



**INSTITUTO LATINOAMERICANO
DE CIENCIAS DE LA VIDA
Y LA NATURALEZA
(ILACVN)**

**CIENCIAS BIOLÓGICAS
ECOLOGÍA Y BIODIVERSIDAD**

**ESTUDIO DE LA NOTACIÓN FUNCIONAL DE SECUENCIAS
RELACIONADAS A LA DORMANCIA Y GERMINACIÓN EN YERBA MATE
(*Ilex paraguariensis* St. Hil)**

DIEGO S. HERNÁNDEZ LÓPEZ

Foz de Iguaçu
2016



**INSTITUTO LATINOAMERICANO
DE CIENCIAS DE LA VIDA
Y LA NATURALEZA
(ILACVN)**

**CIENCIAS BIOLÓGICAS
ECOLOGÍA Y BIODIVERSIDAD**

**ESTUDIO DE LA NOTACIÓN FUNCIONAL DE SECUENCIAS
RELACIONADAS A LA DORMANCIA Y GERMINACIÓN EN YERBA MATE
(*Ilex paraguariensis* St. Hil)**

DIEGO S. HERNÁNDEZ LÓPEZ

Trabajo de Conclusión de Curso
presentado al Instituto
Latinoamericano de Ciencias de
la Vida y la Naturaleza, como
requisito parcial a la obtención
del título de Licenciatura en
Ciencias Biológicas – Ecología y
Biodiversidad.

Orientador: Prof. Dr. Cristian
Antonio Rojas

Foz do Iguaçu
2016

DIEGO S. HERNÁNDEZ LÓPEZ

**ESTUDIO DE LA NOTACIÓN FUNCIONAL DE SECUENCIAS
RELACIONADAS A LA DORMANCIA Y GERMINACIÓN EN YERBA MATE
(*Ilex paraguariensis* St. Hil)**

Trabajo de Conclusión de Curso
presentado al Instituto Latinoamericano
de Ciencias de la Vida y la
Naturaleza, como requisito parcial a la
obtención del título de Licenciatura en
Ciencias Biológicas – Ecología y
Biodiversidad.

BANCA EXAMINADORA

Orientador: Prof. Dr. Cristian Antonio Rojas
UNILA

Prof. Dr. Michel Rodrigo Zambrano Passarini
UNILA

Prof. M.Cs. Marcelo Cezar Pinto
UNILA

Foz do Iguaçu,
18 de julio de 2016

AGRADECIMIENTOS

A mi familia, por estar siempre ahí.

A mis amigos, en especial a Lucía con quien más he colaborado y que me ha ayudado muchas veces.

A mi profesor orientador Cristian, por la oportunidad, por la paciencia y la buena disposición.

A la gente del laboratorio GIGA de Posadas, en especial a Marcos por poner a disposición las secuencias.

A los miembros del tribunal.

A aquéllos que en momentos bajos me han tendido una mano.

A todas aquellas personas que de alguna u otra manera han colaborado en mi sobrevivencia en estos eternos años de graduación.

*La Yerba Mate
despierta a los dormidos,
corrige a los haraganes
y hace hermanas
a las gentes
que no se conocen.*

Eduardo Galeano.

HERNÁNDEZ LÓPEZ, Diego. **Estudio de la notación funcional de secuencias relacionadas a la dormancia y germinación en yerba mate (*Ilex paraguariensis* St. Hil)** 2016. Número de páginas 64. Trabajo de Conclusión de Curso de Graduación en Ciencias Biológicas – Ecología y Biodiversidad – Universidad Federal de Integración Latinoamericana, Foz de Iguazú, 2016.

RESUMEN

La yerba mate es un árbol de la familia Aquifoliaceae descrita por Auguste de Saint Hilaire en 1822 como *Ilex paraguariensis*, presenta gran importancia social, cultural y económica. Pese a esto, poco se conoce sobre la biología de esta especie, la cual no ha pasado por un proceso considerable de selección y mejoramiento genético presentando características como la dormancia y germinación desuniforme indeseables para el cultivo. En este trabajo se hace una búsqueda de secuencias relacionada a estos procesos en un transcriptoma de Yerba de 200.000 secuencias usando como referencia el genoma de *Arabidopsis thaliana*, y programas de análisis de secuencias BioEdit. De esta forma se encontró que existe conservación de la mayoría de los dominios estudiados. Al mismo tiempo una fracción del transcriptoma de yerba fue analizado mediante el programa Blast2go dejando en evidencia las principales vías metabólicas de la especie destacando la de purina y tiamina. También se señala a *Vitis vinifera* como la especie con la cual las secuencias de *Ilex paraguariensis* mostraron mayor homología. Se logró entonces, un mapeo razonable de las secuencias, teniendo en cuenta la separación filogenética entre las especies comparadas. Asimismo la disponibilidad de un transcriptoma de Yerba ofrece la posibilidad de explorar en profundidad las vías metabólicas por detrás de esta planta.

Palabras llave: Bioinformática, BioEdit, Blast2Go, transcriptoma.

HERNÁNDEZ LÓPEZ, Diego. **Estudio de la notación funcional de secuencias relacionadas a la dormancia y generacional en yerba mate (*Ilex paraguariensis* St. Hil)** 2016. Número de páginas 64. Trabajo de Conclusión de Curso de Graduación en Ciencias Biológicas – Ecología y Biodiversidad – Universidad Federal de Integración Latinoamericana, Foz de Iguazú, 2016.

RESUMO

A erva-mate é uma árvore da família Aquifoliaceae descrito por Auguste de Saint Hilaire, em 1822 como *Ilex paraguariensis*, apresenta grande importância social, cultural e econômica, apesar disso, pouco se sabe sobre a biologia desta espécie, a qual não passou por um processo considerável de seleção e melhoramento apresentando características como dormência e germinação não uniforme indesejável para o cultivo. Neste trabalho se faz uma busca de sequências relacionadas com esses processos em uma transcripto de erva de 200.000 sequências utilizando como referência o genoma de *Arabidopsis thaliana*, e programas de análises de sequências BioEdit. Se achou que existe conservação na maioria dos domínios estudados. Ao mesmo tempo uma fração do transcriptoma da erva foi analisado mediante o programa Blast2go deixando em evidência as principais vias metabólicas da espécie, destacando purina e tiamina, também é mostrado *Vitis vinifera* como a espécie com maior similaridade com as sequências de *Ilex paraguariensis*. Foi obtido um mapeamento razoável das sequências, tendo em conta a separação filogenética entre as espécies comparadas, de igual modo a disponibilidade de um transcriptoma de Erva-Mate oferece a oportunidade para explorar a fundo as vias metabólicas, por trás dessa planta.

Palavras-chave: Bioinformática, BioEdit, Blast2Go, transcriptoma.

LISTA DE TABLAS

tabla 1 – Enzimas mapeadas que participan en algunas vías metabólicas en Yerba Mate	46
--	----

LISTA DE GRÁFICOS

Gráfico 1 – N° de secuencias de <i>Arabidopsis</i> encontradas o no en el transcriptoma de <i>llex</i> para los procesos fisiológicos de dormancia y germinación, considerando un umbral mínimo de blast de e50	31
Gráfico 2 – Mejores hits encontrados.....	45

LISTA DE FIGURAS

Figura 1 -Diseño botanico de <i>Ilex</i>	15
Figura 2 - Mapa de distribución geográfica.....	15
Figura 3 - Secuencia AT1G04635.1	32
Figura 4 - Secuencia AT1G07430.1.....	32
Figura 5 - Secuencia AT2G38560.1.....	33
Figura 6 - Secuencia AT1G15080.1.....	33
Figura 7 - Secuencia AT1G18080.1.....	34
Figura 8 - Secuencia AT1G54060.1.....	34
Figura 9 - Secuencia AT1G62750.1.....	35
Figura 10 - Secuencia AT1G05190.....	36
Figura 11 - Secuencia AT1G17070.1.....	36
Figura 12 - Secuencia AT2G01280.1.....	37
Figura 13 - Secuencia AT4G19230.2.....	37
Figura 14 - Secuencia AT1G09970.2	38
Figura 15 - Secuencia AT1G70210.1.....	38
Figura 16 - Secuencia AT2G13540.1.....	39
Figura 17 - Secuencia AT2G34900.1.....	39
Figura 18 - Secuencia AT5G06550.1.....	40
Figura 19 - Secuencia AT1G02780	41
Figura 20 - Secuencia AT1G13740.1	41
Figura 21 - Secuencia AT2G26000.2.....	42
Figura 22 - Secuencia AT4G39850.3.....	43
Figura 23 - Secuencia AT5G66460.1.....	43
Figura 24 -Vía metabólica de Purina.....	47
Figura 25 - Vía metabólica deTiamina.....	47

SUMARIO

1 INTRODUCCIÓN	12
1.1 REVISIÓN BIBLIOGRÁFICA.....	13
1.1.1 Dificultades del Cultivo.....	16
1.1.2 Dormancia	17
1.1.3 Categorías de Dormancia.....	18
1.1.4 Dormancia en <i>Ilex paraguariensis</i>	20
1.1.5 Bioinformática.....	21
1.1.6 Genes Homólogos entre Especies Vegetales.....	23
2 METODOLOGÍA	27
2.1 Transcriptoma.....	27
2.2 PROSPECCIÓN DE GENES INVOLUCRADOS EN EL PROCESO DE DORMANCIA Y GERMINACIÓN.....	28
2.3 ANOTACIÓN FUNCIONAL PARCIAL DE GENES DEL TRANSCRIPTOMA DE <i>Ilex paraguariensis</i>	30
3 RESULTADOS	31
3.1 PROSPECCIÓN DE GENES INVOLUCRADOS EN EL PROCESO DE DORMANCIA Y GERMINACIÓN	31
3.2 ANOTACIÓN FUNCIONAL PARCIAL DE GENES DEL TRANSCRIPTOMA DE <i>Ilex paraguariensis</i>	45
4 DISCUSIÓN	48
5 CONCLUSIÓN	52
BIBLIOGRAFÍA	53
APÉNDICES	58
APÉNDICE A – SECUENCIAS RELACIONADAS A DORMANCIA QUE PRESENTAN LOS MISMOS DOMINIOS EN <i>A.thaliana</i> E <i>I.paraguariensis</i>	59
APÉNDICE B – SECUENCIAS RELACIONADAS A GERMINACIÓN QUE PRESENTAN LOS MISMOS DOMINIOS EN <i>A.thaliana</i> E <i>I.paraguariensis</i>	60
APÉNDICE C – SECUENCIAS RELACIONADAS A DORMANCA CON DOMINIOS QUE ESTAN PRESENTES EN <i>A. thaliana</i> Y NO SE ENCUENTRAN EN <i>I. paraguariensis</i>	62

APÉNDICE D – SECUENCIAS RELACIONADAS A GERMINACIÓN CON DOMINIOS QUE ESTAN PRESENTES EN <i>A. thaliana</i> Y NO SE ENCUENTRAN EN <i>I. paraguariensis</i>	63
APÉNDICE E – SECUENCIAS RELACIONADAS A DORMANCIA QUE PRESENTAN DOMINIOS EN <i>I. paraguariensis</i> QUE NO ESTAN PRESENTES EN <i>Arabidopsis</i>	64
APÉNDICE F - SECUENCIAS RELACIONADAS A GERMINACIÓN QUE PRESENTAN DOMINIOS EN <i>I. paraguariensis</i> QUE NO ESTAN PRESENTES EN <i>A. thaliana</i>	65
APÉNDICE G –SECUENCIAS RELACIONADAS A DORMANCIA QUE NO PRESENTAN DOMINIOS CONSERVADOS EVIDENTES NI EN <i>Arabidopsis thaliana</i> NI EN <i>Ilex paraguariensis</i>	65
APÉNDICE H - SECUENCIAS RELACIONADAS A GERMINACIÓN QUE NO PRESENTAN DOMINIOS CONSERVADOS EVIDENTES NI EN <i>A. thaliana</i> NI EN <i>Ilex</i>	65
APÉNDICE I – SECUENCIAS RELACIONADAS A DORMANCIA EXTRAÍDAS DE TAIR.....	66
APÉNDICE J - SECUENCIAS RELACIONADAS A GERMINACIÓN EXTRAIDAS DE TAIR	67
APÉNDICE K – SECUENCIAS RELACIONADAS A DORMANCIA DE YERBA MATE.....	68
APÉNDICE L- SECUENCIAS RELACIONADAS A GERMINACION DE YERBA MATE.....	69

1 INTRODUCCIÓN

La Yerba Mate tiene varios usos, entre ellos se destaca la preparación de infusiones como el mate, tereré, chimarrão, entre otras bebidas y es usada también en la elaboración de algunos fármacos así como en la industria cosmética (CARDOSO, 2006). Estudios revelaron la presencia de polifenoles como el ácido clorogénico con gran poder antioxidante, alcaloides como cafeína responsable del efecto estimulante del sistema nervioso, saponinas, flavonoides, antocianinas y vitaminas del complejo B, A, C y E (BASTOS *et al.*, 2007, BRACESCO *et al.*, 2011).

Muchas especies de plantas que no han pasado por un proceso de selección y mejoramiento genético presentan características salvajes, siendo que estas condiciones pueden llegar a ser un problema para el cultivo. La Yerba Mate se encuadra en este grupo de plantas. En la actualidad existen pocos programas de mejoramiento de Yerba, entre ellos se destacan el de Embrapa en Brasil e INTA en Argentina. Algunas de las características de la Yerba Mate inherentes a su estado primario de domesticación que obstaculizan su cultivo son la presencia de semillas con baja viabilidad, un bajo índice de germinación y desuniformidad de la misma.

Pese a ser una planta con una gran connotación simbólica, sobre todo para la región sur de América del Sur, que presenta gran importancia social, cultural y económica (DUTRA *et al.*, 2008) son insuficientes las tentativas de la comunidad científica por profundizar en el conocimiento sobre la biología de esta especie. En este contexto se hace interesante impulsar estudios enfocados principalmente en aquellos aspectos limitantes en la producción de *I. paraguariensis*.

1.1 REVISIÓN BIBLIOGRÁFICA.

La Yerba Mate es un árbol de la familia Aquifoliaceae que puede alcanzar entre 8 a 15 metros de altura y fue descrita por Auguste de Saint Hilaire en 1822 como *Ilex paraguariensis* (BRACESCO *et al.*, 2010). Presenta hojas de color verde oscuro, simples, ovoides, con bordes dentados y filotaxia alterna. Son árboles dioicos, con flores pequeñas, unisexuadas, de color blanco y la floración tiene lugar entre los meses de octubre a diciembre. La fructificación se da entre los meses de enero y marzo y los frutos son del tipo bayas, de color violeta en la madurez con 4 o 5 semillas cuya dispersión se da sobre todo por aves (DELLACASSA *et al.*, 2001). Es una especie típica de bosque ombrófilo mixto, aunque también puede encontrarse en bosque estacional semidecidual y en bosque ombrófilo denso. Su área de distribución comprende la región sur de Brasil, noreste de Argentina, región oriental de Paraguay y algunas regiones del norte de Uruguay, ubicándose entre la latitud 22° a 30° S y longitud 48° a 56° W (OLIVEIRA, 1985; GIBERTI, 2011).

La yerba mate es uno de los componentes principales del sistema agrícola tradicional y uno de los más antiguos de la región sur en Brasil, presentando gran importancia en la economía y cultura de los estados del sur. Por un cierto periodo de tiempo fue el principal producto de exportación brasileño y aún en la actualidad su producción representa una de las principales fuentes de renta y empleo para pequeños y medianos productores en la región de distribución geográfica natural de la especie (DE QUADROS, 2009).

La explotación de este cultivo se da en su forma nativa asociada a bosques de araucarias o en plantaciones comerciales. Los productos derivados de su cultivo son utilizados en la elaboración de bebidas como mate, chimarrão, tereré, tés, entre otros, así como aplicaciones industriales como colorantes,

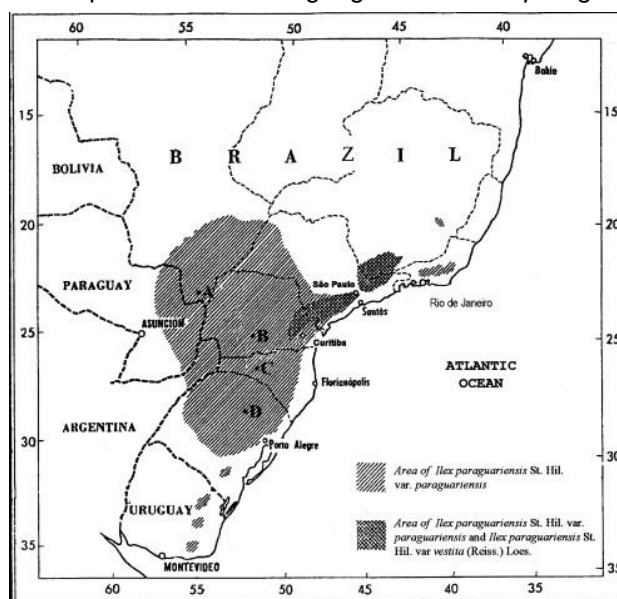
conservantes y en la industria farmacéutica y cosméticos (MAZUKOWSKI, 2000). También se le atribuyen propiedades medicinales como estimulante del sistema nervioso, acción diurética, actuando en casos de cólicos renales (BRAGAGNOLO *et al.*, 1980). Dados los diversos usos de derivados de esta especie, los sistemas de cultivos practicados hasta el momento como la colecta extractivista de los plantíos naturales, no han sido suficientes para abastecer la demanda actual. En este contexto surgió la necesidad de realizar plantíos extensivos de *I. Paraguariensis* aumentando la producción.

