



## **Universidad de Chile**

Facultad de Ciencias Químicas y Farmacéuticas  
Departamento de Bioquímica y Biología Molecular  
Laboratorio de Biología Molecular

### **“ANÁLISIS MOLECULAR Y BIOINFORMÁTICO DE UN MARCADOR DE SEXO DE *BROUSSONETIA PAPYRIFERA* (L.) L’HERIT. EX VENT.”**

**Adrián Rodrigo Fuentes Miranda**

Memoria para optar al Título Profesional de Bioquímico

Director de Tesis

Dr. Sergio Lobos Camus

Laboratorio de Biología Molecular  
Facultad de Ciencias Químicas y Farmacéuticas  
Universidad de Chile

**2015**

*“The cosmos is within us. We are made of star-stuff. We are a way for the universe to know itself.”*

-Carl Sagan

Esta memoria se realizó en el Laboratorio de Biología Molecular del Departamento de Bioquímica y Biología Molecular de la Facultad de Ciencias Químicas y Farmacéuticas de la Universidad de Chile, bajo la dirección del Dr. Sergio Lobos Camus y fue financiada por el proyecto FONDECYT N°1120175 titulado **“Claves para entender la colonización de Isla de Pascua: Patrones de migración de los polinesios en Oceanía Remota a través del análisis genético de la morera de papel (*Broussonetia papyrifera*).”**

Esta memoria ha dado origen a la siguiente comunicación:

**Panel**

**Fuentes, A., Lobos, S., Seelenfreund, D., Seelenfreund, A.**

“Análisis bioinformático del marcador de sexo en la planta *Broussonetia papyrifera*”

XXXI Congreso ANEB (Asociación Nacional de Estudiantes de Bioquímica), 8 de julio a 2 de agosto de 2014, Valparaíso, Chile.

## Agradecimientos

En primer lugar agradezco a toda mi familia por su apoyo incondicional en lo favorable y en lo adverso. Sin mi madre, mi padre o mi hermana nada de lo que he logrado ni en Santiago ni en la vida misma se hubiera podido concretar. A los tres les debo todo lo que soy y lo que llegaré a ser. Los amo.

Gracias a mi querida Tini por estar conmigo en todas: por llegar de Coronel a iluminar mi vida acá en Santiago y por compartir conmigo su compañía, bondad e inteligencia. Las cosas malas que el 2014 puso en mi camino se volvieron nada contigo al lado y lo agradezco infinitamente. Un placer poder haber trabajado contigo aunque haya sido por poco, y haber colaborado en tu trabajo de título. Estoy feliz por tu logro recién alcanzado. Sin duda seguiremos compartiendo buenos momentos en un futuro cercano. Te adoro infinito.

No puedo dejar de agradecer a todos mis amigos de la U por los años de risas y buenos momentos. Gracias a Juan, Daniel, Cris, Pancha, Frenshi, Vale y Aguayo por los no-estudios, las cervezas post pruebas, los chistes y todos los buenos momentos de siempre. Ahora se viene la celebración.

Agradezco a la gente del Laboratorio de Biología Molecular de la Facultad de Ciencias Químicas y Farmacéuticas con los que trabajé, por el buen ambiente de trabajo, especialmente al profesor Sergio por ser mi tutor y a la profesora Daniela por tener siempre muy buena disposición. Agradezco mucho todo el apoyo y comprensión dados en los momentos difíciles que me tocó vivir.

Gracias a Andrea, a la profesora Daniela y al resto del equipo por impulsar el proyecto y por darme la oportunidad de viajar y difundir el proyecto en Rapa Nui.

A Matías Gutiérrez, Ana Morán y a todo el equipo de Bioquímica.cl por ser parte fundamental de mi formación, por creer en mí y darme una base desde donde saltar hacia la ciencia y la vida laboral, por haberme dado la oportunidad de trabajar en ciencia y educación y por ser siempre personas increíbles con las que puedo compartir buenos momentos.

Agradezco a mis amigos de siempre: al Caroca, a Chester y muchos otros, por estar siempre atentos y presentes, y ser los mismos de siempre.

Por último gracias a la Universidad de Chile y a la carrera de Bioquímica por formarme. Me llevo infinitos valores, conocimientos y buenos recuerdos por mi paso por las distintas aulas, laboratorios y bibliotecas.

A todos ustedes muchas gracias, pero muchas gracias.

Y ahora....

...POPOPO por fin puedo decir que soy Bioquímico!!!!

--Adrián.

## ÍNDICE GENERAL

Índice General.....	I
Índice de Figuras.....	IV
Índice de Tablas.....	V
Abreviaturas y siglas utilizadas.....	VII
Resumen.....	VIII
Summary.....	X
<b>1 Introducción.....</b>	<b>1</b>
1.1 Reproducción sexual en plantas.....	1
1.2 Morera de papel.....	3
1.2.1 Descripción.....	3
1.2.2 Importancia económica.....	4
1.2.3 Desarrollo del marcador de sexo.....	5
1.3 Secuencias no codificantes y elementos <i>cis</i> -regulatorios.....	8
1.4 Elementos regulatorios de la transcripción.....	9
1.5 Búsqueda de motivos en secuencias de DNA utilizando herramientas bioinformáticas.....	10
<b>2 Hipótesis.....</b>	<b>12</b>
<b>3 Objetivos.....</b>	<b>12</b>
3.1 Objetivo General.....	12
3.2 Objetivos Específicos.....	12
<b>4 Materiales y Métodos.....</b>	<b>13</b>
4.1 Materiales.....	13
4.1.1 Reactivos e insumos generales.....	13
4.1.2 Material biológico.....	13
4.1.2.1 Muestras de <i>B. papyrifera</i> .....	13
4.1.2.2 Material microbiológico.....	15
4.1.3 Equipos de laboratorio.....	15
4.1.4 Programas herramientas informáticas y bases de datos.....	15

<b>4.2 Métodos.....</b>	<b>16</b>
<b>4.2.1 Amplificación del marcador de sexo de <i>B. papyrifera</i> mediante PCR.....</b>	<b>17</b>
<b>4.2.2 Clonamiento de productos PCR en plasmidio pCR® 2.1-TOPO® TA.....</b>	<b>18</b>
<b>4.2.3 Electrotransformación.....</b>	<b>18</b>
<b>4.2.4 Purificación de plasmidio.....</b>	<b>19</b>
<b>4.2.5 PCR confirmatorio.....</b>	<b>19</b>
<b>4.2.6 Secuenciación.....</b>	<b>20</b>
<b>4.2.7 Análisis de secuencias de productos PCR clonados.....</b>	<b>20</b>
<b>4.2.7.1 Análisis convencional o canónico de la secuencia del marcador de sexo.....</b>	<b>20</b>
<b>4.2.7.2 Búsqueda bioinformática de motivos y genes que poseen los motivos en organismos modelo.....</b>	<b>20</b>
<b>4.2.7.3 Búsqueda bioinformática de factores de transcripción en <i>Saccharomyces cerevisiae</i>.....</b>	<b>21</b>
<b>5 Resultados.....</b>	<b>22</b>
<b>5.1 Amplificación del marcador de sexo de <i>B. papyrifera</i> mediante PCR.....</b>	<b>22</b>
<b>5.2 Clonamiento de productos PCR de la región del marcador de sexo.....</b>	<b>23</b>
<b>5.3 Búsqueda de características de elemento regulatorio de la expresión génica.....</b>	<b>51</b>
<b>5.3.1 Búsqueda de homología con secuencias depositadas en GenBank.....</b>	<b>51</b>
<b>5.3.2 Búsqueda analítica de motivos putativos de unión de factores de transcripción y genes en organismos modelo usando la Suite MEME.....</b>	<b>52</b>
<b>5.3.2.1 Utilización de MEME para buscar motivos en la secuencia del marcador de sexo.....</b>	<b>52</b>
<b>5.3.2.2 Identificación de genes que comparten los motivos encontrados en el marcador de sexo de <i>B. papyrifera</i> mediante utilización de MAST.....</b>	<b>54</b>
<b>5.3.2.3 Búsqueda de factores de transcripción de <i>S. cerevisiae</i> que se unen a los motivos identificados en el marcador de sexo de <i>B. papyrifera</i>.....</b>	<b>67</b>
<b>6 Discusión.....</b>	<b>73</b>
<b>6.1 Búsqueda de homología con otras secuencias depositadas en GenBank.....</b>	<b>73</b>
<b>6.2 Variación en la secuencia del marcador según el sexo del individuo de <i>B. papyrifera</i> ....</b>	<b>74</b>

6.3 Utilización de MEME para buscar motivos en la secuencia del marcador de sexo.....	76
6.4 Factores de transcripción de <i>S. cerevisiae</i> que se unen a los motivos identificados en la secuencia del marcador de sexo de <i>B. papyrifera</i> .....	76
6.5 Proyecciones.....	78
7 Conclusiones.....	79
8 Bibliografía.....	80

## Índice de Figuras

Figura 1: Estrategias reproductivas en plantas.....	2
Figura 2: Imágenes de flores y fruto de <i>B. papyrifera</i> .....	4
Figura 3: Experimento de AFLP para la identificación de un marcador sexo-específico de <i>B. papyrifera</i> .....	6
Figura 4: Secuencia del marcador de sexo de <i>B. papyrifera</i> acceso GenBank: HQ202152.1.....	7
Figura 5: Esquema del método de determinación de sexo en <i>B. papyrifera</i> .....	7
Figura 6: Esquema de la región regulatoria de un gen eucarionte típico.....	10
Figura 7: Origen geográfico de las muestras de <i>B. papyrifera</i> .....	14
Figura 8: Diagrama de flujo de la metodología empleada en esta memoria de título.....	16
Figura 9: Esquema de la amplificación del marcador de sexo mediante PCR.....	17
Figura 10: Amplificación del marcador de sexo de <i>B. papyrifera</i> .....	22
Figura 11: Análisis comparativo de la amplificación del marcador de sexo de <i>B. papyrifera</i> en diferentes especies.....	23
Figura 12: Alineamiento múltiple de las secuencias del marcador de sexo de <i>B. papyrifera</i> .....	25
Figura 13: Alineamiento múltiple de secuencias del marcador de sexo de <i>B. papyrifera</i> por individuo..	28
Figura 14: Alineamiento múltiple de secuencias del marcador de sexo de <i>B. papyrifera</i> , agrupadas según sexo del individuo.....	40
Figura 15: Dendrograma Kimura 80/Neighbor Joining de las secuencias del marcador de sexo de <i>B. papyrifera</i> .....	48
Figura 16: Dendrograma Kimura 80/Neighbor Joining con grupos destacados de secuencias idénticas del marcador de sexo de <i>B. papyrifera</i> .....	49
Figura 17: Dendrograma Kimura 80/Neighbor Joining de las secuencias separadas por sexo.....	50
Figura 18: Búsqueda de secuencias homólogas usando BLAST.....	51
Figura 19: Motivos encontrados en la secuencia del marcador de sexo de <i>B. papyrifera</i> usando MEME.....	53
Figura 20: Esquema de los <i>loci</i> del marcador de sexo en los cromosomas sexuales de <i>B. papyrifera</i> .....	75



