamiento que dan como resultado una mayor endogamia y, finalmente, un impacto negativo en la diversidad genética de las especies de plantas forestales (45).

Estos resultados a través del uso de herramientas bioinformáticas permiten realizar un pre análisis mediante el uso de plataformas y herramientas ómicas en general (46), permitiendo tener hallazgos previos para en lo posterior llegar a estudios más concluyentes en investigaciones en genética, generando futuros planes de mejoramiento genético, en particular a una especie de tanto interés como el Cacao criollo en el Ecuador.

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1.Conclusiones

- El análisis de proteínas mostró un claro efecto de la especiación debido a que cuan mayor fue la cercanía en el linaje genético, mayor fue la relación de proteínas totales entre especie, siendo las *Arabidopsis (thaliana y halleri)* y los *Cacaos (Herrania y Theobroma)* los que se agruparon de forma simultánea, pero también independiente entre sí, conformando dos grupos definidos.
- Algunos clústers mostraron evidencia de que son propios de cada especie, pudiendo ser secuencias génicas claras candidatas a ser analizadas para planes mejoramiento genético, ya que podrían vincularse a genes liderando procesos de resistencia tanto para Aluminio y Cadmio, particularmente en *Theobroma cacao*, la especie objeto de esta tesis.

5.2.Recomendaciones

- Evaluar las relaciones intra e interespecíficas de las proteínas encontradas, lo que permita ver más de cerca la función de estas proteínas y su relación con genes en rutas metabólicas más precisamente identificadas.
- Mediante genómica comparativa y computacional elucidar las funciones de algunos de estos genes en otras especies con el fin de poder considerarlos en un futuro para un estudio de expresión génica en el cultivo de cacao (*Thebroma cacao*)

CAPÍTULO VI

BIBLIOGRAFÍA

6.1. Bibliografía

1. Loor Intriago ML. “Contenido de vitamina C, polifenoles y flavonoides totales en cascarilla de dos variedades de cacao (theobroma cacao L.): nacional y CCN-51”. [Internet] [masterThesis]. Quevedo: Ecuador; 2020 [citado 9 de mayo de 2023]. Disponible en: <https://repositorio.uteq.edu.ec/handle/43000/5935>
2. Efectos del fosforo y azufre sobre el rendimiento de mazorcas, en una plantación de cacao (Theobroma cacao L.) ccn-51, en la zona de Babahoyo [Internet]. [citado 9 de mayo de 2023]. Disponible en: <http://dspace.utb.edu.ec/handle/49000/3358>
3. Chirino L. Agricultura Orgánica en el Trópico y Subtrópico. [citado 23 de mayo de 2023]; Disponible en: https://www.academia.edu/10723422/Agricultura_Org%C3%A1nica_en_el_Tr%C3%B3pico_y_Subtr%C3%B3pico
4. Alloway B. Heavy Metals and Metalloids as Micronutrients for Plants and Animals. En: Heavy metals in soils- trace metals and metalloids in soils and their bioavailability. 2013. p. 195-209.
5. Casteblanco JA. HEAVY METALS REMEDIATION WITH POTENTIAL APPLICATION IN COCOA CULTIVATION. GRANJA Rev Cienc Vida. agosto de 2018;27(1):21-35.
6. Cadmium accumulation and allocation in different cacao cultivars - PubMed [Internet]. [citado 9 de mayo de 2023]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/31078857/>
7. Mite Vivar FA, Carrillo Zenteno MD, Durango Cabanilla WD. Avances del monitoreo de presencia de cadmio en almendras de cacao, suelos y aguas en Ecuador. XII Congreso Ecuatoriano de la Ciencia de Suelo [Internet]. 17 de noviembre de 2010 [citado 9 de mayo de 2023]; Disponible en: <http://repositorio.iniap.gob.ec/handle/41000/5872>
8. Reis GSM, de Almeida AAF, de Almeida NM, de Castro AV, Mangabeira PAO, Pirovani CP. Molecular, Biochemical and Ultrastructural Changes Induced by Pb