Figura 1: Diseño botánico de *Ilex paraguariensis*



Fuente: www.cordobatimes.com

Figura 2: Mapa de distribución geográfica de *Ilex paraguariensis*.



Fuente: Gauer *et al.* 2000

1.1.1 Dificultades del Cultivo:

Pese a su importancia socioeconómica y cultural en la región, el cultivo de yerba ha estado relegado frente a otras culturas como son el maíz, soja, arroz, entre otros. Esta situación ha favorecido al poco interés en planes de mejoramiento genético.

Para la implementación de un sistema de producción aumentada de este cultivo es necesario el desarrollo de estrategias de superación de las limitaciones biológicas. La producción de plantines de yerba mate por semillas enfrenta algunas limitaciones en virtud de la baja calidad genética y fisiológica de sus semillas, (RODIGHERI, 1997) tales como la germinación demorada y desuniforme que puede ir de 100 a 360 días, y un reducido porcentaje de germinación, inferior a 20% (MENNA, 1995; STURION, 1988). Además, en consecuencia al fenómeno de dormancia por el que pasan las semillas, resulta un embrión morfológicamente inmaduro que requiere pasar por un periodo de estratificación para que complete su desarrollo (FOWLER; STURION, 2000). La baja calidad fisiológica de las semillas de yerba es un problema, evidenciado por los frutos maduros que presentan aproximadamente 0,9% de las semillas con embriones maduros. Otra cuestión ligada a las dificultades en la germinación de *I. paraguariensis* está dada por la dureza e impermeabilidad de sus semillas, dificultando el proceso de absorción de agua e intercambio gaseoso desencadenantes del proceso germinativo. Estas cuestiones dificultan la secuencia de los programas de mejoramiento genético en esta especie. Sumado a esto, las semillas recién colectadas presentan un tenor de agua inadecuado para el almacenamiento haciendo necesario un proceso de secado; en este contexto el comportamiento fisiológico de las semillas es descrito de forma contradictoria en la literatura. Borgheti *et al.*, (1990) concluyeron que las semillas

de *I. paraguariensis* presentan comportamiento recalcitrante y por ello no es posible almacenarlas o desecarlas por periodos largos de tiempo sin perder la viabilidad. Por el contrario Ellis *et al.*, (1985) agruparon a la familia Aquifoliaceae como de comportamiento ortodoxo. La importancia de esta cuestión radica en el hecho de que en el caso de ser de comportamiento ortodoxo es posible conservar viable el germoplasma de esta especie por un largo periodo de tiempo (MEDEIROS *et al.*, 2001).

1.1.2 Dormancia:

La dormancia de la semilla es un estadio del ciclo de vida de algunas especies de plantas, presente principalmente en plantas silvestres y especies invasoras. Se caracteriza por la ausencia de la capacidad de la semilla, de germinar en un determinado periodo de tiempo, permitiendo que las especies vegetales sobrevivan a las adversidades. El proceso de dormancia es variable entre las especies vegetales; puede ser de apenas días, meses o de varios años. Inclusive para una misma especie, el período puede variar en función del genotipo y el ambiente donde la semilla se encuentre (RICE *et al.*, 1999). Así, incluso semillas de una misma planta, experimentan intensidades distintas de dormancia. Esta propiedad es una característica adaptativa de las especies en los diferentes ecosistemas, cumpliendo la función de distribuir la germinación a lo largo del tiempo. Es en este sentido, que este fenomeno hace que la germinación no se produzca al mismo tiempo y por el contrario, lo haga por tandas; de manera de que algunas de esas semillas se desarrollen en condiciones optimas para su crecimiento, garantizando la supervivencia de la especie. Sin embargo desde el punto de vista agronómico y de tecnología de semillas, la dormancia es vista como un problema, ya que puede llevar a la pérdida de sembradíos, comprometiendo la implantación de la cultura. Además,

es uno de los factores que contribuyen a la persistencia de las plantas dañinas, dificultando su control y su erradicación con daños económicos para los agricultores (BENECH-ARNOLD *et al.*, 2000). En este contexto cabe resaltar que la mayoría de las plantas cultivadas actualmente es representada por variedades, cultivares e híbridos genéticamente mejorados por procesos de selección artificial, que eliminaron la dormancia, respondiendo a los objetivos de la agricultura moderna, tales como la rapidez y la uniformidad de la germinación de la semilla para la emergencia de la plántula en campo (MOLITERNO, 2008).

1.1.3 Categorías de Dormancia:

Han sido varios los mecanismos descritos por los cuales se da el fenómeno de dormancia. Aspectos como el lugar de origen del carácter, momento en que se establece y atributos morfológicos o fisiológicos de la semilla, determinan el nivel de respuesta. En este sentido, el sitio donde se origina este fenómeno, da lugar a dos tipos de dormancia: embrionaria e impuesta por los tegumentos (CARDOSO, 2009). La dormancia tegumentar o exógena, es provocada por los envoltorios que rodean el embrión, principalmente el endosperma y pericarpio, que impermeabilizan la semilla conformando una barrera al intercambio de gases y agua, al mismo tiempo que ejercen una resistencia mecánica al crecimiento del mismo. La dormancia embrionaria o endógena, se da principalmente en especies forestales y es causada por la existencia de embriones inmaduros o por mecanismos fisiológicos que involucran inhibidores del crecimiento embrionario. Este último tipo de dormancia está dada por una interacción entre promotores e inhibidores de la germinación (KHAN, 1996), donde las hormonas vegetales participan en los mecanismos de dormancia y retomada de la germinación de semillas. Existen evidencias de que el ácido abscísico (ABA) actúa como regulador en la inducción y mantenimiento de la dormancia, mientras

hormonas como las giberelinas tienen un efecto contrario promoviendo la germinación (OLIVEIRA, 2013).

El momento en el que se origina la dormancia es usado como un criterio de clasificación en el que semillas que adquieren el carácter cuando aun están en la planta madre, presentan dormancia primaria. Por otro lado, aquellas semillas que adquieren el carácter posterior a la dispersión o las que presentan un bajo nivel de dormancia primaria, son clasificadas como de dormancia secundaria (BENECH-ARNOLD *et al.*, 2000). Asimismo, semillas que presenten dormancia, pueden adquirir un estado más profundo del carácter debido a las condiciones ambientales desfavorable (THOMPSON *et al.*, 2003). En la dormancia primaria, las semillas maduras son permeables al agua, pero aun así no son capaces de germinar. Evidencias indican que el ácido abscísico (ABA) está relacionado a la manifestación de dicho carácter (FINCH-SAVAGE *et al.*, 2006).

Nikolaeva (1977) propuso uno de los principales sistemas de clasificación de dormancia, diferenciando entre dormancia endógena y exógena. Posteriormente Baskin y Baskin (2004) crearon una clasificación más precisa basada en los conceptos anteriores, donde dormancia fisiológica aparece como la más abundante en la naturaleza y fue subdividida en tres categorías: profunda, intermedia y no profunda (BASKIN *et al.*, 2004, CARDOSO, 2009). En estas categorías se ajusta el concepto de dormancia morfológica, característica de semillas que se separan de la planta madre cuando el embrión aun no está desarrollado. La dormancia morfofisiológica es también evidente en semillas con embriones inmaduros, pero que además presentan componentes fisiológicos que impiden la retomada de actividad metabólica. Finalmente se define la dormancia física, causada por la impermeabilidad que imponen los tejidos que forman la semilla; esta categoría es equivalente a la de dormancia tegumentar antes mencionada.

Vale señalar que en algunos casos la dormancia establecida en una determinada especie, puede estar dada por una combinación de algunas de las categorías antes mencionadas, conocida como dormancia combinada (FINKELSTEIN *et al.*, 2008).

1.1.4 Dormancia en *Ilex paraguariensis*:

Las semillas de muchas especies que no han pasado por un proceso de domesticación, no germinan, aun bajo condiciones ambientales favorables, experimentando un estado de dormancia. Este es el caso de la Yerba Mate, cuyas semillas pasan por un prolongado periodo de dormancia que puede alcanzar de 10 a 12 meses. El proceso de dormancia en *Ilex* se establece en las semillas cuando el fruto aún es inmaduro y se encuentra en la planta, existen factores presentes en el endosperma y endocarpo de las semillas, que inhiben el desarrollo del embrión en la fase conocida como corazón (MEDEIROS, 1998). Estudios de germinación realizados en esta especie, revelaron que embriones provenientes de semillas de frutos blancos (inmaduros) presentaron en media, un índice de germinación de 89,8%, sugiriendo que la acumulación de inhibidores dentro del fruto, comienza cuando éste es aún inmaduro. Esto no ocurre cuando los frutos son de coloración violetas (maduros). El pH también es determinante para el crecimiento embrionario; la literatura señala como valores ideales a pHs entre 5,5 a 6,5 (FERREIRA *et al.*, 1991). Las semillas de *Ilex paraguariensis* se caracterizan además por presentar un envoltorio endurecido que impide la re-hidratación de la semilla e intercambio gaseoso necesarios para iniciar la actividad germinativa. Esta forma de dormancia impuesta por los tegumentos, es superada en forma natural cuando los frutos son ingeridos por aves. El pasaje por el sistema digestivo de éstas, ayuda a debilitar los tejidos que recubren el embrión; de la misma manera se describe la asociación con

hongos existentes en las áreas naturales de donde es nativa la especie, asociados a las semillas, que contribuyen a la germinación, en la medida en que descomponen los tegumentos que la envuelven (CUQUEL *et al.*, 1994, GRIGOLETTI *et al.*, 1999). Este hecho permite deducir que esta especie presenta dormancia tegumentar, y en estado natural es superada por la acción de estos hongos saprofitas (FOWLER *et al.*, 2007).

1.1.5 Bioinformática:

En sentido amplio, la bioinformática es una ciencia dedicada al manejo y análisis de información biológica (DOPAZO *et al.*, 2004). En un sentido más específico, el término hace referencia a las técnicas de manipulación computacional matemáticas, unidas a los análisis de secuencias biológicas (ATTWOOD *et al.*, 2002). Esta es una de las ciencias más relevantes en el área de la biología y se espera que continúe su desarrollo en los años siguientes. Queda claro que se vuelve necesaria la creación de este tipo de herramientas. para el procesamiento sistemático de grandes cantidades de información biológicas, obtenidas a través de la genómica, transcriptómica, proteómica, entre otras (PEÑA *et al.*, 2008).

En la actualidad existen numerosas bases de datos que acumulan grandes cantidades de información sobre secuencias de ADN y proteínas. Desde la década de los 90, más de 190 organismos fueron secuenciados (BUDIMAN *et al.*, 2005) y para el año 2005 más de 400.000 secuencias de ADN habían sido producidas. Esta inmensa cantidad de información es el resultado de nuevas técnicas de secuenciación, sumado al avance tecnológico informático, ya que el procesamiento computacional es fundamental en el estudio de grandes cantidades de información (ROCHA, 2005). Aun cuando los pormenores técnicos relativos a la bioinformática son complejos, ciertamente los resultados

pueden arrojar información de gran utilidad para la biología (ROCHA, 2007). La bioinformática lleva en cuenta cuestiones como la recopilación, organización y almacenamiento de la información biológica en bases de datos; desarrolla herramientas de búsqueda y análisis de información en dichas bases de datos; participa en la generación y uso de programas informáticos para la realización de tareas, que en última instancia permitirán interpretar los resultados y dar un sentido biológico (ATTWOOD *et al.*, 2002). Las bases de datos forman parte esencial de la bioinformática. Es un recurso computacional que permite el almacenamiento y revisión de secuencias de ADN, ARN, proteínas y otras informaciones asociadas. Estas pueden variar en su contenido y calidad (ATTWOOD *et al.*, 2002). La mayoría de las bases de datos poseen también información taxonómica, la cual puede ser obtenida a través de internet. Estas herramientas permiten el libre acceso a la información y vuelven más dinámica la interacción entre centros de investigación e investigadores (ARNDT, 2007; VILLEGAS, 2014).

Algunas de las bases de datos de uso público más conocidas incluyen Genbank, perteneciente al National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov>), The European Molecular Biology Laboratory (EMBL, <http://www.embl.de>), DNA Databank of Japan (DDBJ, <http://www.ddbj.nig.ac.jp/search-es-e.html>) y SynaBASE (<http://www.synamatix.com>).

Éstas son usadas para la caracterización de genomas y transcriptomas de diversos organismos (ANWAR, 2007). Un concepto asociado al manejo de bases de datos es el de mineración de datos, este se refiere a la búsqueda y hallazgo de información valiosa en un gran conjunto de datos (OLAFSSON *et al.*, 2006). La metodología usada en general es automatizada, pero es posible también la realización del proceso manual y así formar categorías y organizar

datos en nuevos conjuntos de información (MUTIS, 2008). Una de las grandes dificultades en la bioinformática es la comparación lineal entre secuencias, conocido como método de alineamiento. Mediante este procedimiento se introducen inserciones en las secuencias para lograr que posiciones equivalentes en secuencias adyacentes puedan ser comparadas (ATTWOOD *et al.*, 2002). Del procedimiento de alineación resultan la predicción de familias de proteínas, estructura y función, la creación de secuencias cortas de ADN (oligonucleótidos) para amplificar regiones de interés. Además, el uso de procesos computacionales para la identificación de genes y comparación entre diferentes bases de datos, logra dar con características únicas de los genes y su expresión (LUSCOMBE *et al.*, 2001).

1.1.6 Genes Homólogos entre Especies Vegetales:

El estudio de las vías metabólicas en especies sin genoma públicamente disponible (o recién secuenciados), se beneficia de la información surgida del secuenciamiento genómico de especies denominadas “modelo”. Existen numerosos métodos para encontrar secuencias similares entre la enorme cantidad de secuencias de las bases de datos. Estos métodos se basan en modelos estadísticos para determinar cuándo estos parecidos se deben a que ambas proteínas comparten un mismo origen y cuándo se deben a parecidos al azar. Se pueden distinguir dos tipos de métodos: los que realizan comparaciones entre pares de secuencias y otros más recientes (de nueva generación), que incluyen información de la familia de proteínas para, por un lado, encontrar homólogos lejanos (aquéllos que se parecen poco) y por otro, para discriminar mejor los parecidos que son fruto del azar de aquéllos que reflejan una homología (GARZON, 2012). Cuando comparamos dos secuencias de forma aislada y observamos unos pocos residuos idénticos (conservados) es más difícil saber si es algo significativo, si no sabemos si esos mismos residuos están conservados

en la familia de proteínas; dicho de otro modo: si los residuos más importantes para la función de la proteínas son los que aparecen idénticos, la confianza de que las proteínas sean homólogas es mayor. Esta información de qué residuos son más importantes sale a la luz con los alineamientos múltiples de secuencias. Estos métodos de nueva generación son capaces de detectar tres veces más homólogos remotos que los tradicionales (PARK *et al.*, 1998). Entre los primeros métodos se encuentran BLAST (ALTSCHUL *et al.*, 1990; 1997) y FASTA (ALTSCHUL *et al.*, 1997). Este método realiza de forma muy rápida una búsqueda de secuencias parecidas en las bases de datos. Para cuantificar los parecidos se determina una puntuación del alineamiento entre las dos secuencias, de las frecuencias esperadas se calculan log-odds de los que se derivan las puntuaciones. BLAST, aplica un marco estadístico (basado en un modelo aleatorio que describe cómo se distribuyen las puntuaciones de parecidos al azar y qué parámetros afectan a esta distribución) para determinar cuán significativa es una determinada puntuación, dadas las características de la secuencia problema de la base de datos y de la matriz de sustitución. Además proporciona un valor esperado que indica para cada puntuación, cuántas veces esperaríamos que por azar apareciese esa determinada puntuación o una mejor en la base de datos utilizada (ALTSCHUL, 1996).