## Índice de Tablas

Tabla 1: Muestras de <i>B. papyrifera</i> usadas en la amplificación del marcador de sexo.....	14
Tabla 2: Partidores utilizados en la amplificación del marcador de sexo.....	17
Tabla 3: Colonias seleccionadas (clones) por cada ejemplar de <i>B. papyrifera</i> .....	24
Tabla 4: Secuencias clonadas por individuo y número de secuencias diferentes identificadas.....	46
Tabla 5: Porcentaje de identidad de las secuencias por individuo de <i>B. papyrifera</i> .....	47
Tabla 6: Genes de <i>A. thaliana</i> que poseen los motivos encontrados en el marcador de sexo de <i>B. papyrifera</i> en sus promotores .....	55
Tabla 7: Genes de <i>S. cerevisiae</i> que poseen los motivos encontrados en el marcador de sexo de <i>B. papyrifera</i> en sus promotores.....	56
Tabla 8: Genes de <i>A. thaliana</i> que poseen el motivo 1 del marcador de sexo de <i>B. papyrifera</i> en sus promotores.....	57
Tabla 9: Genes de <i>A. thaliana</i> que poseen el motivo 8 del marcador de sexo de <i>B. papyrifera</i> en sus promotores.....	57
Tabla 10: Genes de <i>A. thaliana</i> que poseen el motivo 15 del marcador de sexo de <i>B. papyrifera</i> en sus promotores.....	58
Tabla 11: Genes de <i>A. thaliana</i> que poseen una combinación de los motivos del marcador de sexo de <i>B. papyrifera</i> encontrados en sus promotores.....	60
Tabla 12: Genes de <i>S. cerevisiae</i> que poseen el motivo 1 del marcador de sexo de <i>B. papyrifera</i> en sus promotores.....	62
Tabla 13: Genes de <i>S. cerevisiae</i> que poseen el motivo 8 del marcador de sexo de <i>B. papyrifera</i> en sus promotores.....	63
Tabla 14: Genes de <i>S. cerevisiae</i> que poseen el motivo 15 del marcador de sexo de <i>B. papyrifera</i> en sus promotores.....	64
Tabla 15: Genes de <i>S. cerevisiae</i> que poseen una combinación de los motivos encontrados en el marcador de sexo de <i>B. papyrifera</i> en sus promotores.....	66
Tabla 16: Factores de transcripción de <i>S. cerevisiae</i> que se unen al motivo 1 del marcador de sexo de <i>B. papyrifera</i> .....	68
Tabla 17: Función de factores de transcripción de <i>S. cerevisiae</i> que se unen al motivo 1 del marcador de sexo de <i>B. papyrifera</i> .....	68

<b>Tabla 18: Factores de transcripción de <i>S. cerevisiae</i> que se unen al motivo 8 del marcador de sexo de <i>B. papyrifera</i>.....</b>	<b>69</b>
<b>Tabla 19: Función de factores de transcripción de <i>S. cerevisiae</i> que se unen al motivo 8 del marcador de sexo de <i>B. papyrifera</i>.....</b>	<b>70</b>
<b>Tabla 20: Factores de transcripción de <i>S. cerevisiae</i> que se unen al motivo 15 del marcador de sexo de <i>B. papyrifera</i>.....</b>	<b>71</b>
<b>Tabla 21: Función de factores de transcripción de <i>S. cerevisiae</i> que se unen al motivo 15 del marcador de sexo de <i>B. papyrifera</i>.....</b>	<b>72</b>

## Abreviaturas y siglas utilizadas

SNP	Polimorfismo de nucleótido simple ( <i><u>S</u>ingle <u>N</u>ucleotide <u>P</u>olymorphism</i> )
DNA	Ácido desoxirribonucleico ( <i><u>D</u>eoxyribo<u>n</u>ucleic <u>A</u>cid</i> )
RNA	Ácido ribonucleico ( <i><u>R</u>ibo<u>n</u>ucleic <u>A</u>cid</i> )
dNTP	Trifosfato de desoxinucleótido ( <i><u>D</u>eoxy<u>n</u>ucleotide <u>T</u>riphosphate</i> )
EMSA	<i><u>E</u>lectrophoretic <u>M</u>obility <u>S</u>hift <u>A</u>ssay</i>
Taq Pol	DNA polimerasa recombinante de <i>Thermus aquaticus</i>
RNA Pol II	RNA polimerasa del tipo II
MEME	<i><u>M</u>ultiple <u>E</u>M for <u>M</u>otif <u>E</u>licitation</i>
MAST	<i><u>M</u>otif <u>A</u>lignment and <u>S</u>earch <u>T</u>ool</i>
BLAST	<i><u>B</u>asic <u>L</u>ocal <u>A</u>lignment <u>S</u>earch <u>T</u>ool</i>
ZOOPS	<i><u>Z</u>ero <u>O</u>r <u>O</u>ne <u>P</u>er <u>S</u>equence</i>
PCR	Reacción en cadena de la polimerasa ( <i><u>P</u>olymerase <u>C</u>hain <u>R</u>eaction</i> )
AFLP	Polimorfismo de Longitud de Fragmento Amplificado ( <i><u>A</u>mplified <u>F</u>ragment <u>L</u>ength <u>P</u>olymorphism</i> )
SCAR	Región amplificada de secuencia caracterizada ( <i><u>S</u>equence <u>C</u>haracterized <u>A</u>mplified <u>R</u>egion</i> )
SNC	Secuencia no codificante
IPTG	Isopropil-β-D-1-tiogalactopiranosido

## Resumen

La morera de papel (*Broussonetia papyrifera* (L.) L'Herit. ex Vent) es una planta dioica nativa del sudeste asiático, donde se cultiva y utiliza con fines medicinales y como fuente de fibra para la fabricación de papel. Esta especie fue introducida en Oceanía también como fuente de fibra para la fabricación de textiles. Dado el valor económico diferencial de los individuos de morera de papel en China, investigadores de ese país desarrollaron un marcador de sexo de esta planta en base a la técnica de AFLP. Este marcador corresponde a una secuencia de 425 pares de bases.

A partir de esta secuencia en nuestro laboratorio se desarrolló un protocolo de PCR dúplex que permite determinar el sexo de *B. papyrifera*, de modo de generar dos productos de amplificación en individuos masculinos y uno en individuos femeninos. El control interno de este experimento está determinado por el producto PCR de aproximadamente 425 pb, amplificado tanto en individuos masculinos como femeninos. La secuenciación de este producto PCR fue posible en individuos femeninos, pero no en ejemplares de plantas masculinas, probablemente debido a la presencia de más de un producto de amplificación. El marcador de sexo es por tanto, diferente en individuos de distinto sexo de *B. papyrifera*.

En base a estos antecedentes se planteó la siguiente hipótesis: “El marcador de sexo de *B. papyrifera* representa una región de DNA funcional que está involucrado en expresión génica diferencial.” Para verificar la hipótesis se planteó como objetivo caracterizar la secuencia correspondiente al marcador de sexo de *B. papyrifera* en individuos de ambos sexos, e identificar características de elementos regulatorios en *cis* implicados en una expresión génica diferencial, los que podrían promover el desarrollo de fenotipos masculinos y femeninos en esta especie dioica.

Los análisis con este marcador molecular de sexo se realizaron con el material genético de muestras foliares del banco genómico del Laboratorio de Biología Molecular del Departamento de Bioquímica y Biología Molecular de la Facultad de Ciencias Químicas y Farmacéuticas de la Universidad de Chile. Se amplificó el marcador de sexo de *B. papyrifera* de 6 ejemplares femeninos provenientes de Taiwán, Hawái, Vietnam, Rapa Nui y China y seis ejemplares masculinos de Taiwán, Hawái, Vietnam y China. Se clonaron los productos de amplificación para separar las distintas secuencias presentes en cada uno de los individuos. Los clones se secuenciaron y alinearon entre sí, revelándose una multiplicidad de variantes del marcador de sexo en cada individuo. Las secuencias de los individuos masculinos presentan

mayor diversidad genética que las secuencias de individuos femeninos, resultado que se confirmó mediante la construcción de dendrogramas, donde algunas secuencias masculinas formaron clados separados. Por el contrario, las secuencias correspondientes a individuos femeninos poseen una diversidad genética menor.

La ausencia de producto de amplificación en otras especies cercanas (Moráceas), sugiere que el marcador de sexo de *B. papyrifera* es especie específico. La búsqueda de secuencias homólogas indicó que el marcador de sexo no posee homólogos depositados en GenBank.

Al no existir secuencias para comparar estructura y función, se realizó un análisis no canónico de búsqueda de funcionalidad de las secuencias del marcador de sexo de *B. papyrifera* en base a la metodología descrita por Lobos y cols. (2010). En la presente memoria se utilizó la suite MEME para buscar motivos con posible significancia biológica en las secuencias analizadas. Luego, una vez identificados 15 motivos, se buscaron genes de *Saccharomyces cerevisiae* y *Arabidopsis thaliana* que poseen los motivos encontrados en el marcador de sexo de *B. papyrifera*. Los resultados de este análisis indican que las secuencias del marcador de sexo varían en su composición de motivos. Las secuencias masculinas y femeninas comparten los 15 motivos, sin embargo dos de estos motivos están ausentes en determinados secuencias de individuos masculinos. Estas mismas secuencias forman un clado separado en el dendrograma.

Los motivos identificados en el marcador de sexo de *B. papyrifera* se encontraron en diversos genes de *A. thaliana* y *S. cerevisiae*, los cuales podrían estar implicados en el desarrollo de fenotipos diferenciados. También se identificaron los factores de transcripción de *S. cerevisiae* que se unen a dichos motivos del marcador de sexo de *B. papyrifera*. Estos factores de transcripción son de carácter global y regulan diversas funciones en *S. cerevisiae*, aportando información al análisis predictivo presentado en esta memoria.

Los datos de las bases de datos de *S. cerevisiae*, sugieren que existen múltiples factores de transcripción que se unirían a la secuencia del marcador de sexo de *B. papyrifera*. De este modo el trabajo realizado sugiere un rol hipotético de regulador de la expresión génica y además valida el modelo analítico empleado en esta memoria. En el futuro será de interés analizar el contexto genómico de la secuencia estudiada y estudiar *in vitro* la unión de factores de transcripción para comprobar experimentalmente la unión proyectada *in silico* en esta memoria.

## SUMMARY

### **Molecular and bioinformatic analysis of a sex marker of *Broussonetia papyrifera* (L.) L'Herit. Ex Vent.**

Paper mulberry (*Broussonetia papyrifera* (L.) L'Herit. ex Vent) is a dioecious plant native of Southeast Asia, where it is grown and used for medicinal purposes and as a source of fiber for making paper or textiles. This species was introduced into Oceania by the ancient Polynesian voyagers also for making textiles. Given the distinct economic value of male and female *B. papyrifera* individuals in China, Chinese investigators developed a sex marker based on AFLP, corresponding to a 425 bp sequence. Using this sequence as a starting point, a PCR duplex protocol was developed in our laboratory that allows the assessment of gender of *B. papyrifera* plants. In male plants two amplification products are obtained, while only one amplification band is observed in female plants. The 425 bp amplification product observed in both male and female individuals corresponds to an internal control. Sequencing of this PCR product was possible only in female samples, probably due to presence of more than one amplification product in male plants.