- Toxicity in Seedlings of *Theobroma cacao* L. PLoS ONE. 6 de julio de 2015;10(7):e0129696.
9. Ruperty Parraga JE, Sánchez Marceliano JF. Estudio del contenido de metales pesados en cacao (*Theobroma cacao* L.) y su incidencia en la calidad de acuerdo a normativas internacionales considerando su procedencia, tipos y transformación [Internet] [bachelorThesis]. Quevedo: Ecuador; 2021 [citado 9 de mayo de 2023]. Disponible en: <https://repositorio.uteq.edu.ec/handle/43000/6466>
 10. Ordoñez Araque R, Lopez Cortez A, Casa Lopez F, Landines Vera E, Fuentes Pérez EM. Análisis de cadmio, plomo, níquel y arsénico en plantas de cacao y derivados: Industria Alimentaria. *CienciAmérica Rev Divulg Científica Univ Tecnológica Indoamérica*. 2020;9(Extra 4):107-14.
 11. Gustavo E. Cacao orgánico: Guía para productores ecuatorianos. 2010 [citado 9 de mayo de 2023]; Disponible en: <http://repositorio.iniap.gob.ec/handle/41000/4571>
 12. Ferré-Huguet N, Nadal Lomas M, Schumacher M, Domingo JL. Diseño de un software para evaluar los riesgos de exposición ambiental a través del agua, suelos y aire. En: *Derecho de aguas, protección y conservación del medio ambiente*, 2009, ISBN 978-84-9876-647-9, págs 109-124 [Internet]. Tirant lo Blanch; 2009 [citado 9 de mayo de 2023]. p. 109-24. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=3208398>
 13. ¿Qué es la bioinformática y qué aplicaciones tiene en biomedicina? [Internet]. [citado 9 de mayo de 2023]. Disponible en: <https://www.isciii.es/InformacionCiudadanos/DivulgacionCulturaCientifica/DivulgacionISCIIII/Paginas/Divulgacion/Bioinformatica.aspx>
 14. Pérez-Mendoza M. GLUCONEOGÉNESIS: UNA VISIÓN CONTEMPORÁNEA DE UNA VÍA METABÓLICA ANTIGUA. 2012;
 15. Torres Reynoso S. Coe, Sophie D.; Coe, Michael D. La verdadera historia del chocolate. México: Fondo de Cultura Económica, 1999. ISBN 968-16-5569-9. *Rev Digit Univ* [Internet]. 2012 [citado 9 de mayo de 2023];13(6). Disponible en: <https://biblat.unam.mx/es/revista/revista-digital-universitaria/articulo/coe-sophie-d->

coe-michael-d-la-verdadera-historia-del-chocolate-mexico-fondo-de-cultura-economica-1999-isbn-968-16-5569-9

16. Barazarte H, Sangronis E, Unai E. La cáscara de cacao (*Theobroma cacao* L.): Una posible fuente comercial de pectinas. Arch Latinoam Nutr. marzo de 2008;58(1):64-70.
17. Sánchez Ramírez C. El chocolate amargo en la cocina cuencana actual, nuevas recetas [Internet] [bachelorThesis]. Universidad de Cuenca; 2010 [citado 9 de mayo de 2023]. Disponible en: <http://dspace.ucuenca.edu.ec/handle/123456789/1550>
18. Jornaleros, grandes propietarios y exportación cacaotera, 1790-1925 [Internet]. Universidad Andina Simón Bolívar. [citado 9 de mayo de 2023]. Disponible en: <https://www.uasb.edu.ec/publicacion/jornaleros-grandes-propietarios-y-exportacion-cacaotera-1790-1925/>
19. Vera Sigcha CS, Zambrano Mora IA. Evaluación de las características del mucílago de cacao (*Theobroma cacao* L.) en la obtención de alcohol étílico [Internet] [bachelorThesis]. Universidad de las Fuerzas Armadas ESPE. ESPESD. Carrera de Ingeniería Agropecuaria.; 2018 [citado 9 de mayo de 2023]. Disponible en: <http://repositorio.espe.edu.ec/jspui/handle/21000/15798>
20. Organización de las Naciones Unidas para la Alimentación y la Agricultura: FAO en Ecuador | Organización de las Naciones Unidas para la Alimentación y la Agricultura [Internet]. [citado 9 de mayo de 2023]. Disponible en: <https://www.fao.org/ecuador/es/>
21. Vasallao M. Diferenciación y agregado de valor en la cadena ecuatoriana del cacao. Rev Repique. 2017;1(1):23-43.
22. Encuesta de Superficie y Producción Agropecuaria Continua-2020 | [Internet]. [citado 9 de mayo de 2023]. Disponible en: <https://www.ecuadorencifras.gob.ec/encuesta-de-superficie-y-produccion-agropecuaria-continua-2020/>