El método de BLAST resulta muy útil para conocer de forma rápida cuáles son los homólogos cercanos de una proteína. A medida que se secuencian nuevas proteínas, se deposita esta información en bases de datos de acceso público. Paralelamente, se lleva a cabo una anotación de la función de estas proteínas, bien a partir de datos experimentales o bien a partir de los parecidos observados con otras secuencias de función conocida. Una alternativa para la exploración de los mecanismos que están por detrás de las vías metabólicas que dan lugar a la dormancia es el abordaje genómico, donde son

secuenciadas extensas regiones del genoma en cuestión. Puede realizarse a nivel del ADN o a partir de la fracción codificante (utilizando para esto el ARN). Cualquiera que sea el abordaje utilizado, frecuentemente, se hace necesario el uso concomitante de herramientas que consigan ordenar la gran cantidad de información generada en este tipo de proyectos, así como también dividir la fracción de genes que realmente importan para cada caso específico. La Bioinformática ofrece la solución a estas cuestiones. En las última década, los avances de la ingeniería genética y las nuevas tecnologías, condicionaron el surgimiento de esta disciplina que creó vínculos indisolubles entre la Informática y las ciencias biológicas (SOLORZANO *et al.*, 2003). La producción de equipamientos comerciales de alto rendimiento y economía, son capaces de producir grandes cantidades de datos en poco tiempo, como es el caso de la secuenciación masiva en paralelo (*Illumina*), que puede secuenciar en una sola corrida de análisis, más de un billón de bases como afirma HALL en 2007 (BARRETOS, 2008).

Haciendo uso de esta tecnología, recientemente, un grupo de investigación en Yerba Mate elaboró un transcriptoma de 44,907 transcritos (DEBAT *et al.*, 2014), sin embargo las secuencias no están disponibles. Otro grupo de investigadores paralelamente iniciaron un proyecto similar, obteniendo 200.000 secuencias de esta especie a partir de secuenciación de nueva generación (NGS) con *Illumina HiSeqTM-2000* (datos no publicados).

Haciendo uso de esta información disponible y mediante softwares bioinformáticos, este trabajo tiene como objetivo identificar las rutas metabólicas y los genes involucrados en el control y regulación del proceso de germinación y específicamente del fenómeno de dormancia en *Ilex paraguariensis*. Para ello las secuencias obtenidos por nuestros colaboradores de la ciudad de Posadas, Misiones, Argentina, serán analizadas utilizando por

un lado el programa BioEdit v7.2.5 con el cual se hará una prospección de genes relacionados al proceso de dormancia y germinación; por otra parte usando el algoritmo Blast2go que realiza la notación funcional de las 200.000 secuencias y compara los resultados obtenidos con millones de secuencias almacenadas en una base de datos.

Este trabajo visa la contribución de nuevos conocimientos en la biología de Yerba Mate, que por ser representativa de un grupo de plantas que presentan poca importancia comercial, son poco estudiadas, sin embargo no son menos importantes, ya que a este grupo pertenecen la mayor parte de plantas no domesticadas que pueblan los ecosistemas naturales y muchas de ellas pueden presentar mecanismos de dormancia similares a los encontrados en Yerba Mate. De esta manera, conociendo mejor los mecanismos que están por detrás de la dormancia en esta planta, será posible también entender un poco más sobre los mecanismos usados por otras plantas como las recalcitrantes, aportando nuevos conocimientos, por ejemplo en la implementación de bancos de semillas y germoplasma.

2 METODOLOGÍA

2.1 TRANSCRIPTOMA:

En este trabajo se comparan secuencias de un transcripto de *Ilex paraguariensis* usando como referencia una base de datos de genes relacionados a dormancia y germinación de *Arabidopsis thaliana*, conocido organismo modelo y objeto de numerosos trabajos de investigación por contar con su genoma secuenciado disponible. Para la ejecución de dicho trabajo se usa una computadora Hewlett Packard modelo 15-AF131dx AMD® A6 4 core, memoria RAM 4GB, 500GB 5400RPM hard drive, con software Windows 10 Home (64-bit).

Las secuencias utilizadas para este trabajo fueron obtenidas a partir de un secuenciamiento de nuestros colaboradores de la ciudad de Posadas, Misiones, Argentina. El mismo fue realizado a partir de diferentes tejidos y órganos de plantas adultas de *Ilex paraguariensis*. El transcriptoma fue realizado con RNA-Seq total de Illumina (secuenciación de nueva generación o NGS); fue realizado el montaje de genes, utilizando la metodología conocida como *de novo*, la cual es una herramienta apropiada para analizar el transcriptoma, y no necesita la clonación de bibliotecas de ADNc y principalmente no requiere conocimiento *a priori* de la especie, es decir que puede aplicarse sin tener un genoma de referencia. Este montaje derivó en la obtención de 200.000 secuencias finales, que son el objeto de estudio de este trabajo. La descripción y análisis definitivo de dichas secuencias se encuentra en fase de preparación y será enviado para publicación en breve.

2.2 PROSPECCIÓN DE GENES INVOLUCRADOS EN EL PROCESO DE DORMANCIA Y GERMINACIÓN:

Inicialmente se realizó una búsqueda de secuencias de *Arabidopsis thaliana* en la fuente bioinformática TAIR (<https://www.arabidopsis.org/>.html) las cuales son usadas como secuencias de referencia para la comparación. El motor de búsqueda para este caso son las palabras “dormancy” y “germination” en el buscador de la base de datos. De esta manera se obtienen 128 y 170 secuencias respectivamente asociadas a los genes de interés. De todas las secuencias encontradas, son seleccionadas para el estudio de comparación 31 secuencias relacionadas a dormancia y 83 a germinación. La elección de las secuencias usadas se basa en la descripción de las funciones de dichas secuencias. Estas son guardadas en un block de notas (apéndice I y J) así como sus ontologías génicas relacionadas a los procesos biológicos en los que participan.

Seguidamente la base de datos mencionada anteriormente fue cargada al programa de bioinformática BioEdit v7.2.5 disponible en (<http://www.mbio.ncsu.edu/BioEdit/page2.html>) donde es realizado el procesamiento de las mencionadas secuencias. El primer paso fue crear una base de datos con formato “BLAST”, que sirva para realizar alineamientos de secuencia de forma local, o sea, sin recurrir a servidores externos, como el NCBI u otros. Una vez creada esta base de datos las secuencias de *Arabidopsis* previamente rescatadas del consorcio TAIR son analizadas, una a una, mediante el programa BioEdit que produce el alineamiento de las secuencias de ambos organismos y relaciona cada secuencia de *A. thaliana* con una secuencia de *I. paraguariensis*. Con esta finalidad fue utilizada

la modalidad BlastP, que compara la secuencia de aminoácidos de los genes de *Arabidopsis* con cada una de las secuencias de nuestro transcriptoma de *I. paraguariensis*, previa traducción en los seis posibles marcos de lectura.

Esta relación es soportada por un valor esperado o *e-value* que indica en este caso, la probabilidad de que las secuencias comparadas se parezcan en la base de datos por acaso, en lugar de ser por homología (ABASCAL, 2003).

Como medida de corte, la literatura sugiere continuar únicamente con aquellas secuencias que obtuvieron e-valor menor que e-050 (SOLIS-CALERO, 2008)

Las secuencias de aminoácidos de los genes de interés encontrados en nuestro genoma fueron copiadas y llevadas a otra base de datos pública, el N de sus siglas en inglés, *National Center for Biotechnology information* disponible en (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Para el caso de las secuencias de *Arabidopsis thaliana* obtenidas de TAIR se usa BLASTP, sección de la base de datos de NCBI que busca dominios de proteínas en su base de datos utilizando un “query” o anzuelo de proteínas. Como resultado de esta búsqueda se obtiene una representación gráfica de dominios proteicos putativos conservados para cada secuencia.

Para las secuencias de Yerba Mate (apendice K y L), se realizó también en el NCBI la búsqueda de dominios proteicos conservados mediante BLASTX (de nucleótido a proteína). Una vez obtenidos los dominios proteicos conservados de Yerba Mate y de *Arabidopsis*, se realiza la comparación entre ellos.

2.3 ANOTACIÓN FUNCIONAL PARCIAL DE GENES DEL TRANSCRIPTOMA DE *Ilex paraguariensis*.

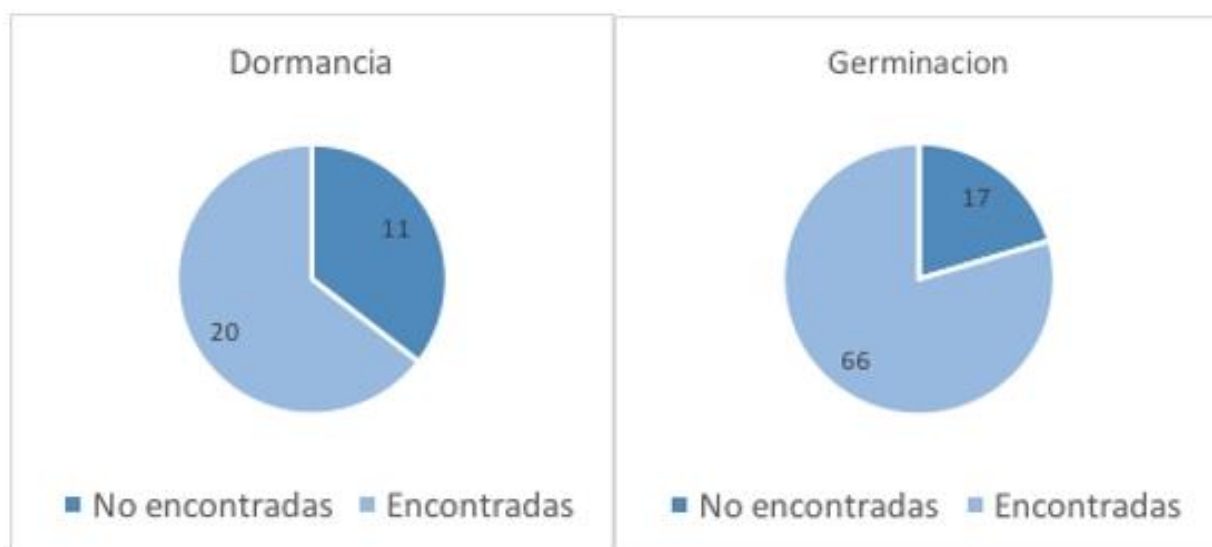
Debido al tiempo demandante para trabajar con el transcriptoma entero, de las 200.000 secuencias del transcriptoma original, fueron separadas aleatoriamente 10.000 y con ellas fue realizada la anotación funcional mediante el programa Blast2go (CONESA *et al.*, 2008), disponible en <http://www.blast2go.com/>. Se trata de una herramienta web con interfase Java, que funciona en cualquier sistemas operativo (Windows, Linux, otros), para análisis funcional de secuencias. Este software usa el *Gene Ontology* (THE GENE ONTOLOGY CONSORTIUM, 2000), como forma de facilitar el intercambio de informaciones disponibles entre la colección de secuencias que serán analizadas y un banco de datos y padroniza la representación de los genes y sus productos para todos los sistemas biológicos (NODA *et al.*, 2010). Estableciendo tres categorías: 1- procesos biológicos en los que el gen o su producto participa, 2- función molecular, 3- lugar de la célula donde el gen o su producto es activado. Blast2go permite realizar búsquedas online con BLAST, búsqueda por secuencias similares, entre otras funciones (BARRETOS, 2008)

3 RESULTADOS

3.1 PROSPECCIÓN DE GENES INVOLUCRADOS EN EL PROCESO DE DORMANCIA Y GERMINACIÓN.

Con la intención de conocer el grado de similitud entre las vías metabólicas involucradas en los procesos de germinación y dormancia entre *Arabidopsis* e *illex paraguariensis* fue realizada una búsqueda en el portal del genoma de *A. thaliana* con los términos “dormancy” y “germination”. Fueron obtenidas 128 y 170 secuencias respectivamente de las cuales se seleccionaron 31 para dormancia y 83 para germinación. De las secuencias elegidas para el análisis de comparación, solo fueron separadas las que atendieron al pre requisito impuesto dado por un e- valor menor a 1×10^{-5} mencionado anteriormente. De esta forma, 20 secuencias para el caso de dormancia y 66 para el caso de germinación fueron consideradas para el análisis como muestra el grafico 1.

Grafico 1: N° de secuencias de *Arabidopsis thaliana* encontradas o no en el transcriptoma de *Illex paraguarienses* para los procesos fisiológicos de dormancia y germinación, considerando un umbral mínimo de blast de e-50



Con el objetivo de facilitar el procesamiento de los resultados, fueron creadas cuatro categorías: la primera hace referencia a proteínas con dominios iguales en ambos organismos; la segunda y tercer incluyen a proteínas que están presentes en un organismo pero no en el otro y viceversa; y la última agrupa proteínas con dominios no conservados en ninguno de los dos organismos.

1- Secuencias que presentan los mismos dominios en *Arabidopsis thaliana* e *Ilex paraguariensis*.

Figura 3: Secuencia AT1G04635.1 relacionada a dormancia involucrada en el desarrollo del embrión que termina en dormancia de la semilla, lipoylation de proteína, procesamiento de tRN en *Arabidopsis* e *Ilex* respectivamente.

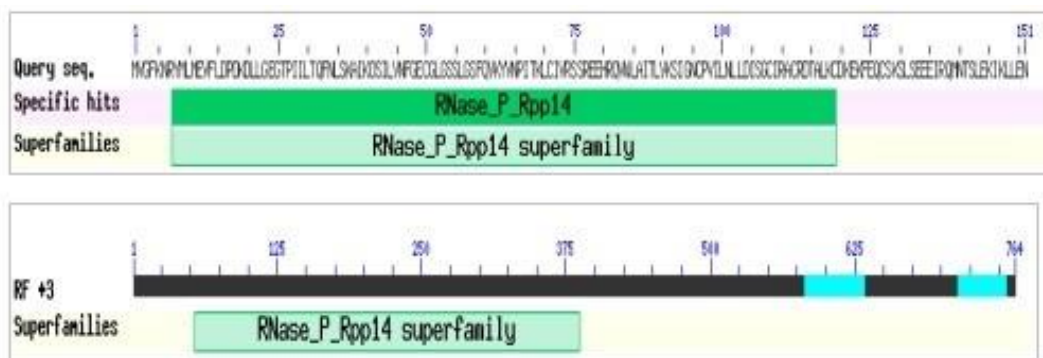


Figura 4: Secuencia AT1G07430.1 relacionada a dormancia involucrada en regulación negativa de la vía de señalización activadas por ácido abscísico, dormancia de semillas, regulación positiva de la vía de señalización mediada por el ácido giberélico y germinación, liberación de las semillas de la dormancia en *Arabidopsis* e *Ilex* respectivamente.

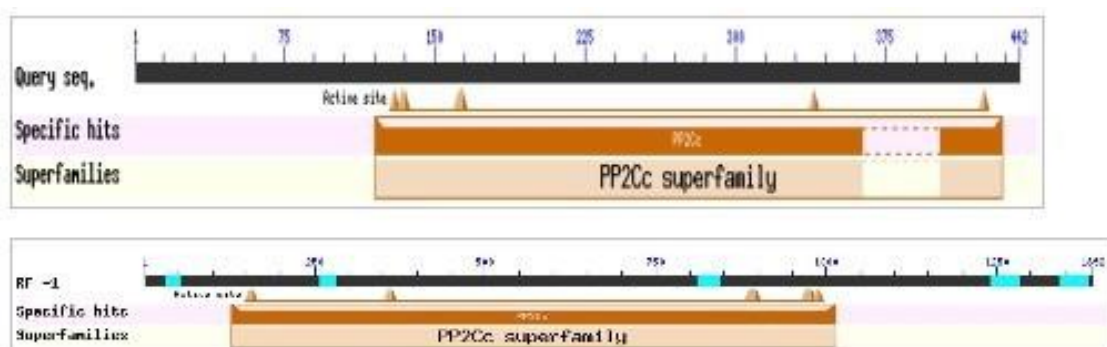


Figura 5: Secuencia AT2G38560.1 relacionada a dormancia involucrada en regulación negativa del desarrollo de la flor, regulación de la transcripción de elongación de promotor ARN polimerasa II, respuesta a la giberelina, proceso de dormancia de las semillas, germinación de semillas en *Arabidopsis* e *Ilex* respectivamente.