Based on these results, we proposed the following hypothesis: "The sex marker of *B. papyrifera* is a functional DNA region involved in differential gene expression". The goal of this work was to characterize the sequence of this sex marker in several individuals sampled in Asia and in the Pacific, and to identify putative regulatory *cis* elements involved in differential gene expression, which may define male and female phenotypes in this dioecious species.

DNA extracted from leaves and stored in the genomic bank of the Molecular Biology Laboratory, Faculty of Chemical and Pharmaceutical Sciences, University of Chile was used for all analyses. The *B. papyrifera* sex marker of six female samples from Taiwan, Hawaii, Vietnam, Rapa Nui and China and of six male samples from Taiwan, Hawaii, Vietnam and China was amplified. These amplification products were cloned to resolve the distinct sequences present in each sample. Clones were sequenced and aligned, revealing multiplicity of sex marker sequences in each sample. Male sequences presented higher genetic variability than female sequences, as seen in dendrograms, as only male sequences cluster in separated groups. The results of this study identified thirty five distinct versions of the sex marker sequence, from a total of forty nine sequences.

The absence of amplification products in closely related species (Moraceae), suggests that the *B. papyrifera* sex marker may be species-specific. A search for homologous sequences in the GenBank database did not identify any similar sequence.

As no homologous sequences are available to compare structure and function, a non-canonical analysis was performed to search for functionality of the *B. papyrifera* sex marker (Lobos *et al.*, 2010). In this thesis, suite MEME was used to search for motifs of biological significance. Results showed that the sex marker sequences contain 15 motifs that are shared by male and female sequences, although two motifs are absent in some male sequences. These sequences also cluster in separate groups in Kimura 80 dendrograms.

Next, identified motifs were searched in *Saccharomyces cerevisiae* and *Arabidopsis thaliana* gene databases containing these motifs. The identified motifs were found in a number of *S. cerevisiae* and *A. thaliana* genes which are involved in the development of differentiated phenotypes. Also, transcription factors (TFs) that bind to these motifs in the sex marker sequence were identified. These TFs regulate diverse functions in *S. cerevisiae*, adding information to the predictive analysis of this work.

These results indicate that the sex marker of *B. papyrifera* corresponds to several genomic sequences, where male sequences present higher genetic diversity than female sequences, as revealed by identity percentages and dendrograms. The sequences found are distinct among female and male sequences. In addition, the sequences derived from male samples that lack the two of the 15 motifs, cluster in separate groups.

To summarize, the data from the *S. cerevisiae* database suggest the presence of TF binding sites in motifs within the sequence of paper mulberry sex marker. Therefore, this work suggests a possible role for this sequence as a regulator of gene expression and also validates the analytical model employed in this thesis. It will be interesting to analyze the genomic context of the sex marker sequence and evaluate the binding of these transcription factors *in vitro*, to experimentally verify the proposed *in silico* binding.

# 1 Introducción:

## 1.1 Reproducción sexual en plantas

La reproducción sexual, es decir la unión de genomas vía meiosis y fusión de gametos, es universal en eucariontes e involucra el funcionamiento de complejos mecanismos de desarrollo de órganos sexuales especializados y diferenciación fenotípica, evitando la autofertilización y la aparición de fenotipos recesivos deletéreos [Bachtrog y cols., 2014].

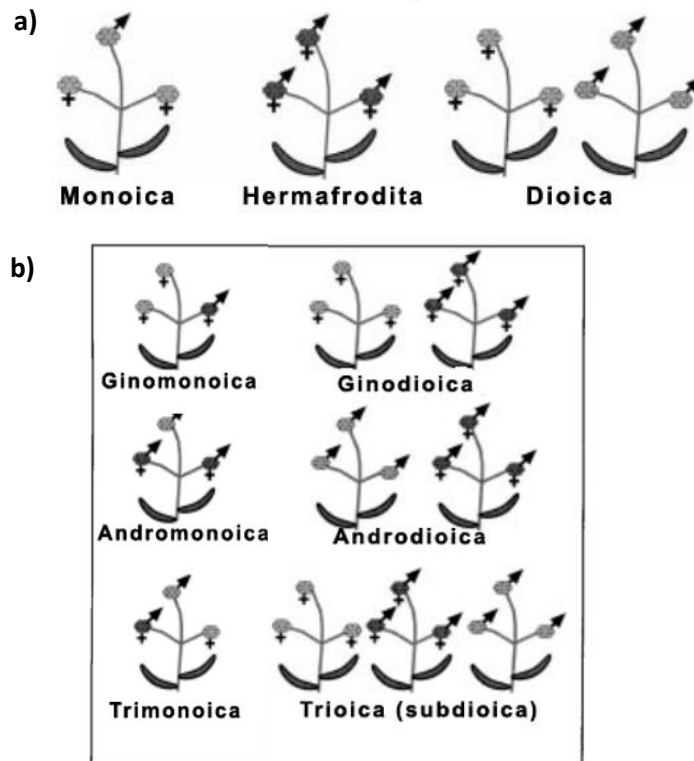
El hermafroditismo es poco frecuente en animales, correspondiente a una cifra cercana al 5% de especies, pues en la mayoría de los animales las especies poseen sexos separados, o sea, individuos ya sea con órganos femeninos o con órganos masculinos [Milewicz y Sawicki, 2012]. En cambio, en plantas el hermafroditismo es muy común, con una cifra de alrededor del 90% de las especies descritas [Milewicz y Sawicki, 2012; Ainsworth, 2000]. Las especies vegetales que componen el 10% aproximado restante presentan diversas estrategias reproductivas.

Estas estrategias de separación de órganos sexuales en plantas incluyen dos casos opuestos: el desarrollo de individuos con ambos órganos sexuales separados en el mismo individuo, las cuales son llamadas monoicas, y el desarrollo de individuos con órganos sexuales masculinos y femeninos en individuos separados, especies que se denominan dioicas (**Figura 1a**). Las especies dioicas corresponden a un 6% aproximado del total de plantas con flores y son más abundantes en dicotiledóneas que en monocotiledóneas (8% vs 5% respectivamente) [Renner y Ricklefs, 1995]. Entre ambas estrategias de separación de órganos sexuales en plantas existen casos intermedios que se ilustran en la **Figura 1b**, como el de las especies androdioicas, que poseen flores masculinas y hermafroditas en individuos separados y las especies ginodioicas, que poseen flores femeninas y hermafroditas en individuos separados. Otras estrategias intermedias incluyen a las especies ginomonoicas, que desarrollan flores femeninas y hermafroditas, y las andromonoicas que poseen flores masculinas y hermafroditas. Las especies trimonoicas poseen flores masculinas, femeninas y hermafroditas en el mismo individuo y por último las especies trioicas o subdioicas que poseen individuos con flores masculinas, femeninas y hermafroditas por separado [Bachtrog y cols., 2014; Charlesworth, 2002; Ainsworth, 2000].

Se desconocen los mecanismos evolutivos y moleculares que han dado origen a las distintas estrategias reproductivas en especies vegetales. Se postula que para alcanzar el desarrollo de individuos con fenotipos sexuales separados en especies dioicas, se asume la aparición en paralelo e independiente de alelos mutados que provoquen esterilidad de órganos sexuales femeninos en individuos masculinos,



así como esterilidad de órganos sexuales masculinos en individuos femeninos [Charlesworth, 1984; Ainsworth, 2000]. La existencia de especies con las estrategias reproductivas intermedias descritas corrobora esta hipótesis y sugiere la idea de la existencia de *loci* determinantes de sexo, tal y como ocurre en animales, aunque estos aún no se describen en plantas [Bachtrog y cols., 2014; Charlesworth, 2002].



**Figura 1: Estrategias reproductivas en plantas.**

Panel a) Se ilustran los individuos de especies monoica, hermafrodita y dioica. Panel b) Estrategias reproductivas consideradas intermedias.

El carácter dioico de las plantas ha surgido en forma independiente en la evolución de distintas familias. Esta visión se ve apoyada también por la existencia de la variedad de estrategias reproductivas en plantas, muchas de ellas consideradas intermedias entre el carácter hermafrodita y el dioico [Charlesworth, 1984]. Entre las especies de plantas dioicas conocidas se encuentran *Cannabis sativa*, *Carica papaya* [Bachtrog y cols., 2014] y *Broussonetia papyrifera* [Matthews, 1996].

## 1.2 Morera de papel:

### 1.2.1 Descripción.

*Broussonetia papyrifera* (L.) L'Herit. ex Vent, conocida comúnmente como morera de papel es una planta nativa del este y sudeste asiático usada con múltiples fines como fuente de fibras en la fabricación de textiles, con fines medicinales y como fuente de materia prima en la obtención de papel en la Antigua China [Ewins, 1987; Clark, 1965], de lo que deriva su nombre vernacular.

Pertenece a la familia Moraceae, que comprende 37 géneros y aproximadamente 1100 especies distribuidas en regiones templadas y tropicales de todo el mundo. Originalmente Carl Linnaeus clasificó esta especie en el género *Morus*, y la denominó *Morus papyrifera*. Esta clasificación fue cuestionada por Pierre Broussonet y Jean Lamarck, por lo que finalmente se situó en un género nuevo denominado *Broussonetia* por el francés Charles-Louis L'Héritier [Barker, 2002].

Su clasificación taxonómica actual es:

- Clase: Equisetopsida
- Subclase: Magnoliidae
- Superorden: Rosanae
- Orden: Rosales
- Familia: Moraceae
- Género: *Broussonetia*
- Especie: *B. papyrifera*

*B. papyrifera* es una planta diploide ( $2n=26$ ) [Oginuma y Tobe, 1995] y dioica, es decir individuos diferentes poseen flores masculinas y femeninas [Matthews, 1996; Ainsworth, 2000]. La Figura 2 muestra una flor masculina (a) que corresponde a una espiga de color amarillo verdoso, una flor femenina (b) de forma globular, vellosa, de color azulado, y el fruto (c) de color rojizo anaranjado, compuesto de frutos pequeños y carnosos reunidos en una infrutescencia globosa. Sus hojas son aserradas, cuya longitud varía entre 8 y 20 cm de longitud. Su forma es altamente variable, incluso dentro de un mismo individuo, pues se observan desde hojas ovadas (sin lóbulos) hasta lobuladas con entre 3 y 5 lóbulos, los que presentan profundidad variable [Seelenfreund y cols., 2011].