23. Manual de Mineralogía. Volumen 1 - Editorial Reverté S.A [Internet]. [citado 9 de mayo de 2023]. Disponible en: https://www.reverte.com/libro/manual-de-mineralogia-volumen-1_89161/
24. Ruiz Esparza Rodríguez MA, Martínez Sánchez R. El aluminio, material trascendente en la historia humana. enero de 2017 [citado 9 de mayo de 2023]; Disponible en: <http://repositorio.utm.mx:8080/jspui/handle/123456789/355>
25. Ciencia e ingeniería de materiales 6ta Edición Donald R. Askeland.pdf [Internet]. pdfcoffee.com. [citado 23 de mayo de 2023]. Disponible en: <https://pdfcoffee.com/ciencia-e-ingenieria-de-materiales-6ta-edicion-donald-r-askelandpdf-2-pdf-free.html>
26. CyD [Internet]. [citado 23 de mayo de 2023]. Disponible en: <https://www.cyd.conacyt.gob.mx/?p=articulo&id=370>
27. Sánchez Barrón G, Sánchez Barrón G. Ecotoxicología del cadmio riesgo para la salud de la utilización de suelos ricos en cadmio [Internet]. 2016 [citado 10 de mayo de 2023]. Disponible en: <https://eprints.ucm.es/id/eprint/50902/>
28. Maddela NR, Kakarla D, García LC, Chakraborty S, Venkateswarlu K, Megharaj M. Cocoa-laden cadmium threatens human health and cacao economy: A critical view. *Sci Total Environ.* 10 de junio de 2020;720:137645.
29. Huaraca-Fernandez JN, Pérez-Sosa L, Bustinza-Cabala LS, Pampa-Quispe NB, Huaraca-Fernandez JN, Pérez-Sosa L, et al. Enmiendas orgánicas en la inmovilización de cadmio en suelos agrícolas contaminados: una revisión. *Inf Tecnológica.* agosto de 2020;31(4):139-52.
30. Chavez E, He ZL, Stoffella PJ, Mylavarapu RS, Li YC, Baligar VC. Chemical speciation of cadmium: An approach to evaluate plant-available cadmium in Ecuadorian soils under cacao production. *Chemosphere* [Internet]. 1 de mayo de 2016 [citado 24 de mayo de 2023];150:57-62. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0045653516301643>

31. Yadav SK. Heavy metals toxicity in plants: An overview on the role of glutathione and phytochelatins in heavy metal stress tolerance of plants. *South Afr J Bot.* 1 de abril de 2010;76(2):167-79.
32. Wang S, Wang Y, Zhang R, Wang W, Xu D, Guo J, et al. Historical levels of heavy metals reconstructed from sedimentary record in the Hejiang River, located in a typical mining region of Southern China. *Sci Total Environ.* 1 de noviembre de 2015;532:645-54.
33. Satarug S, Garrett SH, Sens MA, Sens DA. Cadmium, Environmental Exposure, and Health Outcomes. *Environ Health Perspect.* febrero de 2010;118(2):182-90.
34. Fu D, JIANG L yan, Mason A, XIAO M li, ZHU L rong, LI L zhi, et al. Research progress and strategies for multifunctional rapeseed: A case study of China. *J Integr Agric.* 1 de agosto de 2016;15:1673-84.
35. Charrupi Riascos N, Novoa DM. Estudio ambiental del cadmio y su relación con suelos destinados al cultivo de cacao en los departamentos de Arauca y Nariño. *Ing Ambient Sanit* [Internet]. 1 de enero de 2017; Disponible en: https://ciencia.lasalle.edu.co/ing_ambiental_sanitaria/718
36. Jaramillo R, Medina K, Recalde A, Pastás K, Bedoya D, Ramírez V. Evaluación del contenido de cadmio en suelos destinados al cultivo de cacao en la provincia de Guayas (Ecuador). *Ecuad ES Calid* [Internet]. 2015 [citado 10 de mayo de 2023];1(1). Disponible en: <https://revistaecuadrescoalidad.agrocalidad.gob.ec/revistaecuadrescoalidad/index.php/revista/article/view/73>
37. Alcívar L, Demera M, Salas C. Cadmio en el cultivo de cacao (*Theobroma cacao* L.) y sus efectos ambientales. *Téc Rev Las Agrocienc* ISSN 2477-8982. 31 de marzo de 2022;91.
38. Arévalo-Gardini E, Arévalo-Hernández CO, Baligar VC, He ZL. Heavy metal accumulation in leaves and beans of cacao (*Theobroma cacao* L.) in major cacao growing regions in Peru. *Sci Total Environ.* 15 de diciembre de 2017;605-606:792-800.