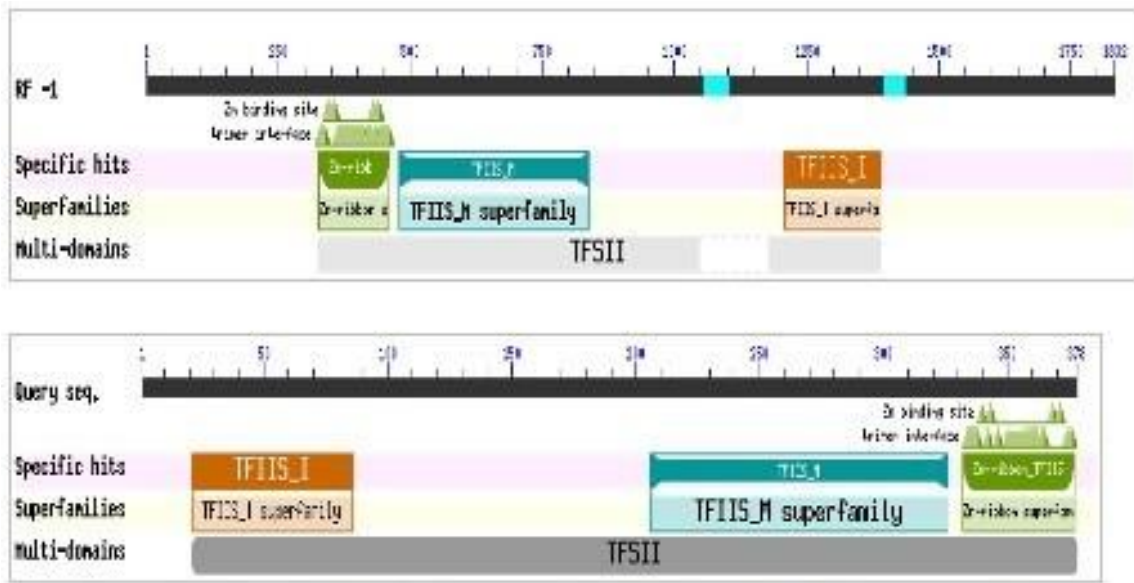


Figura 6: Secuencia AT1G15080.1 relacionada a germinación que participa en la vía de activación de ácido abscísico, la desfosforilación de fosfolípidos, metabolismo de fosfolípidos en *Arabidopsis* e *Ilex* respectivamente.

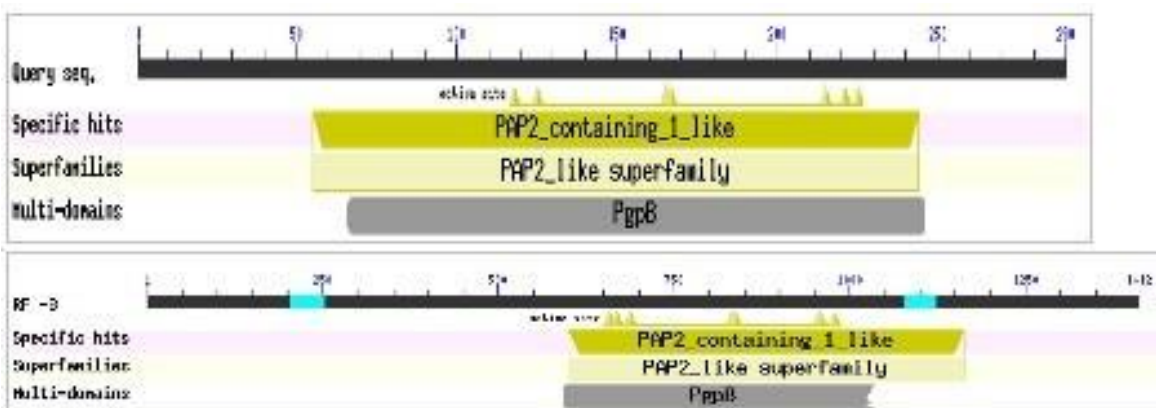


Figura 7: Secuencia AT1G18080.1 relacionada a germinación, involucrada en respuesta a estímulos del ácido abscísico, giberelina mediada por señalización, regulación de la traducción, respuesta ión cadmio, respuesta a las giberelinas, biogénesis de ribosomas, germinación de la semilla, en Arabidopsis e Ilex respectivamente.

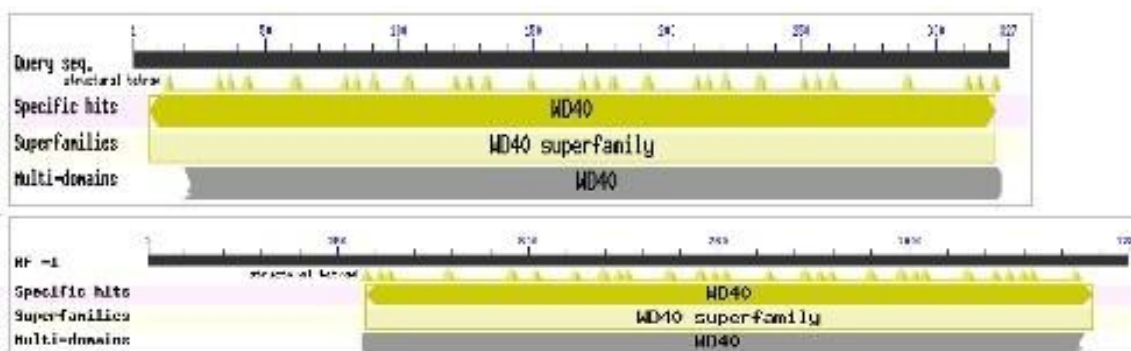


Figura 8: Secuencia AT1G54060.1 relacionada a germinación involucrada en el desarrollo del embrión y dormancia, regulación de la germinación y la transcripción, respuesta a auxina, maduración de la semilla en Arabidopsis e Ilex respectivamente.

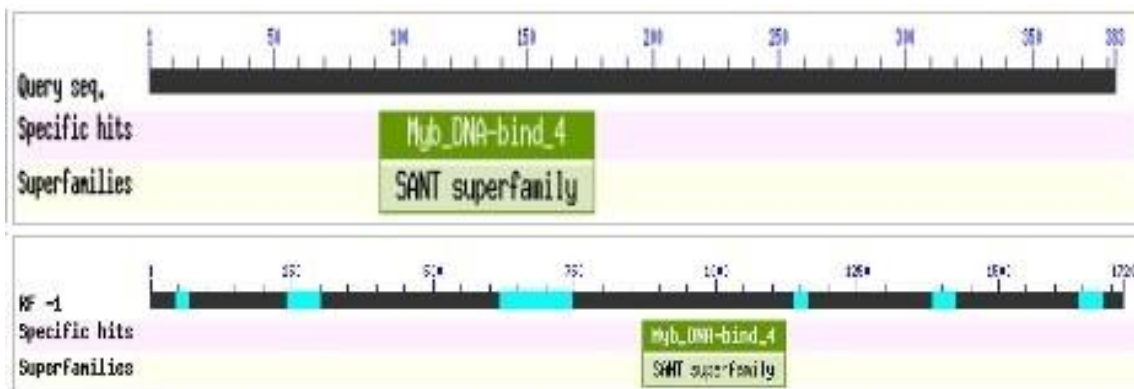
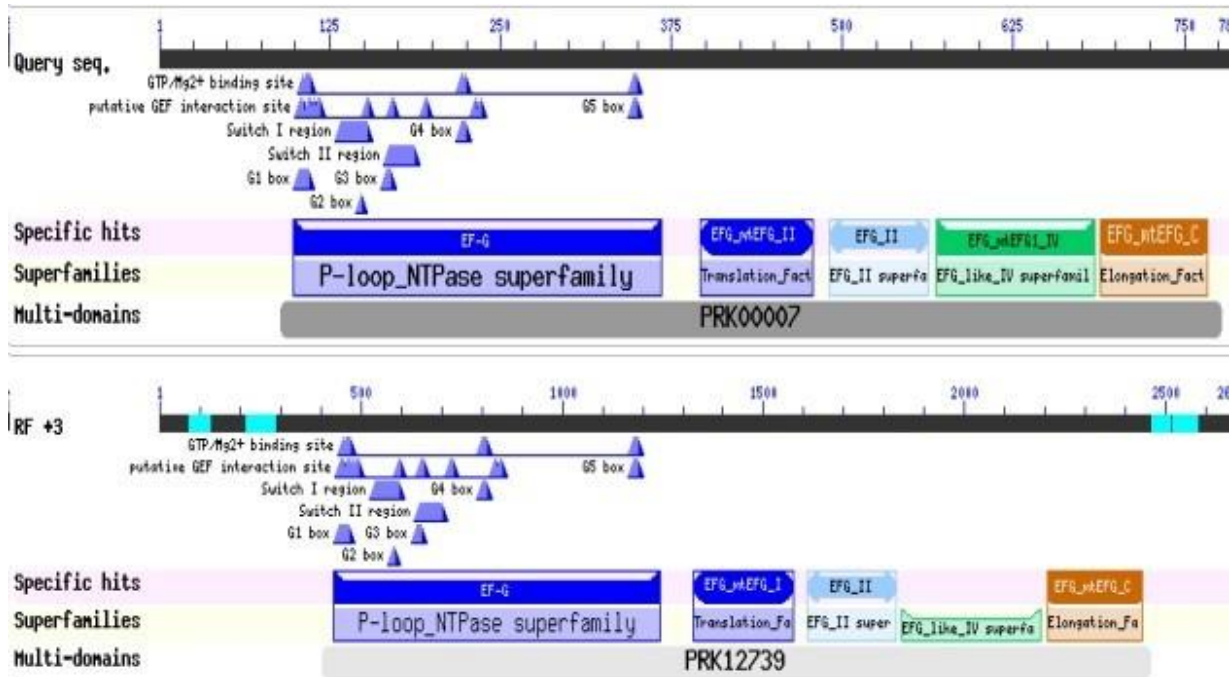


Figura 9: Secuencia AT1G62750.1 relacionada a germinación involucrada en Organización del cloroplasto, desarrollo post-embrionario, desmontaje de ribosoma, germinación de semillas en Arabidopsis e Ilex respectivamente.



Estas secuencias presentan el mismo tipo de dominios proteicos, en la misma secuencia con aproximadamente los mismos tamaños (ver apéndice A y B). En el caso de las secuencias 3, 14, 15, 17 y 19 de dormancia y las secuencias 4, 5, 7, 13, 18, 20, 24, 26, 29, 31, 35, 39, 40, 43, 47, 49, 51, 60, 63, 72, 76, 77 y 78 de germinación se ve la conservación de aminoácidos pertenecientes al sitio catalítico, importantes para la constitución del dominio. Estas regiones sugieren la caracterización de los mismos tipos de enzimas.

2- Secuencias con dominios que están presentes en *Arabidopsis* y no se encuentran presentes en *Illex* :

Figura 10: Secuencia AT1G05190 relacionada a dormancia involucrada el el desarrollo del embrión que termina en la dormancia de las semillas, respuesta a las citoquinas en *Arabidopsis* e *Illex* respectivamente

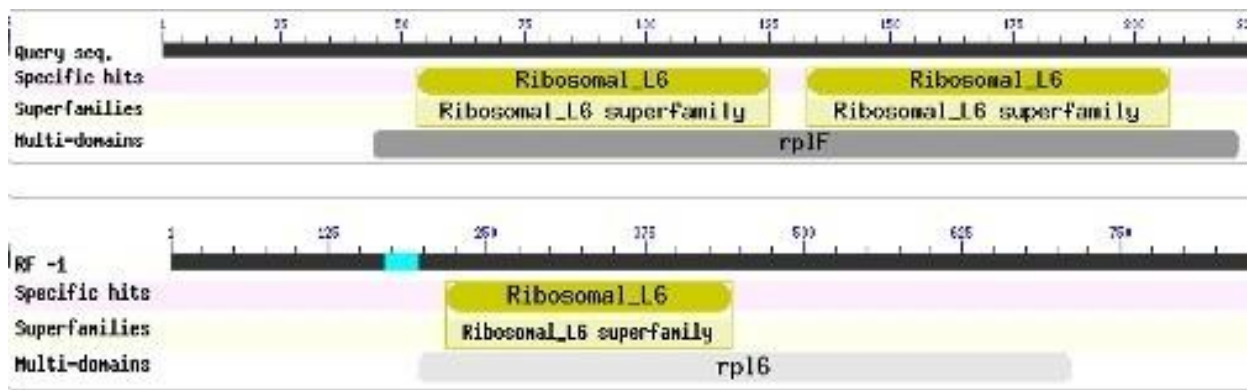


Figura 11: Secuencia AT1G17070.1 relacionada a dormancia involucrada en mRNA de empalme, regulación del ritmo circadiano, desmontaje complejo de spliceosomas en *Arabidopsis* e *Illex* respectivamente

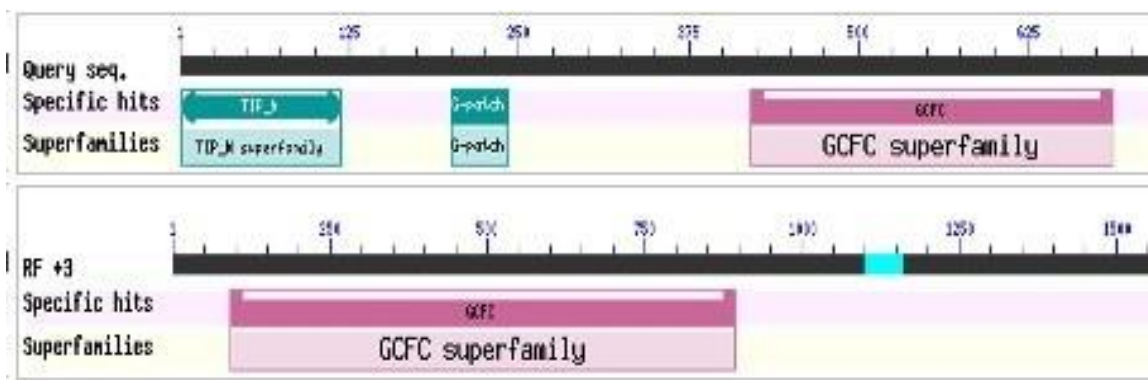


Figura 12: Secuencia AT2G01280.1 relacionada a dormancia involucrada en el fin del desarrollo del embrión y dormancia de las semillas, regulación de la transcripción en *Arabidopsis* e *Ilex* respectivamente

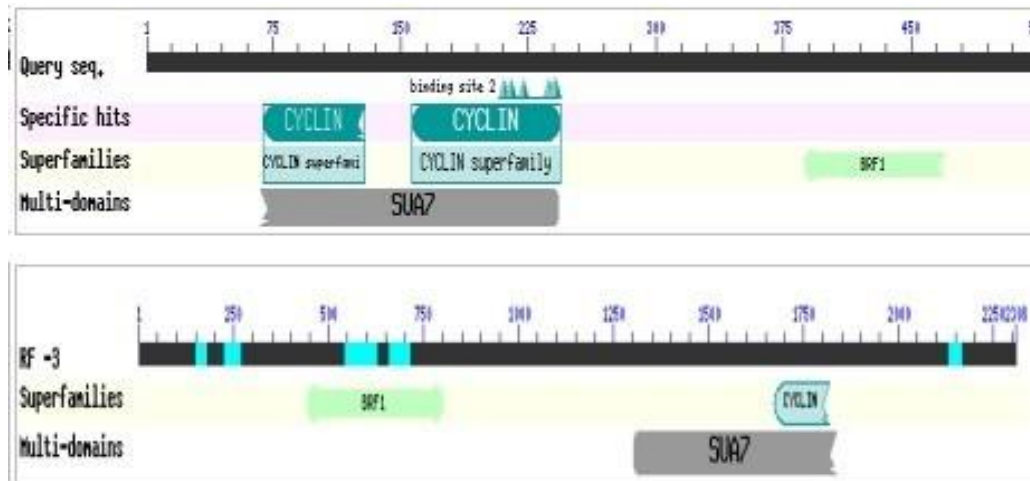


Figura 13: Secuencia AT4G19230.2 Metabolismo del ácido abscísico, biosíntesis de brasinoesteroides, homeostasis de brasinoesteroides, respuesta de defensa a los hongos, proceso de oxidación-reducción, liberación de las semillas de la dormancia, respuesta a la luz roja lejana, metabolismo de esteroides en *Arabidopsis* e *Ilex* respectivamente

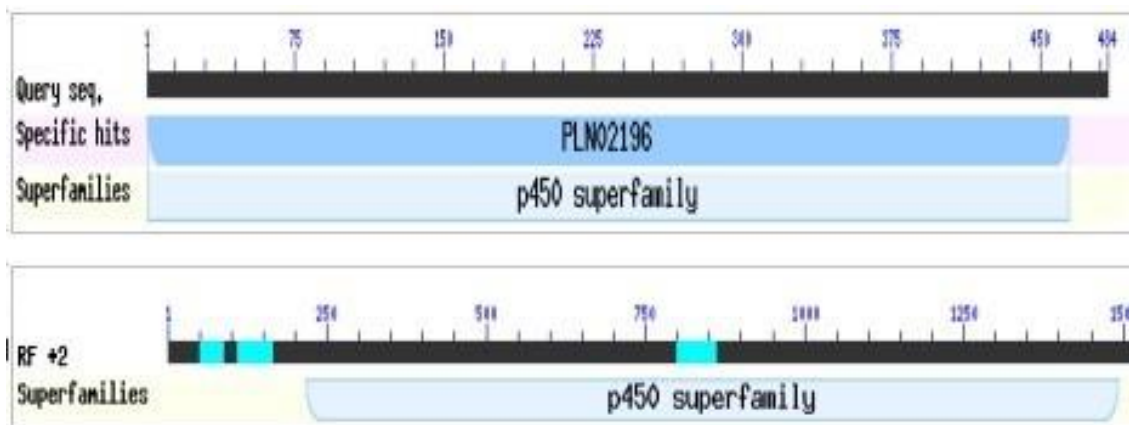


Figura 14: Secuencia AT1G09970.2 relacionada a germinación involucrada en fosforilación de proteínas, respuesta al estrés oxidativo, vía de señalización de la tirosina quinasa del receptor de transmembrana en *Arabidopsis* e *Ilex* respectivamente