Figura 2: Imágenes de flores y fruto de *Broussonetia papyrifera*. a) Flor masculina, b) Flor femenina, c) Fruto

### 1.2.2 Importancia económica.

En su región de origen la morera de papel es usada con múltiples fines. Los frutos se usan para el tratamiento de enfermedades oftálmicas e impotencia [Guo y cols., 2013; Lee y cols., 2001], las hojas, las raíces y la corteza se usan para el tratamiento del sangrado vaginal, edema y otras enfermedades [Guo y cols., 2013]. El descubrimiento de numerosos principios activos en *B. papyrifera* respalda su utilización como planta medicinal tradicional en Asia. Entre los metabolitos de interés farmacológico se ha descrito el papyriflavonol A, un flavonoide prenilado, que inhibe la actividad de fosfolipasas A2 IIA y V secretadas (sPLA2s) y la producción de leucotrieno C4 en mastocitos *in vitro* [Kwak y cols., 2003]. *In vivo* inhibe la reacción de anafilaxis pasiva cutánea (PCA) en modelo murino, demostrando su potencial como compuesto antiinflamatorio y antialérgico [Kwak y cols., 2003]. Además este mismo compuesto demostró poseer actividad antimicrobiana contra hongos y bacterias, incluyendo Gram negativos y Gram positivos [Sohn, 2010]. Otros compuestos de interés son el 5,7,3',4'-tetrahidroxi-3-methoxy-8,5'-diprenilflavona y el brousoflavonol B [Matsumoto y cols., 1985], dos prenilflavonas extraídas de la corteza de *B. papyrifera* que inhiben el crecimiento de células ER (*estrogen receptor*) positivas de cáncer de mama *in vitro* e *in vivo*, exhibiendo potencial como compuesto anticancerígeno [Guo y cols., 2013]. Además, brousoflavonol B inhibe la proliferación celular mediante la inducción del arresto de ciclo celular y muerte celular en células ER negativas de cáncer de mamas *in vitro* [Guo, 2013]. Por su usos, en China los individuos masculinos de morera de papel aparentemente son más valorados [Wang y cols., 2012], pero se desconoce si existe una relación entre el sexo del individuo y la producción de metabolitos secundarios utilizados con propiedades terapéuticas.

Por otra parte, la corteza de *B. papyrifera* se ha usado como materia prima en la fabricación de papel desde el año 105 [Zhang y cols., 2012; Ewins, 1987; Clark, 1965]. Además el árbol se usa para combatir la erosión [Abudakar, 2013], pues presenta resistencia a diversos tipos de estrés abiótico [Zhang

y cols., 2012; Li y cols., 2011; Wu y cols., 2009] y se propaga rápidamente de manera vegetativa [Oyanedel, 2010] siendo considerada invasiva en determinadas regiones donde no es endémica [Wu y cols., 2009].

La morera de papel es una de las fuentes más importantes de fibra vegetal en Polinesia, constituyendo la materia prima para la elaboración de “tapa”, un textil fabricado a partir de su corteza y muy importante en la Polinesia antigua por su valor ritual y económico [Seelenfreund y cols., 2010]. En Rapa Nui es conocida como *mahute* y se usa, tal como en la antigüedad y en otras islas de la Polinesia, en la elaboración de capas, faldas, taparrabos y regalos ceremoniales [Seelenfreund y cols., 2010].

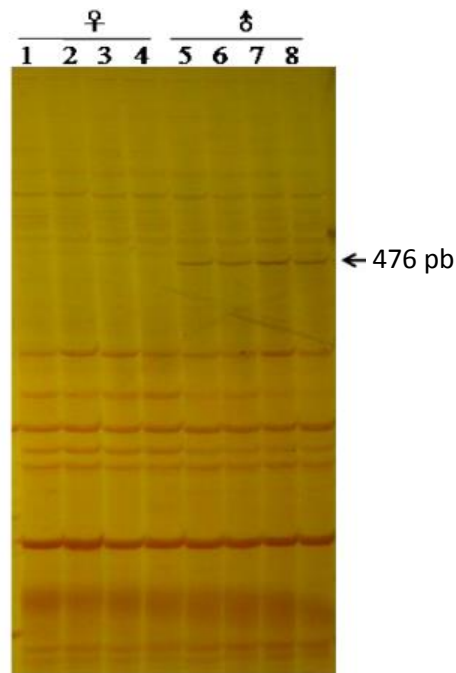
En Polinesia existen principalmente individuos femeninos de morera de papel, cuyo sexo ha sido determinado mediante el método *male marker* de Peñailillo [2014]. Las excepciones corresponden a individuos masculinos de Hawái [Peñailillo, 2014; González; 2014] y un individuo masculino recientemente documentado en la remota isla de Rapa [Payacán, 2015]. Esta distribución de sexo se debe al supuesto traslado de especímenes femeninos en el poblamiento inicial de la Polinesia con la consecuente reproducción vegetativa asexual y una segunda introducción reciente los especímenes de *B. papyrifera* en Hawái [González, 2014, Peñailillo, 2014]. Respecto al origen del individuo masculino de Rapa no hay registro. Además, en la Polinesia los individuos son cortados antes de que lleguen a florecer imposibilitando la identificación de las plantas como individuos femeninos o masculinos, pues no alcanzan el desarrollo floral [Peñailillo, 2014; Oyanedel, 2010].

### 1.2.3 Desarrollo del marcador de sexo.

Resulta entonces útil y necesario identificar el sexo en especímenes de *B. papyrifera* en los cuales no se ha alcanzado el desarrollo floral o en plántulas, para evitar esperar hasta la maduración de las plantas. Por esta razón se desarrolló un marcador molecular de sexo a partir de un análisis de *B. papyrifera* mediante AFLP (*Amplified Fragment Length Polymorphism*) [Wang y cols., 2012]. AFLP es una técnica que permite identificar diversidad genética y desarrollar marcadores moleculares a nivel intraespecífico y que no requiere conocimiento previo del genoma en estudio [Meudt y Clarke, 2007; Schlotterer, 2004], como es el caso de *B. papyrifera*, cuyo genoma aún no se ha secuenciado. Esta técnica consta de 3 etapas y se basa en la amplificación vía PCR de una digestión total de DNA genómico que [Vos, 1995]. Las etapas corresponden a: i) digestión del DNA mediante la acción de dos enzimas de restricción y posterior unión de adaptadores oligonucleotídicos de secuencia conocida, ii) unión de adaptadores y amplificación preselectiva y selectiva mediante PCR de los fragmentos obtenidos luego de la digestión con enzimas de restricción, iii) electroforesis y análisis del patrón de bandas en gel. A partir del perfil de bandeo generado

mediante AFLP es posible aislar productos de amplificación específicos para el desarrollo de marcadores, denominados SCAR. (*Sequence Characterized Amplified Region*). Esta técnica ha sido utilizada con éxito en el desarrollo de marcadores de sexo en plantas cuyos genomas no han sido secuenciados, como *Ficus fulva* [Parrish y cols., 2004] y *Rumex nivalis* [Stehlik y Blattner, 2004].

La técnica de AFLP aplicada al caso de *B. papyrifera* [Wang y cols., 2012], consistió en la digestión de su genoma completo con las enzimas *EcoRI* y *MseI* y posterior amplificación mediante PCR con adaptadores específicos. En la **Figura 3** se observa el patrón de bandeo diferencial obtenido en electroforesis en gel de agarosa. En individuos masculinos se detecta una banda de 476 pb, mientras que en individuos femeninos no hay amplificación detectable [Wang y cols., 2012]). El producto SCAR correspondiente al producto de amplificación de 476 pb fue depositado en la base de datos de secuencias nucleotídicas de GenBank (número acceso GenBank: HQ202152.1, **Figura 4**).



**Figura 3: Experimento de AFLP para la determinación de un marcador sexo-específico de *B. papyrifera*.** Se indica con una flecha la amplificación diferencial del marcador sexo en individuos masculinos (476 pb) respecto de individuos femeninos de *B. papyrifera*

## Broussonetia papyrifera clone G03\_CB100610122\_0197 AFLP marker genomic sequence

GenBank: HQ202152.1

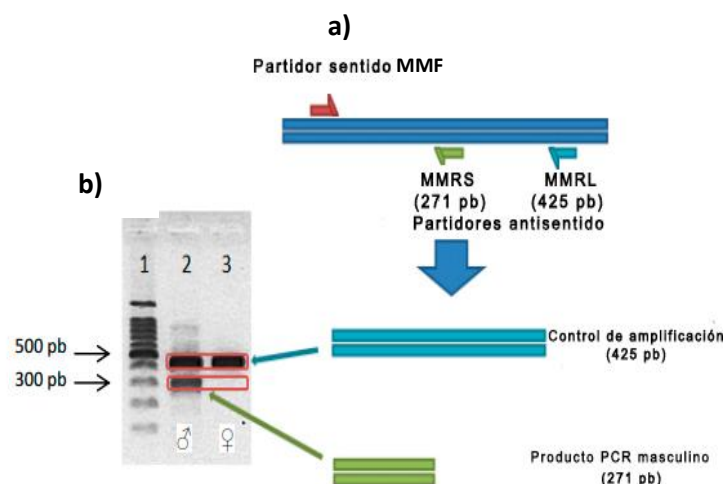
[GenBank](#) [Graphics](#)

>gi|308194304|gb|HQ202152.1| Broussonetia papyrifera clone G03\_CB100610122\_0197 AFLP marker genomic sequence

```
GACTGCGTACCAATTCACCTCAAGCCCTTTGGATCGCGACTTAGAAGCAAATCTCCGAAGCCCCAGTTCTT
CGCCAATAGCTCTTCGCTGAGTGATCTCTGAACTCTTAGGGAGTTACAAGTGTGGACCGATTTGGGCAG
CCGGGAACCTCCGTCATCAAGGCCGTAACCGTGGATCTAAAAAGCCGTTCAATAACGACGACGACGACTT
CTATGGCTACAAGGATTCGACGGCAGCGATATCGACTATGATATCAGTTCCAGTGACGACGACGATGAT
GACGTTGACATGGCTTTTGTATGACGACGACGAGGAGGACGTTGGCAAGAAAGGGAAGAAGAGAAAGTAAA
AGAATTTGAATATCATCATCGTTATTGACCATTGTTTTGGGAAATGTTATTTGGTCTCTGTTTTCCGGA
TTCAAAGTTGGTCTTGCCAGTCTTTGTTCTTGATTACAGTTACTCAGGACTCATC
```

**Figura 4: Secuencia del marcador de sexo de *B. papyrifera*, acceso GenBank: HQ202152.1.** Secuencia del marcador de sexo depositada en GenBank por Wang y cols. (2012).

A partir de la secuencia del marcador SCAR de la **Figura 4**, Peñailillo [2014] diseñó un protocolo de PCR dúplex para la determinación del sexo de individuos de *B. papyrifera*. Se diseñaron tres partidores, denominados MMF, MMRL y MMRS, que generan dos productos de amplificación distintos: la pareja de partidores MMF-MMRL genera un producto de PCR de 425 pb, el cual amplifica en todos los individuos de *B. papyrifera* independientes de su sexo. Por el contrario, la pareja de partidores MMF-MMRS genera un producto de amplificación de 271 pb, el cual solo se obtiene en individuos de sexo masculino. El esquema de la amplificación del PCR dúplex se muestra en la **Figura 5** (adaptado desde [Peñailillo, 2014]).



**Figura 5: Esquema del método de determinación de sexo de *B. papyrifera*.**

a) Esquema de la reacción de PCR-dúplex con partidores y productos PCR generados. b) Gel de agarosa al 2%. Carril 1: Estándar de ADN: Ladder 100 pb, carril 2: BQUCH0137 (♂), carril 3: BQUCH0139 (♀), [Peñailillo, 2014].