39. Agropecuarias INA de I, Pichilingue Q (Ecuador) EET. Bienvenido a la Estación Experimental Tropical Pichilingue. octubre de 1998 [citado 10 de mayo de 2023]; Disponible en: <http://repositorio.iniap.gob.ec/handle/41000/1639>
40. Argout X, Martin G, Droc G, Fouet O, Labadie K, Rivals E, et al. The cacao Criollo genome v2.0: an improved version of the genome for genetic and functional genomic studies. BMC Genomics [Internet]. 15 de septiembre de 2017 [citado 1 de junio de 2023];18(1):730. Disponible en: <https://doi.org/10.1186/s12864-017-4120-9>
41. Rodríguez-Santiago B, Armengol L. Tecnologías de secuenciación de nueva generación en diagnóstico genético pre- y postnatal. Diagnóstico Prenat [Internet]. 1 de abril de 2012 [citado 1 de junio de 2023];23(2):56-66. Disponible en: <https://www.elsevier.es/es-revista-diagnostico-prenatal-327-articulo-tecnologias-secuenciacion-nueva-generacion-diagnostico-S2173412712000273>
42. Rondón JB, Campo LJC. Aportes al conocimiento del género *Herrania* (Sterculiaceae) en Venezuela. Acta Bot Venezuelica [Internet]. 2006 [citado 1 de junio de 2023];29(2). Disponible en: http://saber.ucv.ve/ojs/index.php/rev_abv/article/view/795
43. Rondón JB, Campo LJC. Aportes al conocimiento del género *Herrania* (Sterculiaceae) en Venezuela. Acta Bot Venezuelica [Internet]. 2006 [citado 1 de junio de 2023];29(2). Disponible en: http://saber.ucv.ve/ojs/index.php/rev_abv/article/view/795
44. Villamar-Torres R, Bayas BO, Torres-Navarreete Y, Zambrano-Vega C, Jazayeri SM. Comparative computational genomic analysis between wild and domesticated species of cacao and coffee. J Pharm Negat Results [Internet]. 16 de septiembre de 2022 [citado 1 de junio de 2023];13(3):535-51. Disponible en: <https://www.pnrjournal.com/index.php/home/article/view/664>
45. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun [Internet]. 3 de junio de 2014 [citado 1 de junio de 2023];5(1):4006. Disponible en: <https://www.nature.com/articles/ncomms5006%20>.

46. Jazayeri SM, Villamar Torres RO. Genomic and transcriptomic approaches toward plant selection. J Sci Res Rev Cienc E Investig [Internet]. 2017 [citado 1 de junio de 2023];2(8):54-64. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=7344281>

CAPÍTULO VII

ANEXOS

7.1.Evidencia fotográfica



Phytozome 13
THE PLANT GENOMICS RESOURCE

[JGI Home](#)
[JGI Data Portal](#)
[Tools ▾](#)
[Projects ▾](#)
[Genomes ▾](#)
[Cart](#)
[Contact](#)
[Subscribe](#)
[Login](#)

Welcome to **Phytozome ▾**

[Overview](#)
[Release Notes](#)
[News](#)