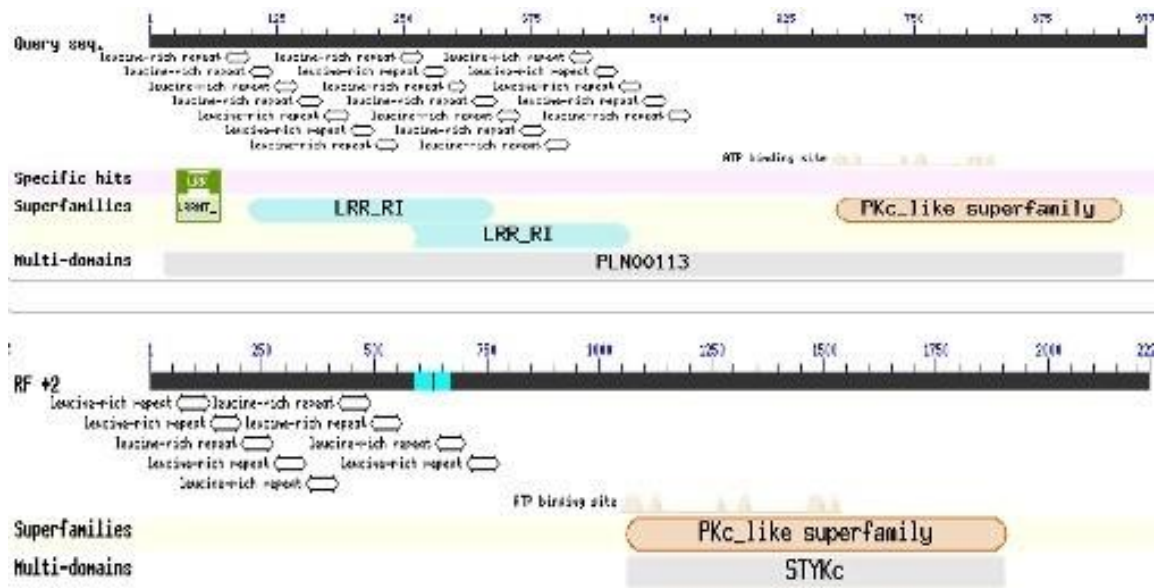


Figura 15: Secuencia AT1G70210.1 relacionada a germinación involucrada en la regulación del ciclo celular en *Arabidopsis* e *Ilex* respectivamente

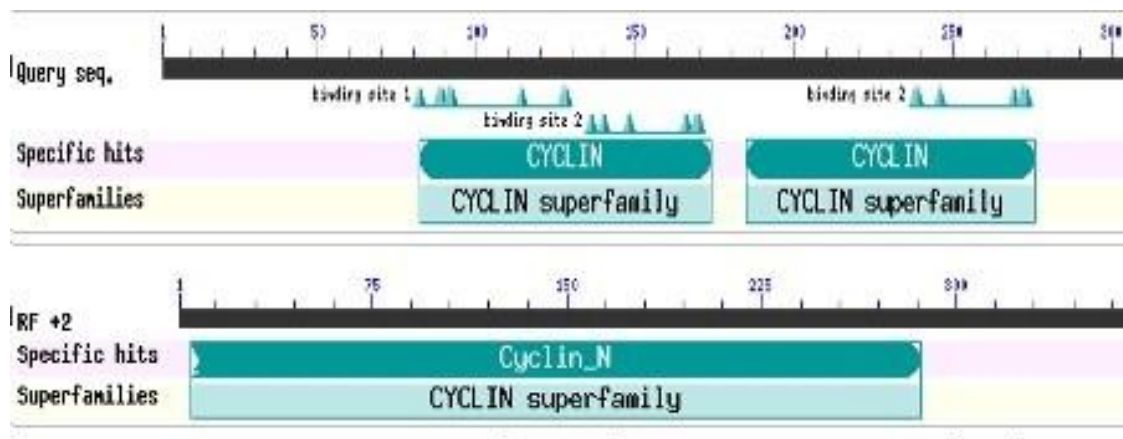


Figura 16: Secuencia AT2G13540.1 relacionada a germinación involucrada en el empalme de ARN, a través de la escisión y ligadura endonucleolítica, fotoperiodismo, floración, morfogénesis, procesamiento miARN primaria, respuesta al ácido abscísico, traducción en *Arabidopsis* e *Ilex* respectivamente

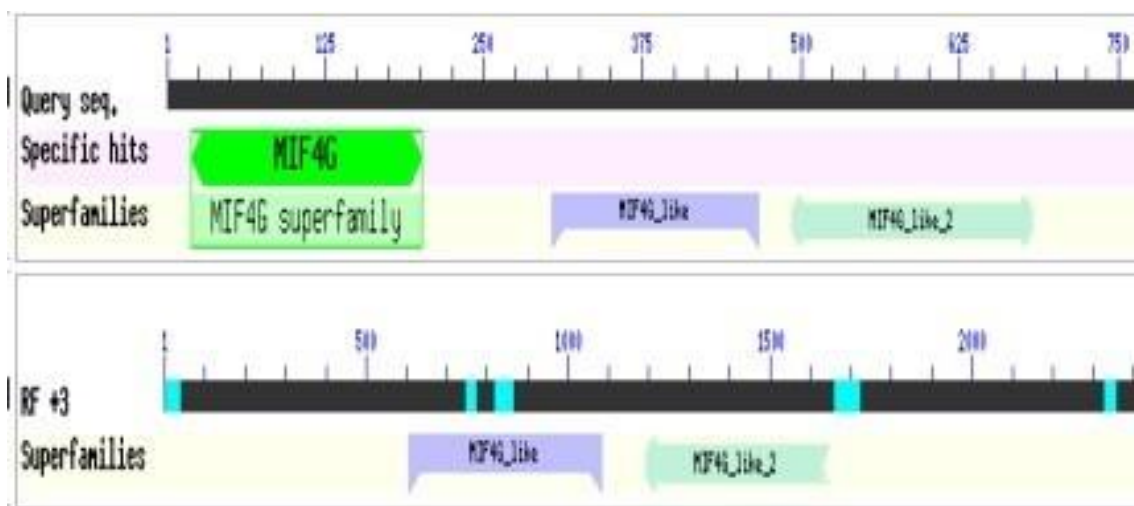


Figura 17: Secuencia AT2G34900.1 relacionada a germinación involucrada en la regulación positiva de la germinación en *Arabidopsis* e *Ilex* respectivamente.

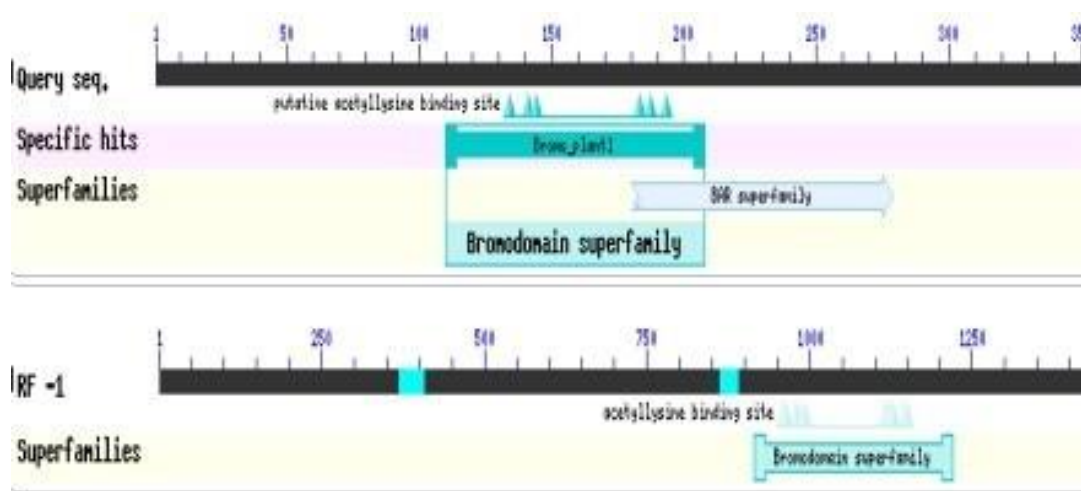
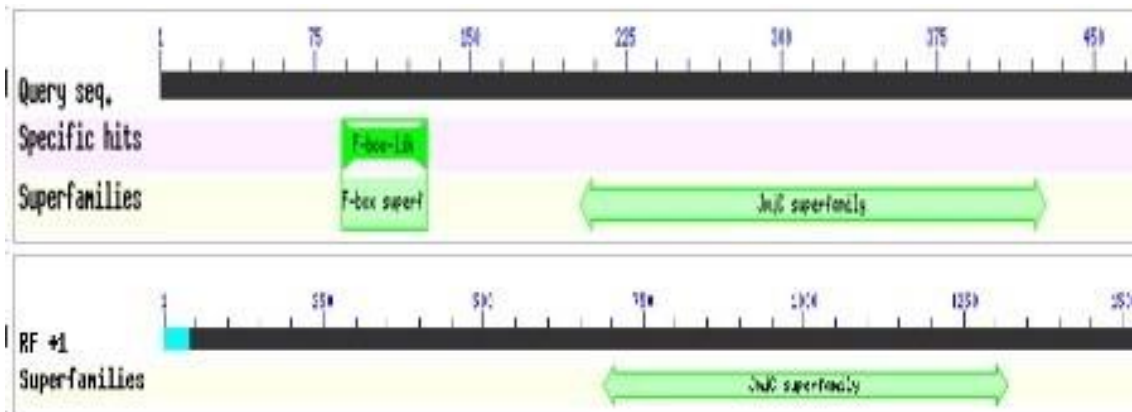


Figura 18: Secuencia AT5G06550.1 relacionada a germinación involucrada en la vía de señalización de receptores de superficie celular, histona H4R3 de metilación, regulación positiva de la germinación de las semillas en *Arabidopsis* e *Ilex* respectivamente.



En esta categoría se pueden ver duplicaciones de dominios en las secuencias de *Arabidopsis* (ver apéndice C y D). La secuencia 5 de dormancia presenta dominios proteicos involucrados en el empalme de ARNm, regulación del ritmo circadiano, desmontaje del complejo de spliceosoma que no está presente en Yerba Mate. Con respecto a las secuencias relacionadas a la germinación se puede mencionar la secuencia 27 que presenta varios dominios duplicados y cuya descripción indica la participación en la proliferación celular, la modificación de la cromatina, vía de señalización activadas por citoquinina, regulación negativa de la vía de señalización activadas por ácido abscísico, regulación negativa de la transcripción, regulación del desarrollo de las raíces laterales, respuesta a auxinas, giberelinas respuesta a, desarrollo de las raíces.

3- Proteínas que presentan dominios en *Ilex paraguariensis* que no están presentes en *Arabidopsis thaliana* (apendice E y F):

Figura 19: Secuencia AT1G02780 relacionada a dormancia involucrada en el desarrollo del embrión que termina en dormancia de las semillas, biogénesis de ribosomas, traducción en *Arabidopsis* e *Ilex* respectivamente.

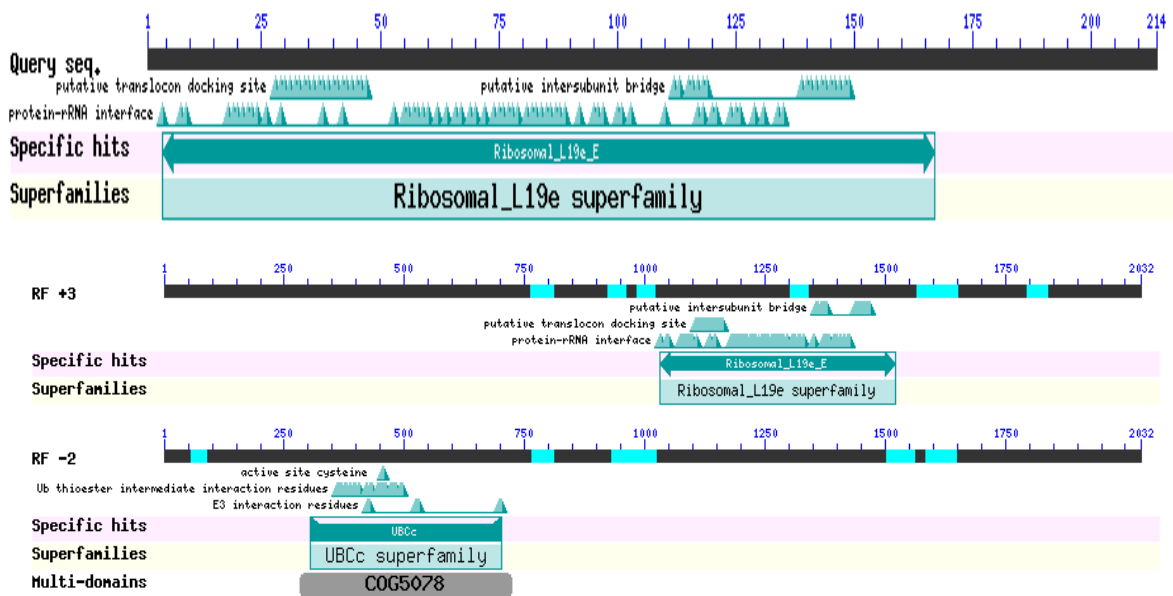


Figura 20: Secuencia AT1G13740.1 relacionada a germinación involucrada en respuesta al ácido abscísico, respuesta a la privación de agua en *Arabidopsis* e *Ilex* respectivamente.

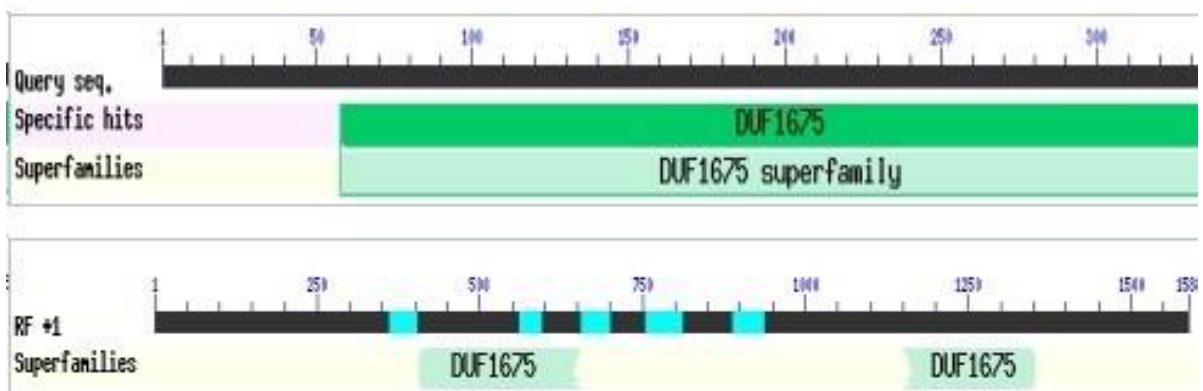
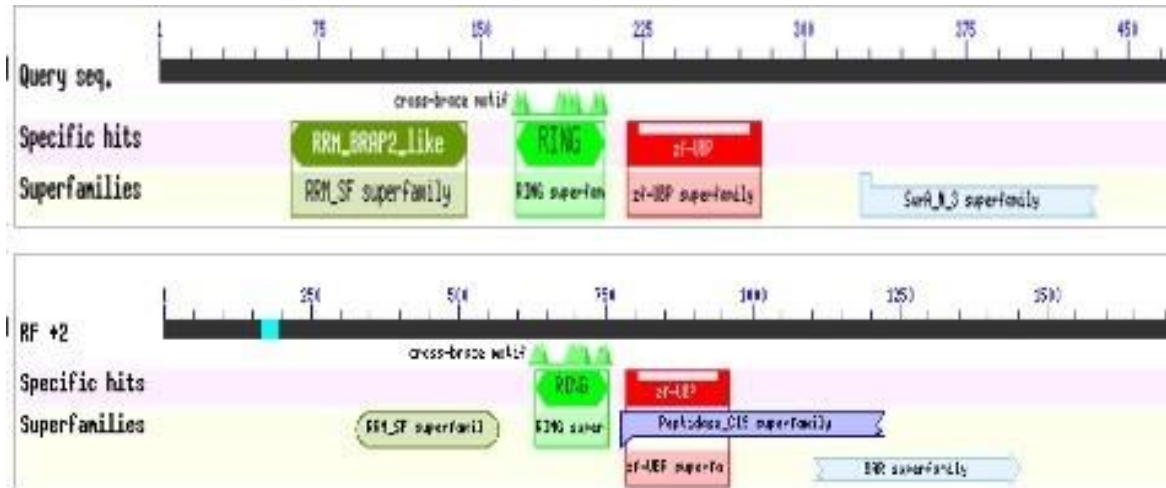


Figura 21: Secuencia AT2G26000.2 relacionada a germinación involucrada a la regulación de la germinación de las semillas en *Arabidopsis* e *Ilex* respectivamente.