La amplificación diferencial observada en el experimento de AFLP indica que el marcador posee sitios polimórficos en su interior [Wang y cols., 2012], que implican cambios puntuales de la secuencia del genoma masculino de la planta respecto del femenino, y que derivan en especificidades de PCR diferenciales en función de la temperatura [Peñailillo, 2014]. Debido a que no se ha descrito homología con secuencias de proteínas conocidas, existe la posibilidad de que el marcador de sexo corresponda a una región regulatoria de la expresión génica, que está clasificada dentro de las llamadas SNC (Secuencias no codificantes).

### 1.3 Secuencias no-codificantes y elementos *cis*-regulatorios

Un desafío central en la genómica funcional y evolutiva ha sido determinar las partes de un genoma que están bajo restricción selectiva. Las regiones codificantes de proteínas son relativamente fáciles de identificar, mientras que otros elementos funcionales, como las regiones de regulación transcripcional y post-transcripcional pueden ser cortas y carentes de motivos evidentes o patrones de secuencias que permitan detectarlos en un solo genoma [Haudry y cols., 2013].

En vertebrados aproximadamente el 5% del DNA corresponde a secuencias codificantes y existen evidencias de que el 95% restante corresponde a secuencias *cis*-regulatorias, que controlan cuánto, cuándo y dónde se va a expresar un gen [Haudry y cols., 2013]. Se cree que estas SNCs (Secuencias No-Codificantes) participan en diversas funciones regulatorias, incluyendo la iniciación de la transcripción y el procesamiento del RNA (por ejemplo, corte y empalme o la localización del mRNA). En eucariontes la regulación de la expresión puede ocurrir a distintos niveles: en la transcripción, procesamiento y estabilización del RNA y en la traducción. La regulación de la transcripción ocurre a través de la unión de proteínas regulatorias, llamadas factores de transcripción (TF, por su sigla en inglés) a las secuencias regulatorias del gen, los cuales se conocen como elementos regulatorios en *cis* o, simplemente, elementos *cis*-regulatorios [Rustici y Gardner, 2006].

En plantas, el desarrollo de un fenotipo característico también está dirigido por programas de expresión génica. El funcionamiento celular y sus mecanismos reguladores en plantas son intermedios en complejidad respecto de los de *S. cerevisiae* y *C. elegans* [Haudry y cols., 2013]. Esto es consistente con la hipótesis de que las plantas obtienen complejidad regulatoria a través de la duplicación de genes o la totalidad del genoma en lugar de un aumento en la complejidad en la regulación vía regiones no codificantes, entendiéndose así la baja relación SNC/gen que poseen las plantas [Haudry y cols., 2013]. Con este antecedente se explica que muchas SNC pasen desapercibidas en enfoques canónicos de búsqueda o identificación [Haudry y cols., 2013].

#### 1.4 Elementos regulatorios de la transcripción.

Los elementos genómicos fundamentales de la regulación de la transcripción o elementos regulatorios en *cis* son:

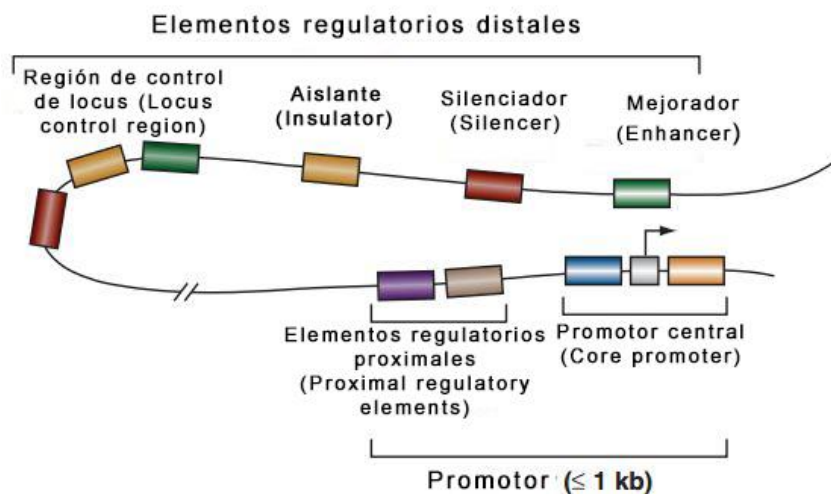
**Promotor central (*Core promoter*):** Es la región inmediatamente río arriba del inicio de transcripción de un gen, que sirve como lugar de anclaje para la maquinaria basal de la transcripción como el complejo de preiniciación (PIC). En organismos eucarióticos el promotor central habitualmente posee motivos conservados como la caja TATA (*TATA box*), un elemento iniciador (*Initiator*), un elemento promotor río abajo (*Downstream Promoter Element*, DPE), un elemento central río abajo (*Downstream Core Element*, DCE), un elemento de reconocimiento de factor de transcripción IIB (*TFIIB-Recognition Element*, BRE) y un elemento motivo 10 (*Motif Ten Element*, MTE) [Maston y cols., 2006].

**Elementos regulatorios proximales (*Proximal Regulatory Elements*):** Es la región ubicada río arriba del promotor central que tiene una longitud de solo unos cientos de pares de bases. Habitualmente posee islas CpG, que son regiones de DNA ricas en GC [Maston y cols., 2006].

**Elementos regulatorios distales (*Distal Regulatory Elements*):** Son regiones que regulan la expresión de un gen a grandes distancias. Poseen sitios de unión de factores de transcripción, pero ubicados a distancias de hasta 1 Mb [Maston y cols., 2006]. Ejemplos de estos elementos distales son los mejoradores (*Enhancers*) que activan la transcripción [Blackwood y Kadonaga, 1998], silenciadores (*Silencers*) que actúan reprimiendo la transcripción [Ogbourne y Antalis, 1998], aislantes (*Insulators*) que son regiones que actúan evitando que los genes se vean afectados por la transcripción de un gen vecino, y regiones de control de *locus* (*Locus Control Region*) que son regiones regulatorias tejido-específicas que actúan sobre cluster de genes [Maston y cols., 2006].

Las secuencias regulatorias mencionadas son conocidas genéricamente como elementos *cis*-regulatorios [Prats y cols., 1990] y su localización se esquematiza en la **Figura 6**.





**Figura 6: Esquema de la región reguladora de un gen eucarionte típico.**

El promotor, compuesto por una región central (*core promoter*) y elementos proximales (*proximal promoter elements*) posee un largo habitual de menos de 1 kb. Los elementos regulatorios distales (*distal regulatory elements*) que pueden incluir mejoradores (*enhancers*), silenciadores (*silencers*), aislantes (*insulators*) y regiones de control de locus (*locus control region*) pueden estar hasta 1 Mb río arriba e incluso en regiones río debajo del sitio de inicio de la transcripción [Maston y cols., 2006].

### 1.5 Búsqueda de motivos en secuencias de DNA utilizando herramientas bioinformáticas

Un motivo es un patrón común en un grupo de secuencias de ácidos nucleicos que comparten alguna propiedad biológica de interés, por ejemplo, corresponder a un sitio de unión para una proteína regulatoria [Lobos y cols., 2011]. El hallazgo de ellos en secuencias de DNA y la identificación de estos como un posible elemento *cis*-regulatorio de la transcripción es una de las grandes aplicaciones de la bioinformática, pues se constituye en un primera etapa de la investigación de secuencias nuevas con eventual función regulatoria de la expresión génica como es el caso del marcador de sexo de *B. papyrifera*, modelo de estudio de esta memoria.

Una herramienta muy utilizada en la búsqueda de motivos en secuencias de DNA es el programa MEME, el cual se basa en el algoritmo *expectation maximization* (EM) [Tanaka y cols., 2014], que permite la búsqueda de motivos estadísticamente significativos en las secuencias ingresadas. De esta forma MEME puede descubrir motivos o patrones no evidentes, que podrían corresponder a sitios de unión en regiones regulatorias en un determinado grupo de secuencias biológicamente relacionadas. El análisis de estas secuencias mediante MEME, cuya longitud idealmente no debiese superar los 1000 pb [Bailey y cols.,

**2006; Bailey y Elkan., 1994]**, entrega como resultado motivos o secuencias de DNA descritos mediante “logos”. Los *logos* son una representación gráfica del grado de conservación del motivo entre las distintas secuencias alineadas entregando un consenso a partir de la diversidad de secuencias **[Sakai y Aerts, 2014]**. Las secuencias obtenidas como resultado del análisis MEME pueden ser sometidas a un posterior estudio mediante MAST. El programa MAST es una herramienta que permite buscar los motivos identificados por MEME en bases de datos de organismos secuenciados y acota la búsqueda a regiones promotoras o río arriba de genes de estos organismos de un largo fijo de 1200 pb **[Bailey y cols., 2006; Lobos y cols., 2011]**. Este último análisis permite determinar si el o los motivos encontrados por MEME poseen efectivamente la funcionalidad de un elemento *cis*-regulatorio y su rol en la regulación de la expresión génica.

Considerando que el marcador de sexo no posee homología con otras secuencias en GenBank y que su secuencia es diferente entre individuos masculinos y femeninos en nuestro modelo de estudio *B. papyrifera*, se propone evaluar si este marcador es o no un posible elemento regulatorio de la expresión génica que conlleve el desarrollo de los distintos sexos en esta especie.

En base a las consideraciones anteriores se plantean las siguientes interrogantes:

**¿Posee el marcador de sexo de *Broussonetia papyrifera* características de elemento regulatorio de la transcripción?, y ¿Es posible encontrar diferencias en la secuencia del marcador de sexo entre individuos masculinos y femeninos que permitan entender una expresión génica diferencial en esta especie dioica?**

## 2 Hipótesis

“El marcador de sexo de *Broussonetia papyrifera*, desarrollado a partir de la técnica de AFLP, representa una región de DNA funcional que está involucrado en la expresión génica diferencial en individuos de sexo masculino y femenino.”

## 3 Objetivos:

### 3.1 Objetivo general

Caracterizar la secuencia correspondiente al marcador de sexo de *B. papyrifera* en individuos de ambos sexos, identificando características de elementos regulatorios en *cis* implicados en una expresión génica diferencial que podría promover el desarrollo de fenotipos masculinos y femeninos en esta especie dioica.

### 3.2 Objetivos específicos:

3.1.1 Aislar las secuencias correspondientes al marcador de sexo de *B. papyrifera* de individuos femeninos y masculinos mediante PCR y posterior clonamiento de vectores.

3.1.2 Caracterizar las secuencias obtenidas correspondientes al marcador de sexo de *B. papyrifera* mediante análisis bioinformático convencional para identificar elementos regulatorios existentes en bases de datos correspondientes a genomas secuenciados.

3.1.3 Caracterizar las secuencias obtenidas correspondientes al marcador de sexo de *B. papyrifera* mediante análisis bioinformático no convencional para identificar elementos regulatorios putativos.

## 4 Materiales y Métodos

### 4.1 Materiales

#### 4.1.1 Reactivos e insumos generales

**Axygen Inc., Corning Life Sciences** (California, EE.UU.): Sistema comercial AxyPrep™ Plasmid Miniprep.

**Biotium Inc.** (California, EE.UU.): GelRed™ Nucleic Acid Gel Stain 10.000X.

**Fermelo Biotec** (Santiago, Chile): Agarosa Lafken p.a.

**Invitrogen, Life Technologies** (California, EE.UU.): TOPO TA cloning® kit, Ultra Pure™ IPTG.

**Promega®** (Wisconsin, EE.UU.): DNA polimerasa GoTaq, MgCl<sub>2</sub>25mM, tampón Green GoTaq Flexi 5X, dNTPs (dATP, dCTP, dTTP y dGTP), X-Gal.