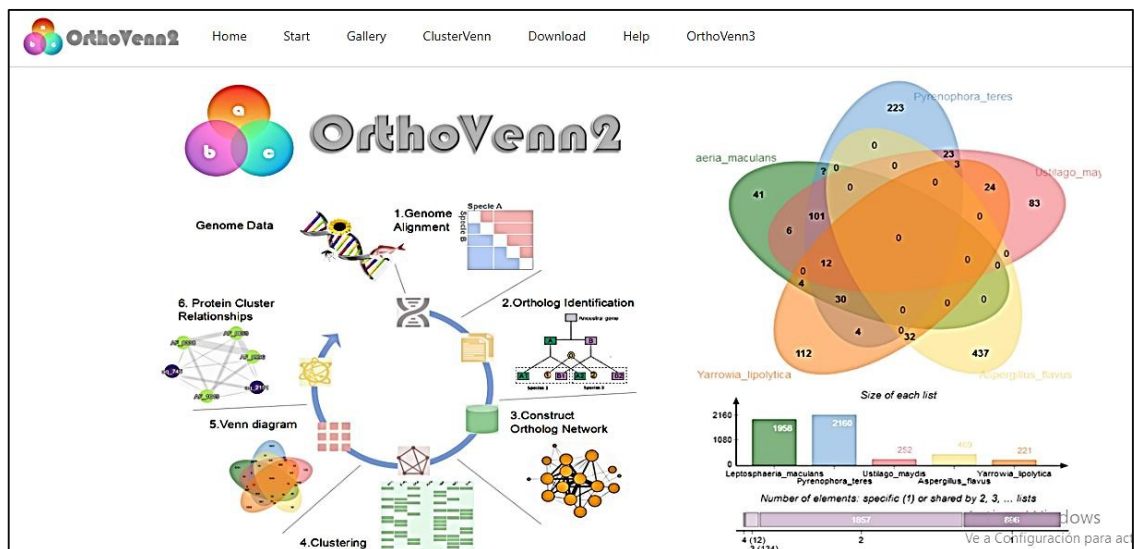
Recent Genome Releases

Genome	Common name	Release Date
Brassica rapa ssp. trilocularis R500 v2.1	Yellow sarson	May 2, 2023
Brachypodium mexicanum v2.1		May 2, 2023
Eschscholzia californica var. Aurantia Orange King Plant1.1 HAP1 v1.1	California poppy	May 2, 2023
Eschscholzia californica var. Aurantia Orange King Plant1.1 HAP2 v1.1	California poppy	May 2, 2023
Corylus americana var. winkler v1.1	American hazelnut Winkler	May 2, 2023
Corylus americana var. rubra v1.1	American hazelnut Dub.	May 2, 2023

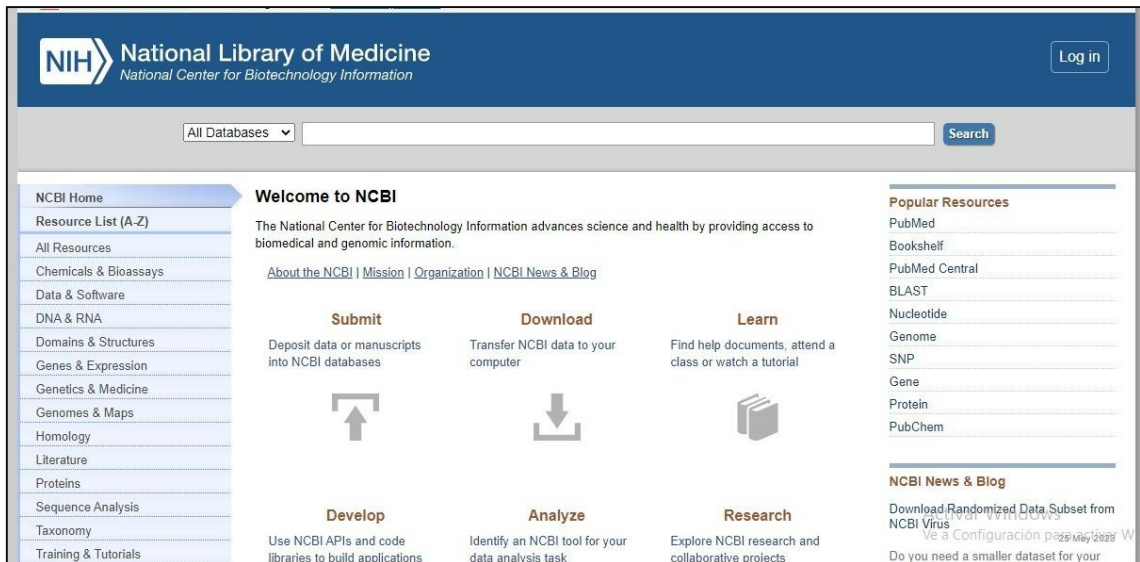
Phytozome, the Plant Comparative Genomics portal of the Department of Energy's Joint Genome Institute, provides JGI users and the broader plant science community a hub for accessing, visualizing and analyzing JGI-sequenced plant genomes, as well as selected genomes and datasets that have been sequenced elsewhere. By integrating this large collection of plant genomes into a single resource and performing comprehensive and uniform annotation and analyses, Phytozome facilitates accurate and insightful comparative genomics studies.

Ve a Configuración para activar Windows.