En esta categoría encontramos la secuencia 3 de germinación (apendice F) la cual presenta dominios duplicados involucrados en funciones relacionadas a respuestas del ácido abscísico y estrés hídrico en Yerba mate, mientras que la secuencia 28 y 36 presenta dominios no caracterizados en *Arabidopsis* relacionados a la regulación de la germinación de semillas.

4- Proteínas que no presentan dominios conservados evidentes ni en *Arabidopsis thaliana* ni en *Ilex paraguariensis* (apendice G y H):

Figura 22: Secuencia AT4G39850.3 relacionada a dormancia involucrada en la Beta-oxidación de ácidos grasos, proceso catabólico de lípidos, regulación positiva de la germinación de las semillas, proceso de biosíntesis de ubiquinona en *Arabidopsis* e *Ilex* respectivamente.

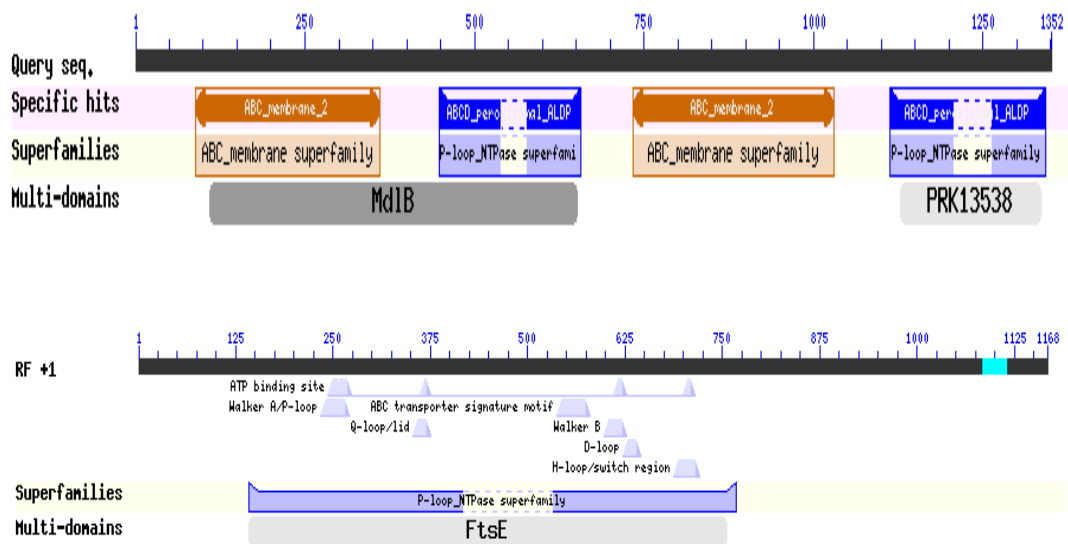
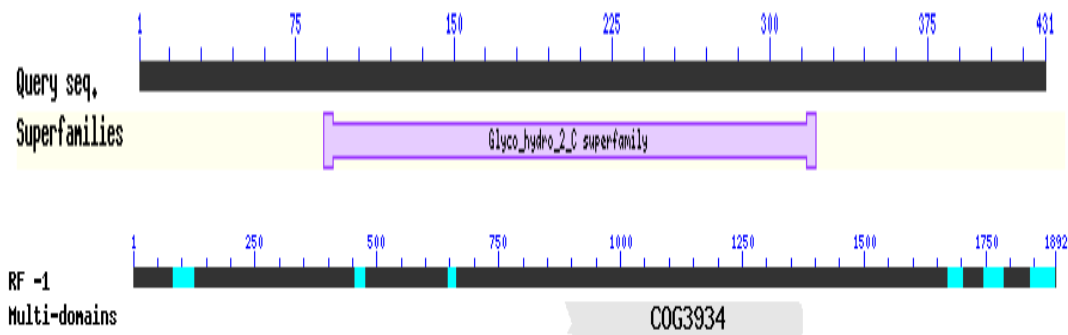


Figura 23: Secuencia AT5G66460.1 relacionado germinación involucrado a procesos catabólico de manano, germinación de semillas en *Arabidopsis* e *Ilex* respectivamente.



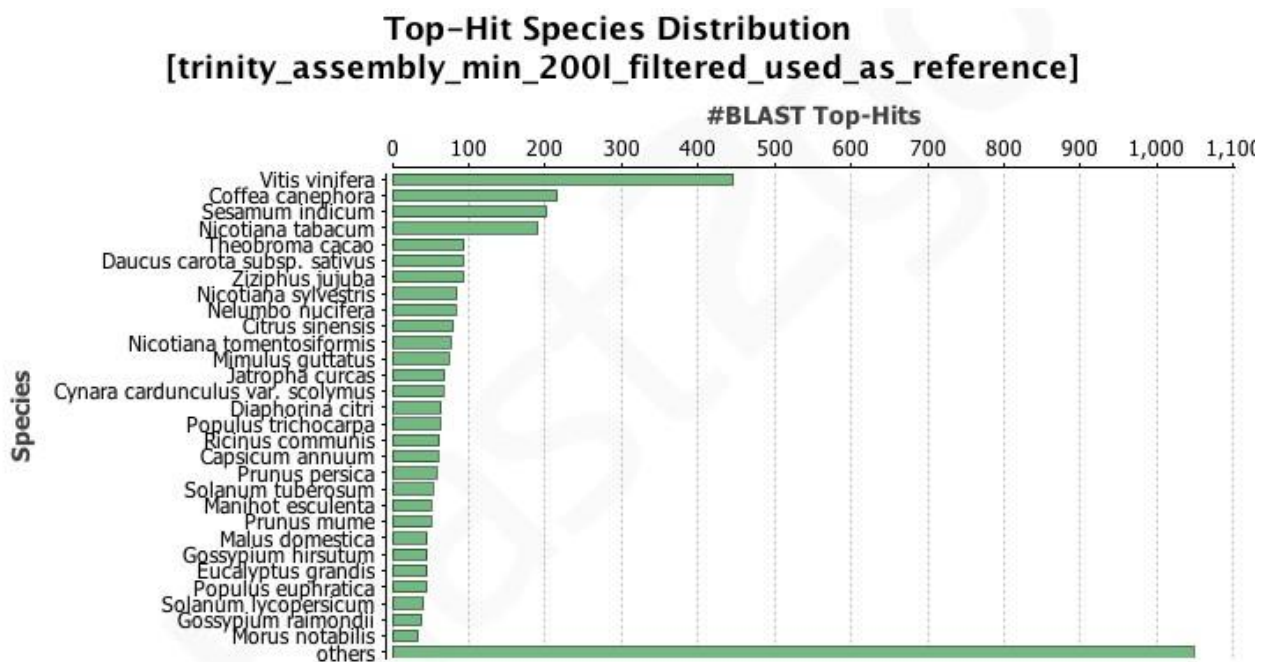
En esta categoría se encuentra la secuencia 25 de dormancia, en Arabidopsis está relacionada a funciones de beta-oxidación de ácidos grasos, catabolismo de lípidos, regulación positiva de la germinación de las semillas, biosíntesis de ubiquinona, mientras que la proteína equivalente en Yerba presenta dominios que están implicadas en la importación de ácidos grasos de cadena larga al peroxisoma, responsables de la translocación de una variedad de compuestos a través de las membranas biológicas (transportadores ABC), transporte y montaje del citocromo C.

La secuencia 83 de germinación se relaciona a proceso catabólico de manano y germinación de semillas en Arabidopsis mientras que la secuencia correspondiente a Yerba Mate describe una proteína con dominios involucrados en el transporte y metabolismo de carbohidratos.

3.2 ANOTACIÓN FUNCIONAL PARCIAL DE GENES DEL TRANSCRIPTOMA DE *Ilex paraguariensis*:

Con la finalidad de conocer el perfil de las vías metabólicas más importantes de *Ilex paraguariensis* fue realizado el análisis de la anotación funcional de una muestra (aproximadamente 5%) del transcriptoma descrito en material y métodos. El gráfico 2 muestra que el organismo con el cual las secuencias de *Ilex paraguariensis* demostraron mayor cantidad de homologías fue *Vitis vinifera*, la uva, seguida por *caffea canephora*.

Gráfico 2 – Mejores Hits encontrados para las secuencias de yerba



En la tabla 1 son citadas la cantidad de enzimas de *Ilex paraguariensis* mapeadas. A pesar de tratarse de una muestra pequeña, fueron mapeadas secuencias del transcriptoma en varias vías metabólicas. Se destacan, por la cantidad de secuencias involucradas, las vías del metabolismo de las purinas (figura 24) y la del metabolismo de Tiamina (figura 25)

Tabla 1: Enzimas mapeadas que participan en algunas vías metabólicas en Yerba Mate.

Vías metabólicas	Secuencias	Enzimas
Purine metabolism	257	161
Thiamine metabolism	224	4
Starch and sucrose metabolism	76	25
Aminobenzoate degradation	50	4
Pyrimidine metabolism	37	10
Amino sugar and nucleotide sugar metabolism	36	18
Phenylpropanoid biosynthesis	31	7
Glycolysis / Gluconeogenesis	30	16
Oxidative phosphorylation	25	6
Pentose and glucuronate interconversions	23	10
Galactose metabolism	22	13
Inositol phosphate metabolism	22	11
Pyruvate metabolism	22	13
Glutathione metabolism	21	10
Cysteine and methionine metabolism	17	12
Arginine and proline metabolism	16	9
Drug metabolism - cytochrome P450	16	5
Fructose and mannose metabolism	15	9
Glycerophospholipid metabolism	15	12
Glycerolipid metabolism	15	10
Methane metabolism	15	7
Phosphatidylinositol signaling system	15	7
Lysine degradation	14	2
Arachidonic acid metabolism	14	6
Metabolism of xenobiotics by cytochrome P450	14	5
Fatty acid degradation	13	7
Cyanoamino acid metabolism	13	4
Aminoacyl-tRNA biosynthesis	12	6
Pentose phosphate pathway	12	6
Carbon fixation in photosynthetic organisms	12	5
Tryptophan metabolism	12	5
Ascorbate and aldarate metabolism	11	7
Tyrosine metabolism	11	8
Valine, leucine and isoleucine degradation	10	5
Isoquinoline alkaloid biosynthesis	10	6

Figura 24: Vía del metabolismo de purina: responsable de las características organolépticas de la especie.

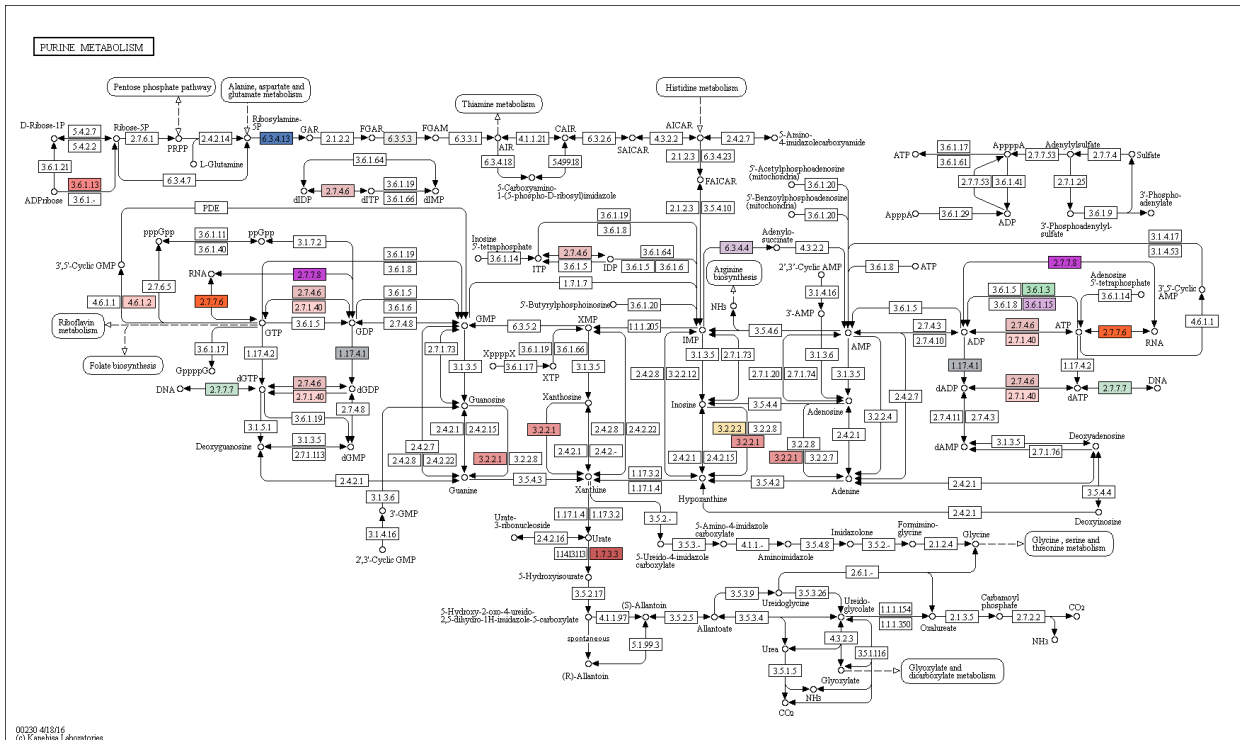
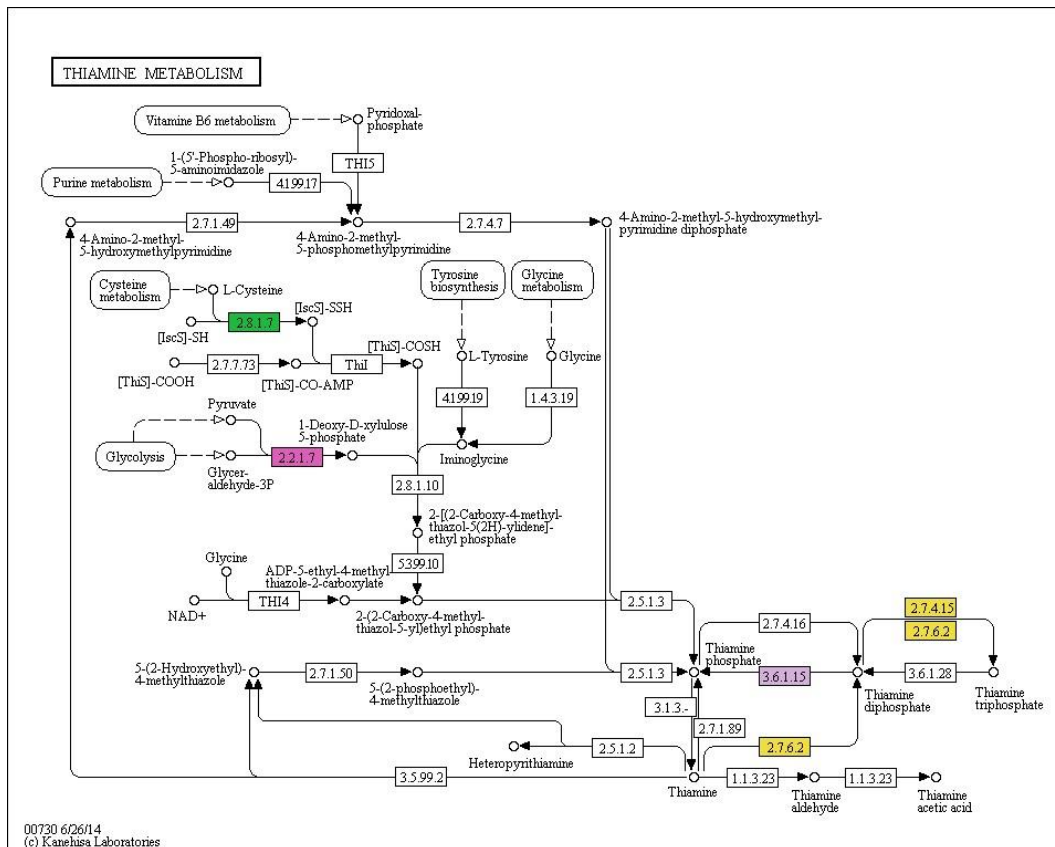


Figura 25: Vía de metabolismo de tiamina, involucrada en la respuesta a estrés biótico y abiótico, como cofactor de varios procesos bioquímicos.