**Sigma** (Missouri, EE.UU.): Ampicilina (sal sódica)

**Tampón TBE 5X:** Tris 445 mM, ácido bórico 445 mM, EDTA 12,5 mM, ajustado a pH 8,4.

**Thermo Scientific** (Pensilvania, EE.UU.): Tampón de carga 6X, Estándar de peso molecular de ADN 1 Kb Plus.

#### 4.1.2 Material biológico

##### 4.1.2.1 Muestras de *B. papyrifera*

Se utilizaron 10 muestras de DNA genómico de individuos de *B. papyrifera* obtenidas mediante una variante del método de extracción de DNA desarrollado por Lodhi y cols. [1994]. En la **Tabla 1** se indica la clave, el sexo y origen geográfico de cada muestra, que se muestra en el mapa de la **Figura 7**.



y las muestras de Vietnam (BQUCH0137, BQUCH0138, BQUCH0139 y BQUCH0140) fueron enviadas por el Dr. Peter Matthews (National Museum of Ethnology, Osaka, Japan). Las muestras fueron colectadas en terreno y almacenadas en presencia de sílica gel a temperatura ambiente.

#### 4.1.2.2 Material microbiológico

Se utilizaron células *E. coli* DH5 $\alpha$  electrocompetentes y plasmidios pCR<sup>®</sup> 2.1-TOPO<sup>®</sup> TA (Life Technologies) provistos en el sistema comercial TOPO<sup>®</sup> TA cloning kit.

#### 4.1.3 Equipos de laboratorio

**Bio-Rad** (California, EE.UU.): Equipo de electrotransformación: *Capacitance extender, Pulse Controller* y *Gene Pulser*<sup>™</sup>.

**Lab-Line instruments** (Kerala, India): Agitador termorregulado Environ Shaker 3527-1.

**Wealtec** (Nevada, EE.UU.): Transiluminador MD-20 UV,  $\lambda = 312$  nm.

**Savant Instruments Inc.** (New York, EE.UU.): Concentrador SpeedVac ISS110-230.

**Techne** (New Jersey, EE.UU.): Termociclador TC-512.

#### 4.1.4 Programas, herramientas informáticas y bases de datos

**Invitrogen Corporation:** Vector NTI Advance<sup>™</sup> 10, AlignX

**CLC Bio, Qiagen<sup>®</sup>:** CLC Sequence Viewer 7

**NCBI** (National Center for Biotechnology Information, U.S. Library of Medicine): VecScreen: *Screen a Sequence for Vector Contamination*.

**NCBI** (National Center for Biotechnology Information, U.S. Library of Medicine): BLAST<sup>®</sup> (*Basic Local Alignment Search Tool*).

**Portuguese Fundação para a Ciência e Tecnologia:** Yeabstract (Yeast Search for Transcriptional Regulators And Consensus Tracking).

**SGD** (Saccharomyces Genome Database)/**Intermine:** Yeastmine.

**Phoenix Bioinformatics Corporation:** TAIR (The Arabidopsis Information Resource).

**National Biomedical Computation Resource:** MEME Suite.

## 4.2 Métodos

La metodología realizada en este trabajo se resume en el diagrama de flujo de la **Figura 8**, donde se muestran las etapas en orden descendente: la primera parte experimental está conformada por la amplificación, clonamiento y secuenciación del marcador de sexo (Objetivo específico 1), y la parte bioinformática consistió en el análisis mediante alineamientos y construcción de dendrogramas (Análisis convencional: Objetivo específico 2) para concluir con la búsqueda de motivos en MEME, la compilación de genes en MAST y la anotación GO de genes de *Arabidopsis thaliana* y *Saccharomyces cerevisiae* usando TAIR y YEASTMINE respectivamente (Análisis no convencional: Objetivo específico 3). Luego para el caso de *Saccharomyces* se buscaron los factores de transcripción que se unen a los motivos encontrados usando Yeabstract.

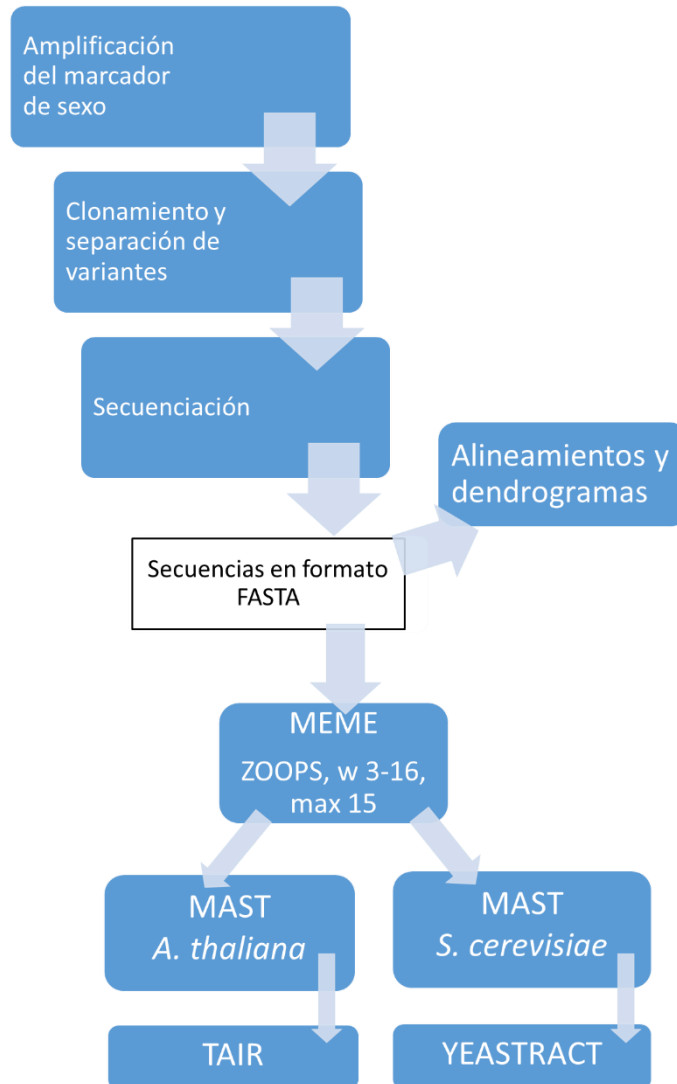
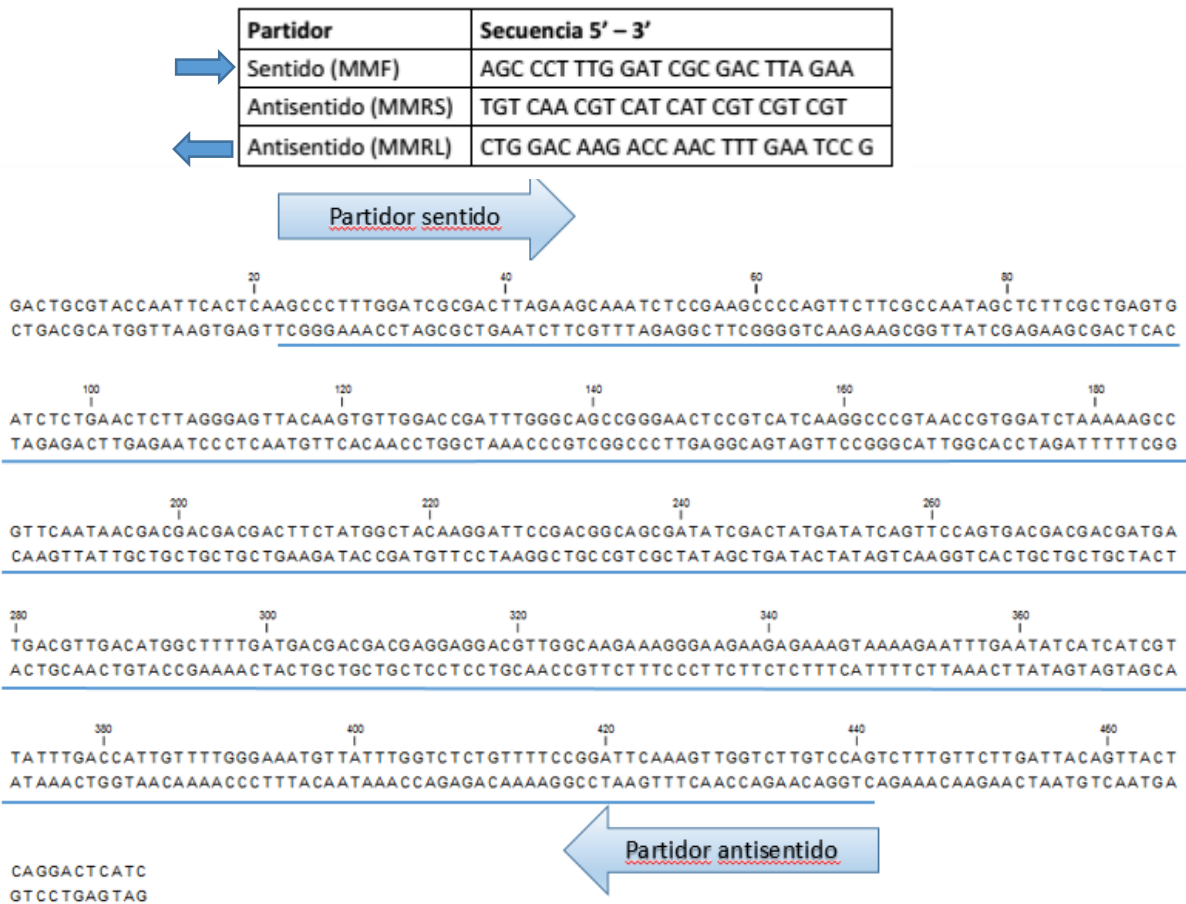


Figura 8: Diagrama de flujo de la metodología empleada en esta memoria de título.

#### 4.2.1 Amplificación del marcador de sexo de *B. papyrifera* mediante PCR

Se amplificó el marcador de sexo de las muestras de *B. papyrifera* utilizando los partidores MMF y MMRL, sentido y antisentido respectivamente, desarrollados por J. Peñailillo en el laboratorio de Biología Molecular de la Facultad de Ciencias Químicas y Farmacéuticas de la Universidad de Chile [Peñailillo, 2014] a partir de la secuencia HQ202152.1. La secuencia de los partidores se muestra en la **Tabla 2**. El esquema de la amplificación de la secuencia HQ202152.1 mediante PCR se muestra en la **Figura 9**, donde se destaca el producto PCR de 425 pares de bases.

**Tabla 2: Partidores usados en la amplificación del marcador de sexo**



**Figura 9: Esquema de la amplificación del marcador de sexo mediante PCR.**

Se muestra la secuencia H1202152.1 con sus dos hebras. Se destaca con una línea azul la región amplificada de la secuencia, que sería de 425 pares de bases usando los partidores sentido MMF y antisentido MMRL respectivamente, diseñados por [Peñailillo, 2014].