Fotografía 1
Programa JGI (Phytozone13)

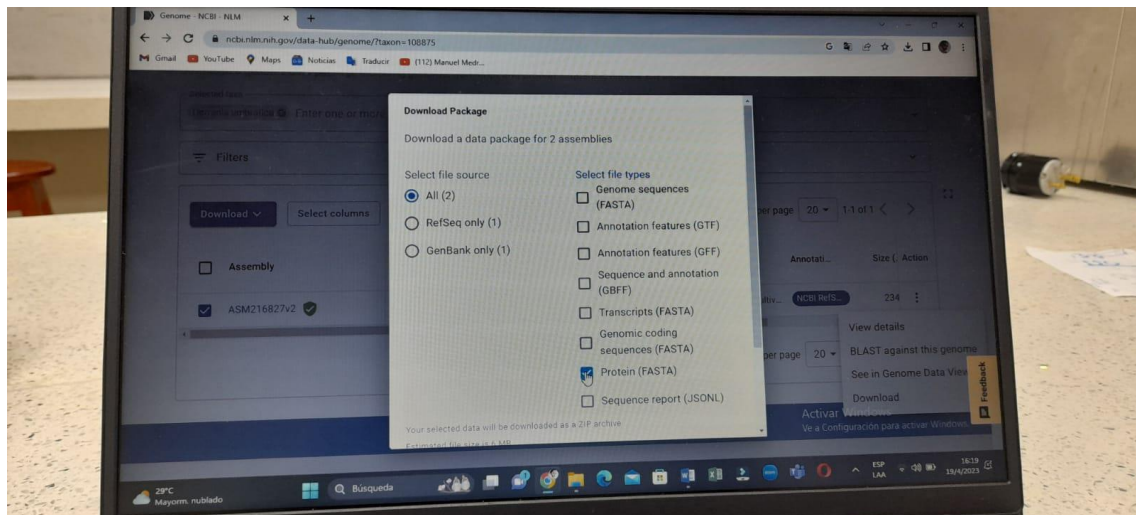


Fotografía 2
Programa *OrthoVenn2*



Fotografía 3

Programa de NCBI (National center for Biotechnology Information)

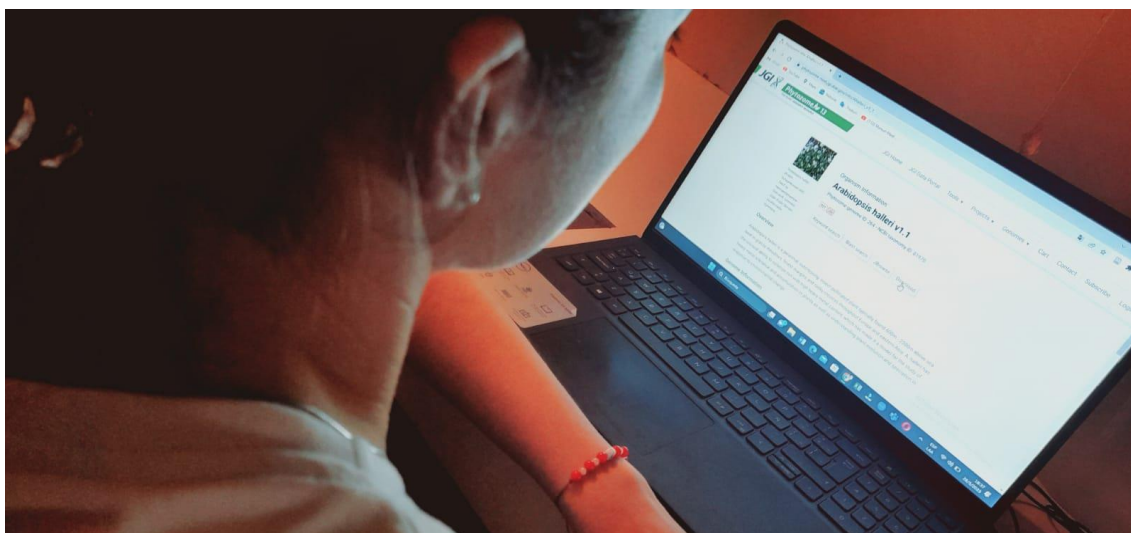


Fotografía 4

Descarga de proteína en NCBI



Fotografía 5
*Análisis de bioinformático de
rutas metabólicas*



Fotografía 6
Descarga de proteína de (Arabidopsis thaliana) en Phytozome