4 DISCUSIÓN:

La Ontología génica, como su nombre lo indica desarrollar una ontología para definir la función de los genes en un organismo a partir de tres conceptos generales: componente celular, función molecular y proceso biológico. Los procesos biológicos implican generalmente transformaciones químicas o físicas que ocurren por la acción de un conjunto de funciones moleculares organizadas (VIERA *et al.*, 2010). Esta información fue consultada y para el caso de secuencias involucradas en dormancia de este estudio pudo verse una correspondencia entre mayor grado de conservación de dominios en aquellos genes con funciones relacionadas a la promoción de la dormancia.

Al ser comparadas secuencias que pertenecen a organismos diferentes pero cumplen las mismas funciones, es posible observar el grado de similitud y que las secuencias de aminoácidos se corresponden en posición para algunas regiones. Las comparaciones realizadas entre Yerba mate y *Arabidopsis* ejemplifican este fenómeno. No obstante el hecho de ser plantas pertenecientes a familias diferentes varias secuencias mantienen gran conservación de sus dominios. Esta situación queda de manifiesto en las secuencias de la primer categoría de resultados (apendice A y B) aun cuando las funciones puedan ser diferentes. Siguiendo con los preceptos de la teoría evolutiva podemos asumir que dichas secuencias son similares porque derivan de un ancestral común y las particularidades que las diferencian son producto del acúmulo de mutaciones.

Las categorías 2 y 3 (apendice C, D,E y F) muestran dominios particulares presentes en uno u otro organismo, estos dominios encontrados dependen directamente del organismo modelo elegido, en este caso *Arabidopsis thaliana*, pero vale destacar que en un escenario diferente donde el genoma de *Ilex paraguariensis* se encontrara disponible, el repertorio de dominios involucrados

podría ser muy diferente teniendo en cuenta que esta especie debe tener una dinámica metabólica diferente. En la categoría 4 (apendice G y H) no hay conservación a nivel de dominio pero si a nivel de secuencia de aminoácidos como lo demostro el analisis BLAST. Es evidente que la homologia detectada mediante la comparacion de secuencias esta dispersa de forma difusa y no es suficiente para evidenciar los dominios presentes en *Arabidopsis*. Una posibilidad es que exista conservación a nivel de conformación tridimensional de proteínas como fue visto en el trabajo de Debat *et al.*, (2014) en el que una búsqueda inicial de la proteína cafeína sintasa mediante BLAST no arrojó resultados satisfactorios. Posteriormente fue comparada la estructura tridimensional de la cafeína sintasa encontrada en *Ilex paraguariensis* con la estructura de esta misma proteína en plantas de Café, *Coffea arabica* sugiriendo un alto grado de conservación cuando superpuestos.

Una manera de categorizar proteínas es usando como base las relaciones evolutivas existentes entre proteínas homólogas, creando una jerarquía en superfamilias, familias y subfamilias (YONA *et al.*, 1999). Superfamilias incluyen proteínas con un origen evolutivo común y en consecuencia estructura similar, dentro de las superfamilias están contenidas familias, y dentro de éstas, subfamilias formadas por proteínas con funciones comunes. Estas categorías jerárquicas inferiores pueden ser denominadas como ortólogos. Este término hace referencia a secuencias homólogas que derivan de un ancestro común y se han separado por especiación. Estas copias divergentes de un mismo gen en las especies derivadas resultantes son ortólogos. Siguiendo la misma definición de Best Bidireccional Hits empleada por SOLIS-CALERO (2008) la cual dice que dos genes son ortólogos si luego de una búsqueda de similaridad en las secuencias de ambos organismos resultan ser más parecidos entre sí cuando comparados con otros genes, podemos inferir que la mayoría de secuencias

comparadas en nuestro trabajo pueden ser vistas como ortólogas, del mismo modo que en el mencionado trabajo de referencia otro requisito impuesto fue el nivel de similaridad presente e-value superior a 1×10^{-5} . Los genes ortólogos generalmente tienen la misma función aunque esto no es regla. Es pertinente suponer que dada la gran similitud general en todas las secuencias analizadas, estamos frente a esta situación, las interacciones génicas que llevan adelante el proceso de dormancia y germinación en ambos organismos son presumiblemente homólogos. Este presupuesto trae consigo el interrogante de si *Arabidopsis* es un buen modelo de comparación o sería oportuno establecer comparaciones con especies más emparentadas al género *Ilex*. En este contexto vale mencionar que un total de 6 secuencias no se encontraron punto de comparación siendo que 2 de ellas corresponden a dominios que aún no fueron caracterizados en *Arabidopsis*. Este resultado está en concordancia con el de OLORIZ *et al.*, (2005) en el que compararon fragmentos de secuencias de Caña de azúcar con una base de datos utilizando el algoritmo BLAST, en su búsqueda por homologías encontraron 5 genes codificantes de proteínas ribosomales, en este trabajo no se encontraron dominios conservados para una de las secuencias, atribuyendo este hecho posiblemente, al pequeño tamaño de la secuencia que no coincide con la región más conservada de la proteína.

La existencia de dominios proteicos no mapeados revela que el proceso de germinación y dormancia no está completamente entendido incluso en un organismo modelo como *Arabidopsis*, lo que deja la posibilidad de futuros estudios en el tema enfocados en la búsqueda de estos dominios no caracterizados en diferentes bases de datos disponibles.

La anotación funcional llevada a cabo por el programa Blast2go permitió conocer la potencialidad del transcriptoma aquí estudiado, a partir del análisis de una muestra de 5% de las secuencias. Uno de los resultados más curiosos fue el

hecho de presentar mayor homología, globalmente, con *Vitis vinifera*. Específicamente, de las secuencias genómicas de organismos ya secuenciados, las que más se parecían a las presentes en el transcriptoma de *Ilex paraguariensis* analizado fueron las de la uva. Esto puede deberse a dos factores: un error de muestreo, considerando que aquí fueron analizados 5% del transcriptoma o al hecho de la importancia en estas dos especies de la producción de polifenoles. Ambas especies tienen destaque en sus propiedades nutricionales por la producción de estos compuestos (XIA *et al.*, 2010; CHANDRA *et al.*, 2004)

Las vías más representadas en esa muestra fueron las del metabolismo de las purinas y la del metabolismo de Tiamina. En el primer caso las enzimas de esa vía producen metabolitos involucrados en las características organolépticas que caracterizan a la especie y que probablemente la han tornado agradable y propicia para el uso alimenticio, por lo que era de cierta forma esperado una estructura genómica que reflejase esta característica. En el caso del metabolismo de la Tiamina, podemos destacar el hecho de que el producto final de esta vía tiene como finalidad la Tiamina que además de estar involucrado en varios procesos bioquímicos como cofactor, diversos autores citan su participación en procesos de respuesta a varios tipos de estrés biótico y abiótico (GOYER *et al.*, 2010). Otro hecho destacable es que las cuatro enzimas de esta vía encontradas en la muestra están representadas por 224 secuencias, pudiendo ser explicado por un error en el proceso de montaje, o por una familia multigénica característica del género o familia.

5 CONCLUSIÓN:

La Yerba Mate se posiciona como un desafío al estudio de los mecanismos envueltos en su biología y la escasa información disponible dificulta la implementación de programas de mejoramiento genético que conlleven a un aumento en la productividad de dicho cultivo. En este trabajo se aborda la temática mediante la comparación de un organismo modelo filogenéticamente distante, lo cual a pesar de ser la mejor opción disponible, no es una comparación ideal.

Si bien se ha conseguido llevar a cabo un mapeo razonable de las secuencias involucradas en dormancia y germinación, queda la duda si las secuencias homólogas en *Ilex* desempeñan las mismas funciones, considerando que para algunas de las secuencias no fueron encontrados dominios conservados ya sea por la distancia filogenética entre estas dos especies comparadas o porque efectivamente no existe la secuencia en Yerba Mate.

Se puede percibir que la maquinaria que lleva a cabo los procesos de germinación y dormancia en *Ilex* fueron mínimamente evidenciados lo cual no es suficiente para tener un entendimiento satisfactorio de la totalidad del proceso.

La disponibilidad de un transcriptoma de 200.000 secuencias de *Ilex paraguariensis* ofrece la posibilidad de explorar más en profundidad las vías metabólicas por detrás de estos y otros procesos hasta el momento poco estudiados, el 5% de dicho transcripto analizado fue suficiente para mapear algunas rutas enzimáticas involucradas en estos y otros procesos.

En este tipo de estudios la bioinformática se presenta como una herramienta que auxilia y da guía en la búsqueda de información sobre los grupos de estudio. No obstante para especies que se encuentran aisladas filogenéticamente como el caso de la Yerba Mate, la biología molecular ofrece soporte valioso en el estudio de la expresión génica que en definitiva ayuda a dar una perspectiva general sobre las especies estudiadas.

6 BIBLIOGRAFÍA:

ABASCAL, F. **Análisis de genomas. Métodos para la predicción y anotación de la función de las proteínas.** Universidad Autónoma de Madrid, 2003.

ALTSCHUL. S; KARLIN. F, Samue, Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. **Proceedings of the National Academy of Sciences**, vol. 87, no 6, p. 2264-2268, 1990

ARNDT, T. 2007. Visual software tools for bioinformatics. **Journal of Visual languages & Computing.** doi:10.1016/j.jvlc..06.001, 2007

ATTWOOD, TK; PARRY. D. **Introducción a la Bioinformática.** Prentice Hall, España. pp. 1-36, 209, 2002.

BACKES, P.; IRGANG, B. **Árvores do Sul: guia de identificação e interesse ecológico.** 1. ed., [S.I.]: Instituto Souza Cruz, 326 p. 2002.

BASKIN, C.C. & BASKIN, J.M. **A classification system for seed dormancy.** seed Science research, 14: 1-16. 2004.

BASTOS, D.H.M., et al., **Phenolic Antioxidants Identified by ESI-MS from Yerba maté (*Ilex paraguariensis*) and Green Tea (*Camelia sinensis*) Extracts.** molecules,;p. 423-432. 12 ,2007.

BENECH-ARNOLD, Roberto L., et al. **Environmental control of dormancy in weed seed banks in soil.** Field Crops Research, vol. 67, no 2, p. 105-122, 2000.

BRASESCO, N. et al., Recent advances on *Ilex paraguariensis* research: minireview. **journal of ethnopharmacology**, v. 136, p. 378-384, 2010.

BRACESCO. N, SANCHEZ. A. G et al., Recent advances on *Ilex paraguariensis* research: minireview. **Journal of Ethnopharmacol**, 136(3): p. 378-384, 2011.

BUDIMAN, M; LOW. R, et al. **Sequencing of the oil palm genespace.** **Proceedings of the Pipoc International Palm Oil Congress**, malaysia. pp. 628-639. SC. 2005

BRAGAGNOLO, N. **Manual técnico de erva-mate.** Secretaria de Estado da Agricultura: EMATER-PR: Instituto de Terras e Cartografia, 1980.

CHANDRA, S. Polyphenolic compounds, antioxidant capacity, and quinone reductase activity of an aqueous extract of *Ardisia compressa* in comparison to

mate (*Ilex paraguariensis*) and green (*Camellia sinensis*) teas. **Journal of agricultural and food chemistry** 3583-3589.11, 2004

CONESA. A; GÖTZ. S, Blast2GO: A comprehensive suite for functional analysis in plant genomics. **International journal of plant genomics**, 2008.

CARDOSO Jr, E.L. **Teores de metilxantinas e compostos fenólicos em extratos de erva-mate (*Ilex paraguariensis* St. Hil.)**. Tese (Doutorado em agronomia). Universidade Estadual de Maringá. Maringá, p.124, 2006.

CUQUEL, F.L., CARVALHO, CHAMA. H, **Avaliação de Métodos de Estratificação para a Quebra de Dormência de Sementes de Erva-mate**. Sci. agric. 51(3): p. 415-421, 1994.

CARDOSO, V.J.M., Conceito E Classificação Da Dormência Em Sementes. **Oecologia Australis**,13(04): p. 619-631. 2009.

DOPAZO, J., & VALENCIA, A. **Bioinformática y genómica**.(2004)

DUTRA, L. F.; HANSEL, F. A.; WENDLING, I. **Introdução ao cultivo in vitro de erva-mate (*Ilex paraguariensis*)**. Embrapa Florestas. Boletim de pesquisa e desenvolvimento, 2008.

DE QUADROS, K. M. **Propagación vegetativa de Yerba Mate (*Ilex paraguariensis* Saint Hilaire–Aquifoliaceae)**.2009

DELLACASSA, E. and BANDONI.L., El mate. **Revista de Fitoterapia**, p. 269-278. 1(4) 2001.

DEBAT, H.J., *et al.*, Exploring the genes of yerba mate (*Ilex paraguariensis* A. St.-Hil.) by NGS and de novo transcriptome assembly. **PLoS One**, p. e109835. 9(10) 2014.

FOWLER, J. A. P; BIANCHETTI, A. **Dormência em sementes florestais**. Colombo, Brazil: Embrapa Florestas, 2000.

FOWLER, J.A.P., STURION J.A, *et al.* **Varição do desenvolvimento Embrionário das Sementes de Erva-mate**. Pesq. Flor. bras. colombo,(54): p. 105-108, 2007.

FINKELSTEIN, R., *et al.*, Molecular aspects of seed dormancy. **Annu Rev Plant Biol.**, 59: p. 387-415, 2008.

FINCH-SAVAGE, W.E. and G. LEUBNER-METZGER, **Seed dormancy and the control of germination**. *New Phytol*, 171(3): p. 501-23. 2006.

FERREIRA, A. G.; CUNHA, G. G.; SILVEIRA, T. S.; HU, C. Y. *In vitro* germination of immature embryos of *Ilex paraguariensis* St. Hil. **Phyton**, v. 52, n. 1, p. 27-32, 1991.

GRIGOLETTI JUNIOR, A.; RODIGHERI, H. R.; MOSELE, S. H.; WIELEWSKI, P. **Estimativa de danos causados por doenças em viveiros de erva-mate, nos Estados do Paraná e Rio Grande do Sul.** Colombo: EMBRAPA-CNPQ, 3 p. 1999.

GIBERTI, G.C., La "Yerba Mate" (*Ilex paraguariensis*, Aquifoliaceae) en tempranos escritos rioplatenses de Bonpland y su real distribución geográfica en Sudamérica austral. **Bonplandia**, 2011. 20(2): p. 203-212.

GOYER, A. **Thiamine in plants: aspects of its metabolism and functions.** Phytochemistry, 71.14 (1615-1624). 2010.

GARZÓN MARTÍNEZ, G. A. **Anotación del transcriptoma foliar de la uchuva (*Physalis peruviana* L.): fruta promisorio de la familia solanaceae.** 2012.

KHAN, A.A. Induction of dormancy in nondormant seeds. **Journal of American Society for Horticultural Science** 119: 408-413, 1996.

LUSCOMBE, N.; GREENBAUM, D.; GERSTEIN, M. **What is Bioinformatics: An introduction and overview.** Disponible en <http://www.csb.yale.edu> 2001. Acceso en 10 de mayo de 2016

MEDEIROS, A.C.d.S.; SILVA, L.C., **Efeitos da secagem na viabilidade das sementes.** Bol. Pesq. FI Colombo, (42): p. 35-46. 2001

MEDEIROS, A.C.d.S., **Dormência em sementes de erva-mate (*Ilex paraguariensis* St. Hil.).** Embrapa, 36: p. 1-25, 1998.

MENNA, A. B. Proposta para ação extensionista na cultura da erva-mate. In: WINGE, H.; FERREIRA, A. G.; MARIATH, J. E. de A.; TARASCONI, L. C. (Org). **erva-mate: biologia e cultura no Cone Sul.** Porto Alegre: Ed. da Universidade federal do Rio Grande do Sul, p. 235-239, 1995.

MOLITERNO, E. A. P. **Variabilidade genética e a eficiência de seleção no caráter dormência de sementes em aveia-preta (*Avena strigosa* Schreb.).** 2008.

MUTIS, G.; ROCHA, P. J.; RESTREPO, S. Generación de la plataforma bioinformática para palma de aceite de Cenipalma. **Revista Palmas**, vol. 29, no 2, p. 19-28. 2008.

MAZUCHOWSKI, J.Z. **Incorporação e exportação de biomassa e de nutrientes pela erva mate *Ilex paraguariensis* St. Hil.** Mestrado – Disciplina "Ciclagem de nutrientes em ecossistemas florestais". Universidade federal do Paraná. Curitiba, 2001.

OLIVEIRA, H. **Qualidade fisiológica e expressão de genes durante o desenvolvimento de sementes de pimenta habanero (*Capsicum chinense* JACQUIM.)**. 2013.

OLIVEIRA, Y.M.M. & ROTTA, E. **A área de distribuição natural de erva-mate (*Ilex paraguariensis* St. Hil.). Seminário sobre atualidades e perspectivas florestais, 10. Silvicultura da erva-mate (*Ilex paraguariensis* St. Hil.)**. Curitiba, 1983, *Anais*. (EMBRAPA-CNPQ Documentos, 15) p. 17-36. Curitiba, 1985.