Las amplificaciones se realizaron en un volumen de reacción total de 20  $\mu$ L, compuesto por 2  $\mu$ L de solución de DNA genómico 10 ng/ $\mu$ L, 4  $\mu$ L de tampón de PCR 5X, 2  $\mu$ L de  $MgCl_2$  25 mM, 0,5  $\mu$ L de solución de dNTPs 25 mM c/u, 1  $\mu$ L del partidor sentido MMF 5  $\mu$ M (5 pmoles), 1  $\mu$ L del partidor



antisentido, MMRL, 5  $\mu$ M (5 pmoles) y 0,8  $\mu$ L de DNA polimerasa GoTaq Flexi 5U/ $\mu$ L (4U). El programa de amplificación en el termociclador consistió en: Desnaturación inicial a 94°C durante 5 minutos, seguida de 32 ciclos con una etapa de desnaturalización a 94°C durante 1 minuto, una etapa de apareamiento a 53°C durante 1 minuto y una etapa de elongación a 72°C durante 1 minuto, concluyendo con una extensión final a 72°C durante 7 minutos. Las reacciones de PCR se realizaron en tubos de 200  $\mu$ L en termociclador TC-512 e incluyeron un control negativo de amplificación (volumen de agua en reemplazo de solución molde de DNA).

Los productos de PCR se separaron por electroforesis en geles de agarosa al 1,5% en tampón TBE 0,5X. En cada bolsillo del gel se cargó 5  $\mu$ L de producto de PCR y 1  $\mu$ L de Gel Red 100x. En todos los geles de análisis de los productos de PCR se incluyó en un carril 3  $\mu$ L de estándar de peso molecular 1 kb Plus y 1,5  $\mu$ L de Gel Red 100X. La electroforesis se realizó a 90V hasta que el frente de corrida alcanzó el final del gel. Las bandas de los geles se visualizaron en un transiluminador UV ( $\lambda = 312$  nm).

#### **4.2.2 Clonamiento de productos PCR en plasmidio pCR® 2.1-TOPO® TA**

Se clonaron los productos PCR en el plasmidio comercial pCR® 2.1-TOPO® TA según el protocolo del fabricante. Se mezclaron 3  $\mu$ L de producto PCR, 1  $\mu$ L de solución salina (*Salt Solution*), 1  $\mu$ L de agua y 1  $\mu$ L de solución de vector en un volumen final de 6  $\mu$ L. Se mezcló suavemente y se incubó a temperatura ambiente durante 5 minutos, para luego detener la reacción al llevar a hielo. Luego, la mezcla de clonamiento se diluyó con agua a un volumen total de 24  $\mu$ L.

#### **4.2.3 Electrotransformación**

Se tomó 2  $\mu$ L de cada dilución de clonamiento y se adicionaron a un tubo con 50  $\mu$ L de bacterias *E. coli* DH5 $\alpha$  electrocompetentes en glicerol al 10%. Se transfirió la mezcla a una cubeta con separación de electrodos de 0,1 cm. La electrotransformación se realizó con los siguientes parámetros: Gene Pulser a 25  $\mu$ F, Pulse Controller a 200  $\Omega$  con un campo de 12,5 kV/cm. Se removió la cubeta desde la cámara, se recuperaron las células en 1 mL de medio SOC, se resuspendieron con pipeteo suave y luego se transfirieron a tubos de ensayo estériles. Estos cultivos se incubaron en agitador Environ Shaker 3527-1 a 37°C y 200 RPM durante una hora. Luego se sembró 150  $\mu$ L de cultivo en placas agar LB/Ampicilina con 50  $\mu$ L X-gal (Indicador, 20 mg/ml) y 50  $\mu$ L de IPTG (Inductor, 100 mM) en la superficie.

Las placas se incubaron a 37°C durante 24 hr. Se confirmó la presencia del inserto por la formación de colonias blancas (sistema de selección mediante  $\alpha$  complementación) y se hizo un recuento de las colonias recombinantes por placa.

#### **4.2.4 Purificación de plasmidio.**

Se identificaron y picaron las colonias blancas de *E. coli*, se inocularon en 5 mL de medio LB ampicilina 50  $\mu\text{g}/\text{mL}$  líquido y se incubaron en agitador a 37°C y 200 RPM durante 16 hrs. Los plasmidios se purificaron usando el sistema comercial AxyPrep™ Plasmid Miniprep según el protocolo del fabricante.

#### **4.2.5 PCR confirmatorio.**

Se verificó la presencia del inserto en los plasmidios mediante PCR siguiendo el mismo protocolo de la amplificación inicial (volumen de reacción total de 20  $\mu\text{L}$ , compuesto por 2  $\mu\text{L}$  de solución de DNA genómico 10  $\text{ng}/\mu\text{L}$ , 4  $\mu\text{L}$  de tampón de PCR 5X, 2  $\mu\text{L}$  de  $\text{MgCl}_2$  25 mM, 0,5  $\mu\text{L}$  de solución de dNTPs 25 mM c/u, 1  $\mu\text{L}$  del partidor sentido MMF 5  $\mu\text{M}$  (5 pmoles), 1  $\mu\text{L}$  del partidor antisentido, MMRL 5  $\mu\text{M}$  (5 pmoles) y 0,8  $\mu\text{L}$  de DNA polimerasa GoTaq 5U/ $\mu\text{L}$  (4U). El programa de amplificación en el termociclador consistió en: Una desnaturalización inicial a 94°C durante 5 minutos, seguida de 32 ciclos con una etapa de desnaturalización a 94°C durante 1 minuto, una etapa de apareamiento a 53°C durante 1 minuto y una etapa de elongación a 72°C durante 1 minuto, concluyendo con una extensión final a 72°C durante 7 minutos. Las reacciones de PCR se realizaron en tubos de 200  $\mu\text{L}$  en termociclador TC-512 e incluyeron un control negativo de amplificación (volumen de agua en reemplazo de solución molde de DNA).

Los productos PCR se separaron por electroforesis en geles de agarosa al 1,5% en tampón TBE 0,5X. En cada bolsillo del gel se cargó 5  $\mu\text{L}$  de producto de PCR y 1  $\mu\text{L}$  de Gel Red 100x. Las electroforesis de los productos de PCR incluyeron un carril con 3  $\mu\text{L}$  de estándar de peso molecular 1 kb Plus y 1,5  $\mu\text{L}$  de Gel Red 100X. La electroforesis se realizó a 90V hasta que el frente de corrida alcanzó el final del gel. Las bandas de los geles se visualizaron en transiluminador UV ( $\lambda = 312 \text{ nm}$ ).

#### 4.2.6 Secuenciación:

Los plasmidios recombinantes se deshidrataron y se secuenciaron en MacroGen Inc (Seoul, Korea; <http://dna.macrogen.com/eng/>) usando el servicio de secuenciación Standard y partidores universales M13.

#### 4.2.7 Análisis de secuencias de productos PCR clonados

Se analizaron las secuencias obtenidas y se eliminaron los extremos correspondientes a los extremos del plasmidio usando la herramienta *VecScreen: Screen a Sequence for Vector Contamination* (<http://www.ncbi.nlm.nih.gov/tools/vecsreen/>) y el software Vector NTI Advance™ 10. Se editaron nucleótidos cuya señal era de mala calidad de acuerdo a lo observado en el electroferograma.

Las secuencias se analizaron mediante las siguientes estrategias bioinformáticas:

##### 4.2.7.1 Análisis convencional o canónico de la secuencia del marcador de sexo:

- Alineamiento con la herramienta AlignX Vector NTI Advance™ 10, la cual emplea una variante del algoritmo Clustal W.
- Alineamiento y agrupamiento de las secuencias con el software CLC Sequence Viewer 7 (CLC Bio), el cual emplea los algoritmos MUSCLE y Clustal W para el alineamiento y Kimura 80/Neighbor Joining para la construcción de dendrogramas, respectivamente.
- Alineamiento con BLAST® (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) para determinar homología con la secuencia HQ202152.1 y otras secuencias depositadas en GenBank.

##### 4.2.7.2 Búsqueda bioinformática de motivos y genes que poseen los motivos en organismos modelo:

Esta estrategia está basada en el trabajo de **Lobos y cols. [2011]**, que consistió en la búsqueda de motivos putativos con una probable significancia biológica usando el programa MEME y luego, estos motivos se buscaron en promotores de genes identificados con el software MAST. En la presente memoria se utilizaron las bases de datos de *Arabidopsis thaliana* y *Saccharomyces cerevisiae*. Una vez identificados los genes cuyas regiones río arriba poseen los motivos identificados en la secuencia del marcador de sexo usando MEME, se buscó su anotación GO (*Gene Ontology*), con el fin de establecer una aproximación a la

funcionalidad de los genes. Para esto se utilizó la base de datos TAIR (The Arabidopsis Information Resource) para *A. thaliana* y YEASTRACT y YEASTMINE para *S. cerevisiae*.

#### **4.2.7.3 Búsqueda bioinformática de factores de transcripción en *Saccharomyces cerevisiae*.**

Luego de compilados los genes de *S. cerevisiae* que poseen los motivos del marcador de sexo de *B. papyrifera*, se realizó un análisis en la base de datos de *S. cerevisiae*, YEASTRACT, que permite la búsqueda de factores de transcripción (TFs) que se unen a los motivos encontrados y su grado de conservación. Se buscó además el rol biológico de los TFs identificados también mediante YEASTRACT.

## 5 Resultados.

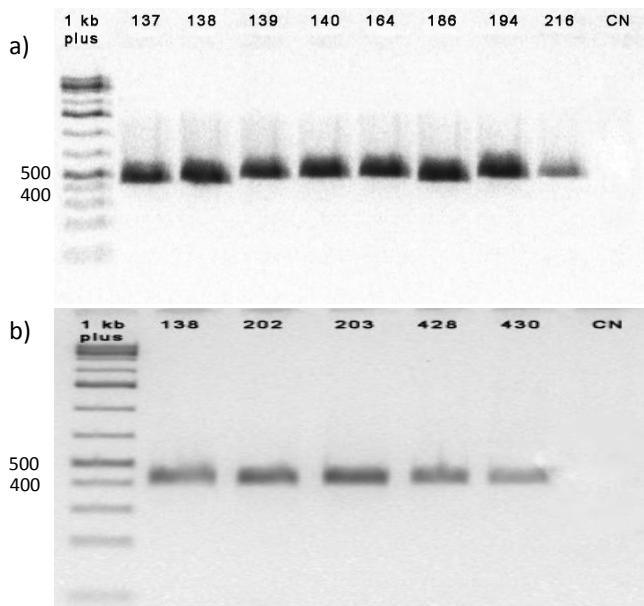
Para un análisis que conduzca a aislar y caracterizar diferentes secuencias del marcador de sexo en individuos distintos de *B. papyrifera* se escogieron 12 individuos según los siguientes criterios de selección:

- Incluir individuos masculinos y femeninos,
- Incluir individuos de distinto origen geográfico: muestras asiáticas y de Oceanía,
- Incluir individuos de distintas características genéticas, definidas previamente mediante microsatélites [Peñailillo, 2014].

En base a estos criterios se seleccionaron las muestras correspondientes a 6 individuos femeninos y 6 individuos masculinos, como se indica en la **Tabla 1 (Ver página 14)**.