OLAFSSON, S; Li, X; Wu, S. Operations Research and Data Mining. **European journal of operational research**. doi:10.1016/j.ejor.09.023.2006.

OLORIZ, M. I., *et al.* **Comparación de fragmentos de genes que codifican para proteínas ribosomales de caña de azúcar**. 2005

PEARSON, W.R; LIPMAN, D.J. **Improved tools for biological sequence comparison. Proceedings of the National Academy of Sciences**, vol. 85, no 8, p. 2444-2448, 1988.

PEÑA, M; CARRIZOSA, M. S; *et al.* **La biotecnología, motor de desarrollo para la Colombia de 2015**. Colciencias, primera edición. 264P, 2008.

RICE, DINSOFRE. Dormência em sementes de arroz vermelho. **Ciência Rural**, vol. 29, no 3, 1999.

ROCHA, P. J. **Aportes de la biotecnología al cultivo de la palma de aceite en Pipoc**. Palmas. 26 (4): 53-59. 2005.

ROCHA, P. J. **Biotecnología en el cultivo de palma de aceite: Aspectos sobresalientes en Pipoc**. Palmas. 28(3): 47-55, 2007.

RODIGHERI, H. R. **Rentabilidade Econômica Comparativa entre Plantios florestais e Sistemas Agroflorestais com Erva-mate, Eucalipto e Pinus e as culturas do Feijão, Milho, Soja e Trigo**. Embrapa 26: 1-36, 1997.

SOLORZANO, L.P, **Impacto de la Bioinformática en las ciencias biomédicas**. ACIMED, 11(4). 2003.

SOLÍS-CALERO, C. Identificación in silico de un grupo de secuencias ortólogas conservadas (COS) de *Ipomoea batatas*. **Revista Peruana de Biología**, vol. 15, no 1, p. 79-83, 2008.

THOMPSON, K., *et al.* **Are seed dormancy and persistence in soil related?** Seed Science Research 13: 97-100, 2003.

VILLEGAS, P. **Análisis Genómico y Transcriptómico de *Zea mays* basados en Datos de Next Generation Sequencing**. 2014

VIERA, I; CASTELLANOS, M. *et al.* **Un acercamiento a la ontología de genes y sus aplicaciones.** 2008.

XIA, EN-QIN, *et al.* **Biological activities of polyphenols from grapes.** International journal of molecular sciences : 622-646/ 11.2, 2010.

APÉNDICE

APENDICE A- SECUENCIAS RELACIONADAS A DORMANCIA QUE PRESENTAN LOS MISMOS DOMINIOS EN *A.thaliana* E *I.paraguariensis*

N.º de secuencia	<i>Arabidopsis thaliana</i>	<i>Ilex paraguariensis</i>
2	AT1G04635.1	>comp124415_c0_seq1
3	AT1G04635.1	>comp119138_c0_seq9
7	AT1G48270.1	>comp111449_c0_seq1
14	AT2G34850.1	>comp104688_c0_seq1
15	AT2G38560.1	>comp112445_c0_seq3
16	AT2G45330.1	>comp106287_c1_seq1
17	AT3G01220.1	>comp66361_c0_seq1
19	AT3G50500.2	>comp115878_c0_seq10
27	AT5G37630.1	>comp120119_c1_seq16

APENDICE B: SECUENCIAS RELACIONADAS A GERMINACIÓN QUE PRESENTAN LOS MISMOS DOMINIOS EN *A.thaliana* E *I.paraguariensis*

N.º de secuencias	<i>Arabidopsis thaliana</i>	<i>Ilex paraguariensis</i>
4	AT1G15080.1	>comp108879_c0_seq2
5	AT1G18080.1	>comp103955_c0_seq1
7	AT1G23260.1	>comp100097_c0_seq1
9	AT1G54060.1	>comp93304_c0_seq1
12	AT1G62720.1	>comp108442_c0_seq1
13	AT1G62750.1	>comp94047_c0_seq1
14	AT1G64990.1	>comp118158_c0_seq3
18	AT1G72770.1	>comp99174_c0_seq1
20	AT1G79750.1	>comp122314_c0_seq1
24	AT2G20180.2	>comp120777_c1_seq11
25	AT2G22660.2	>comp109562_c0_seq2
26	AT2G23070.1	>comp104135_c0_seq1
29	AT2G26980.4	>comp102980_c0_seq2
31	AT2G33150.1	>comp103028_c0_seq1
35	AT2G40180.1	>comp107322_c0_seq2
38	AT2G47790.1	>comp91715_c0_seq2
39	AT3G01220.1	>comp66361_c0_seq1
40	AT3G02600.1	>comp103993_c0_seq3
43	AT3G11410.1	>comp115017_c0_seq5
47	AT3G46620.1	>comp87396_c0_seq1
49	AT3G48330.1	>comp105819_c0_seq6
50	AT3G49810.1	>comp117729_c0_seq6
51	AT3G50500.2	>comp115878_c0_seq10

56	AT4G19230.2	>comp103902_c0_seq1
57	AT4G21540.1	>comp112886_c0_seq2
60	AT4G25640.2	>comp122697_c0_seq1
61	AT4G27630.3	>comp118158_c0_seq3
63	AT5G01540.1	>comp121419_c0_seq4
64	AT5G02250.1	>comp120994_c0_seq6
68	AT5G13200.1	>comp88689_c0_seq1
71	AT5G42890.1	>comp62572_c1_seq1
72	AT5G43280.1	>comp121116_c0_seq8
73	AT5G45340.1	>comp115002_c1_seq2
76	AT5G51760.1	>comp110321_c0_seq2
77	AT5G57390.1	>comp103177_c0_seq1
78	AT5G59550.1	>comp87396_c0_seq1
80	AT5G63080.1	>comp111779_c0_seq3

APENDICE C : SECUENCIAS RELACIONADAS A DORMANCA CON DOMINIOS QUE ESTAN PRESENTES EN *A. thaliana* Y NO SE ENCUENTRAN EN *I. paraguariensis*

Nº de secuencias	<i>Arabidopsis thaliana</i>	<i>Ilex paraguariensis</i>
4	AT1G05190	>comp122647_c0_seq1
5	AT1G17070.1	>comp121989_c0_seq14
9	AT1G67890.1	>comp120214_c1_seq4
12	AT2G01280.1	>comp121096_c0_seq3
23	AT4G19230.2	>comp103902_c0_seq1
26	AT5G13200.1	>comp88689_c0_seq1
31	AT5G66880.1	>comp115878_c0_seq6

APENDICE D : SECUENCIAS RELACIONADAS A GERMINACIÓN CON DOMINIOS QUE ESTAN PRESENTES EN *A. thaliana* Y NO SE ENCUENTRAN EN *I. paraguariensis*

Nº de secuencias	<i>Arabidopsis thaliana</i>	<i>Ilex paraguariensis</i>
2	AT1G09970.2	>comp120308_c1_seq2
11	AT1G60190.1	>comp108427_c0_seq1
16	AT1G66730.1	>comp117442_c0_seq2
17	AT1G70210.1	>comp119398_c0_seq11
22	AT2G13540.1	>comp119011_c0_seq2
23	AT2G18790.1	>comp112987_c0_seq1
27	AT2G25170.14	>comp113893_c0_seq1
33	AT2G34900.1	>comp114371_c1_seq2
41	AT3G03450.1	>comp122144_c1_seq1

41	AT3G03450.1	>comp122144_c1_seq1
42	AT3G08030.1	>comp96231_c0_seq1
44	AT3G22490.1	>comp107253_c1_seq2
46	AT3G27380.1	>comp92060_c0_seq1
48	AT3G47620.1	>comp96423_c0_seq1
55	AT3G59220.1	>comp106558_c0_seq1
62	AT4G39850.3	>comp107570_c0_seq1
65	AT5G04040.1	>comp115042_c0_seq1
66	AT5G06550.1	>comp115472_c0_seq12
67	AT5G08450.1	>comp109530_c0_seq2
69	AT5G40650.1	>comp92060_c0_seq1
79	AT5G61780.1	>comp90082_c0_seq1

APENDICE E - SECUENCIAS RELACIONADAS A DORMANCIA QUE PRESENTAN DOMINIOS EN *I. paraguariensis* QUE NO ESTAN PRESENTES EN *A. thaliana*.

Nº de secuencias	<i>Arabidopsis thaliana</i>	<i>Ilex paraguariensis</i>
1	AT1G02780	>comp97230_c0_seq2

APENDICE F - SECUENCIAS RELACIONADAS A GERMINACIÓN QUE PRESENTAN DOMINIOS EN *I. paraguariensis* QUE NO ESTAN PRESENTES EN *A. thaliana*.

Nº de secuencias	<i>Arabidopsis thaliana</i>	<i>Ilex paraguariensis</i>
3	AT1G13740.1	>comp120343_c0_seq1
28	AT2G26000.2	>comp112172_c0_seq3
36	AT2G42160.1	>comp112172_c0_seq3
54	AT3G54810.2	>comp112926_c0_seq2

APENDICE G - SECUENCIAS RELACIONADAS A DORMANCIA QUE NO PRESENTAN DOMINIOS CONSERVADOS EVIDENTES NI EN *Arabidopsis thaliana* NI EN *Ilex paraguariensis*.

N de secuencias	<i>Arabidopsis thaliana</i>	<i>Ilex paraguariensis</i>
25	AT4G39850.3	>comp121594_c0_seq6

APENDICE H- SECUENCIAS RELACIONADAS A GERMINACIÓN QUE NO PRESENTAN DOMINIOS CONSERVADOS EVIDENTES NI EN *Arabidopsis thaliana* NI EN *Ilex paraguariensis*.

Nº de secuencias	<i>Arabidopsis thaliana</i>	<i>Ilex paraguariensis</i>
83	AT5G66460.1	>comp91665_c0_seq1

APENDICE I- SECUENCIAS RELACIONADAS A DORMANCIA EXTRAÍDAS DE TAIR.

- 1>AT1G02780.1
- 2>AT1G04635.1
- 3>AT1G07430.1
- 4>AT1G05190.1
- 5>AT1G17070.1
- 6>AT1G28330.5
- 7>AT1G48270.1
- 8>AT1G54070.1
- 9>AT1G67890.1
- 10>AT1G70910.1
- 11>AT1G72100.1
- 12>AT2G01280.1
- 13>AT2G33830.2
- 14>AT2G34850.1
- 15>AT2G38560.1
- 16>AT2G45330.1
- 17>AT3G01220.1
- 18>AT3G02860.2
- 19>AT3G50500.2
- 20>AT4G00260.1
- 21>AT4G11040.1
- 22>AT4G13560.1
- 23>AT4G19230.2
- 24>AT4G36930.1
- 25>AT4G39850.3
- 26>AT5G13200.1
- 27>AT5G37630.1
- 28>AT5G39750.1
- 29>AT5G45830.1
- 30>AT5G56780.1
- 31>AT5G66880.1

APENDICE J – SECUENCIAS RELACIONADAS A GERMINACIÓN EXTRAIDAS DE TAIR .

1>AT1G03790.1	30>AT2G27380.1	62>AT4G39850.3
2>AT1G09970.2	31>AT2G33150.1	63>AT5G01540.1
3>AT1G13740.1	32- >AT2G33590.1	64>AT5G02250.1
4>AT1G15080.1	33>AT2G34900.1	65>AT5G04040.1
5>AT1G18080.1	34>AT2G36270.1	66>AT5G06550.1
6>AT1G18100.1	35>AT2G40180.1	67>AT5G08450.1
7>AT1G23260.1	36>AT2G42160.1	68>AT5G13200.1
8>AT1G28300.1	37>AT2G46590.2	69>AT5G40650.1
9>AT1G54060.1	38>AT2G47790.1	70>AT5G41560.1
10>AT1G55870.1	39>AT3G01220.1	71>AT5G42890.1
11>AT1G60190.1	40>AT3G02600.1	72>AT5G43280.1
12>AT1G62720.1	41>AT3G03450.1	73>AT5G45340.1
13>AT1G62750.1	42>AT3G08030.1	74>AT5G45830.1
14>AT1G64990.1	43>AT3G11410.1	75>AT5G48870.1
15>AT1G65690.1	44>AT3G22490.1	76>AT5G51760.1
16>AT1G66730.1	45>AT3G23830.1	77>AT5G57390.1
17>AT1G70210.1	46>AT3G27380.1	78>AT5G59550.1
18>AT1G72770.1	47>AT3G46620.1	79>AT5G61780.1
19>AT1G76150.1	48>AT3G47620.1	80>AT5G63080.1
20>AT1G79750.1	49>AT3G48330.1	81>AT5G64210.1
21>AT1G80730.1	50>AT3G49810.1	82>AT5G65420.3
22>AT2G13540.1	51>AT3G50500.2	83>AT5G66460.1
23>AT2G18790.1	52>AT3G51810.1	
24>AT2G20180.2	53>AT3G54770.1	
25>AT2G22660.2.	54>AT3G54810.2	
26>AT2G23070.1	55>AT3G59220.1	
27>AT2G25170.1	56>AT4G19230.2	
28>AT2G26000.2	57>AT4G21540.1	
29>AT2G26980.4	58>AT4G24210.1	
	59>AT4G25140.1	
	60>AT4G25640.2	
	61>AT4G27630.3	

APENDICE K – SECUENCIAS RELACIONADAS A DORMANCIA DE YERBA MATE

1- >comp97230_c0_seq2

2->comp124415_c0_seq1

3->comp119138_c0_seq9

4->comp110416_c0_seq3

5->comp121989_c0_seq14

6- no se uso

7->comp111449_c0_seq1

8-no se uso

9->comp120214_c1_seq4

10-no se uso

11-no se uso

12->comp121096_c0_seq3

13-no se uso

14->comp104688_c0_seq1

15->comp104688_c0_seq1

16->comp106287_c1_seq1

17->comp66361_c0_seq1

18->comp92282_c0_seq1

19->comp115878_c0_seq10

20-no se uso

21-no se uso

22-no se uso

23->comp103902_c0_seq1

24-no se uso

25->comp121594_c0_seq6

26->comp88689_c0_seq1

27-no se uso

28-no se uso

29->comp121638_c0_seq3

30->comp115878_c0_seq6

31->comp115878_c0_seq6

APENDICE L – SECUENCIAS RELACIONADAS A GERMINACION DE YERBA MATE

1->comp102998_c0_seq1	31->comp103028_c0_seq1	63->comp121419_c0_seq4
2->comp120308_c1_seq2	32->comp102136_c0_seq1	64->comp120994_c0_seq6
3->comp120343_c0_seq1	33->comp114371_c1_seq8	65->comp115042_c0_seq1
4->comp108879_c0_seq2	34-no se uso	66->comp115472_c0_seq12
5->comp103955_c0_seq1	35->comp107322_c0_seq2	67->comp109530_c0_seq2
6-no se uso	36->comp112172_c0_seq3	68->comp88689_c0_seq1
7->comp100097_c0_seq1	37-no se uso	69->comp92060_c0_seq1
8-no se uso	38->comp91715_c0_seq2	70-no se uso
9->comp93304_c0_seq1	39->comp66361_c0_seq1	71->comp62572_c1_seq1
10->comp118634_c0_seq13	40->comp103993_c0_seq3	72->comp121116_c0_seq8
11->comp108427_c0_seq1	41->comp122144_c1_seq1	73->comp115002_c1_seq2
12->comp108442_c0_seq1	42->comp96231_c0_seq1	74-no se uso
13->comp94047_c0_seq1	43->comp115017_c0_seq5	75-no se uso
14->comp118158_c0_seq3	44->comp107253_c1_seq2	76->comp110321_c0_seq2
15-no se uso	45-no se uso	77->comp103177_c0_seq1
16->comp117442_c0_seq2	46->comp92060_c0_seq1	78->comp87396_c0_seq1
17->comp119398_c0_seq11	47->comp87396_c0_seq1	79->comp90082_c0_seq1
18->comp99174_c0_seq1	48->comp96423_c0_seq1	80->comp111779_c0_seq3
19->comp109922_c0_seq6	49->comp105819_c0_seq6	81->comp123185_c0_seq1
20->comp122314_c0_seq1	50->comp117729_c0_seq6	82-no se uso
21- no se uso	51->comp115878_c0_seq10	83->comp91665_c0_seq1
22->comp119011_c0_seq2	52-no se uso	
23->comp112987_c0_seq1	53-no se uso	
24->comp120777_c1_seq11	54->comp112926_c0_seq2	
25->comp109562_c0_seq2	55->comp106558_c0_seq1	
26->comp104135_c0_seq1	56->comp103902_c0_seq1	
27->comp113893_c0_seq1	57->comp112886_c0_seq2	
28->comp112172_c0_seq3	58-no se uso	
29->comp102980_c0_seq2	59-no se uso	
30->comp121936_c0_seq6	60->comp122697_c0_seq1	
	61->comp118158_c0_seq3	
	62->comp107570_c0_seq1	