### 5.1 Amplificación del marcador de sexo de *B. papyrifera* mediante PCR.

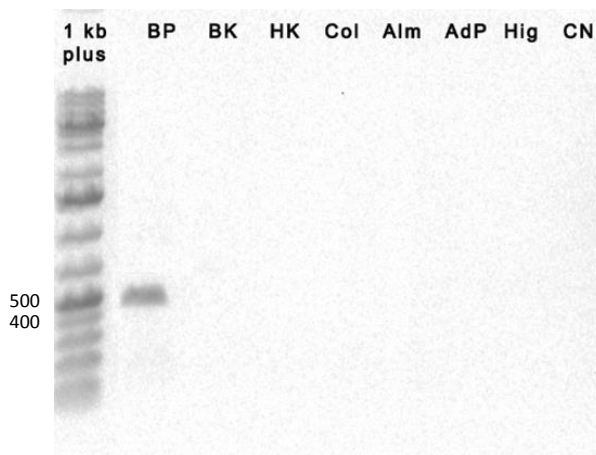
Se amplificó la región del marcador de sexo usando los partidores diseñados por Peñailillo [2014] en individuos masculinos y femeninos de *B. papyrifera*, obteniéndose la banda esperada de aproximadamente 425 pb como se muestra en la **Figura 10**.



**Figura 10: Amplificación del marcador de sexo de *B. papyrifera*.**

**a)** Carril 1 kb plus: Estándar de peso de ADN 1 kb plus, 137: Taiwán ♂, 138: Taiwán ♂, 139: Taiwán, 140: Taiwán ♀, 164: Hawái ♀, 186: Hawái ♂, 194: Hawái ♂, 216: Rapa Nui ♀. CN: Control negativo de amplificación (agua). **b)** Carril 1 kb plus: Estándar de peso de ADN 1 kb plus, 138: Taiwán ♂, 202: Vietnam ♀, 203: Vietnam ♂, 428: China ♀, 430: China ♂, CN: Control negativo de amplificación (agua). Electroforesis en gel de agarosa al 1,5%.

Con el fin de determinar si esta secuencia es particular de *B. papyrifera*, se realizó una PCR en otras especies empleando los mismos partidores usados para amplificar el marcador de sexo. Para ello se escogieron otras especies del mismo género como *Broussonetia kazinoki* y el híbrido Hime-Kozo (*B. papyrifera*-*B. kazinoki*), así como especies de otros géneros o tribus de la familia de las Moraceas como árbol del pan (*Artocarpus altilis*) e higuera (*Ficus carica*). Como controles se incluyeron especies de otras familias como coliflor (*Brassica oleracea*) de las Brassicaceae y almendro (*Prunus dulcis*) de la familia de las Rosaceas. En ninguna de estas especies se observa el producto de amplificación, sugiriendo que la amplificación del marcador de sexo de *B. papyrifera* es específica (**Figura 11**).



**Figura 11: Amplificación del marcador de sexo de *B. papyrifera* en diferentes especies.**

Carril 1 kb plus: Estándar de peso de ADN 1 kb plus, BP: Muestra BQUCH016 de *B. papyrifera* (Rapa Nui ♀), BK: *Broussonetia kazinoki*, HK: híbrido Hime-Kozo (*B. papyrifera*-*B. kazinoki*), Col: coliflor (*Brassica oleracea*), Alm: almendro (*Prunus dulcis*), AdP: árbol del pan (*Artocarpus altilis*), Hig: higuera (*Ficus carica*). CN: control negativo de amplificación (agua). Electroforesis en gel de agarosa al 1,5%.

## 5.2 Clonamiento de productos PCR de la región del marcador de sexo.

Previamente Peñailillo [2014] secuenció el producto de amplificación de individuos femeninos y masculinos. En los primeros se obtuvieron señales definidas en el electroferograma, mientras que en los segundos se obtuvieron señales superpuestas, y por tanto secuencias ilegibles. Lo anterior se interpreta en primera instancia como que los individuos femeninos poseen una sola versión de la región amplificada, mientras que los masculinos poseen dos o más secuencias, las cuales provocarían la obtención de una secuencia que no es posible interpretar. Para separar las distintas variantes amplificadas del marcador de sexo presentes en un mismo individuo se clonaron los productos de PCR de ejemplares femeninos y masculinos en el vector comercial pCR® 2.1-TOPO® TA. Una vez ligado el producto de amplificación al vector, se transformaron células *E. coli* DH5α con el plasmidio recombinante y se seleccionaron las

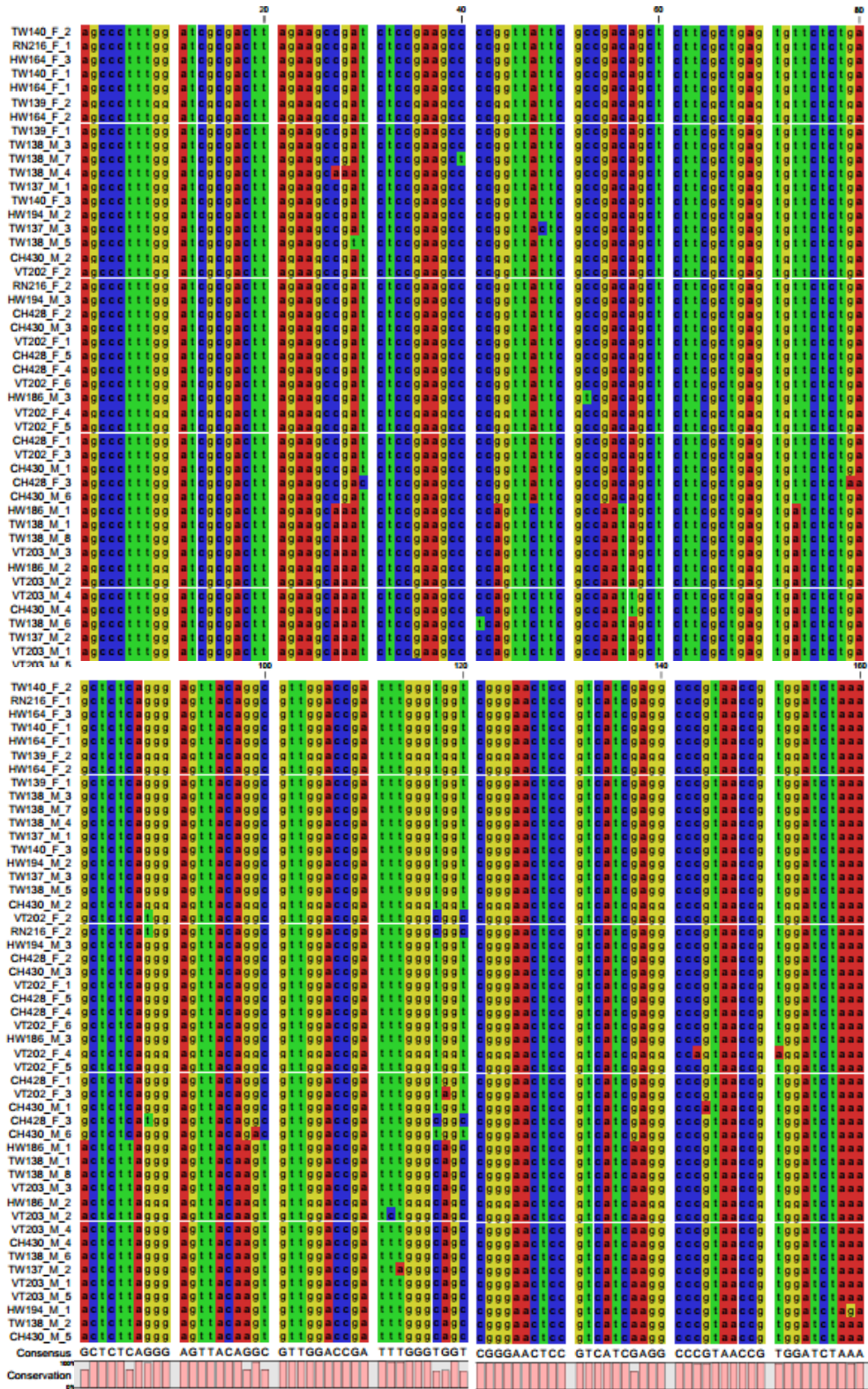
colonias blancas (con inserto) por  $\alpha$  complementación. En la **Tabla 3** se indica el número de colonias recombinantes seleccionadas por muestra. Se tomó el mayor número posible de colonias blancas para maximizar el total de secuencias a analizar.

**Tabla 3: Colonias seleccionadas (clones) de cada ejemplar de *B. papyrifera*.**

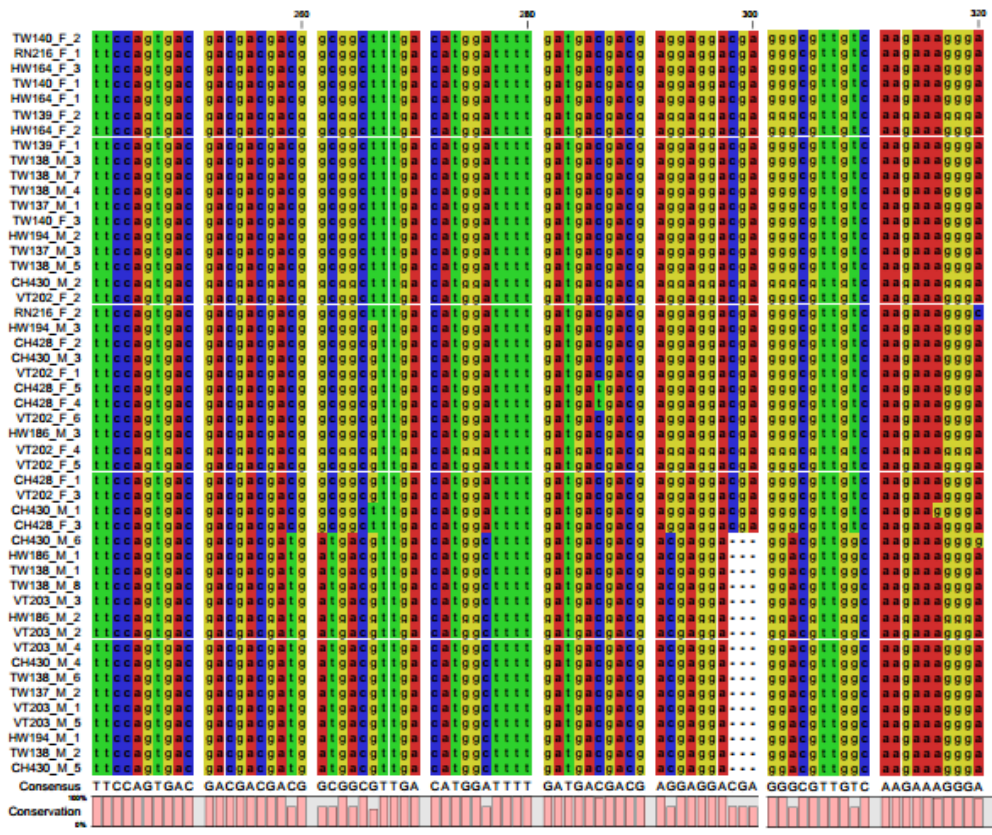
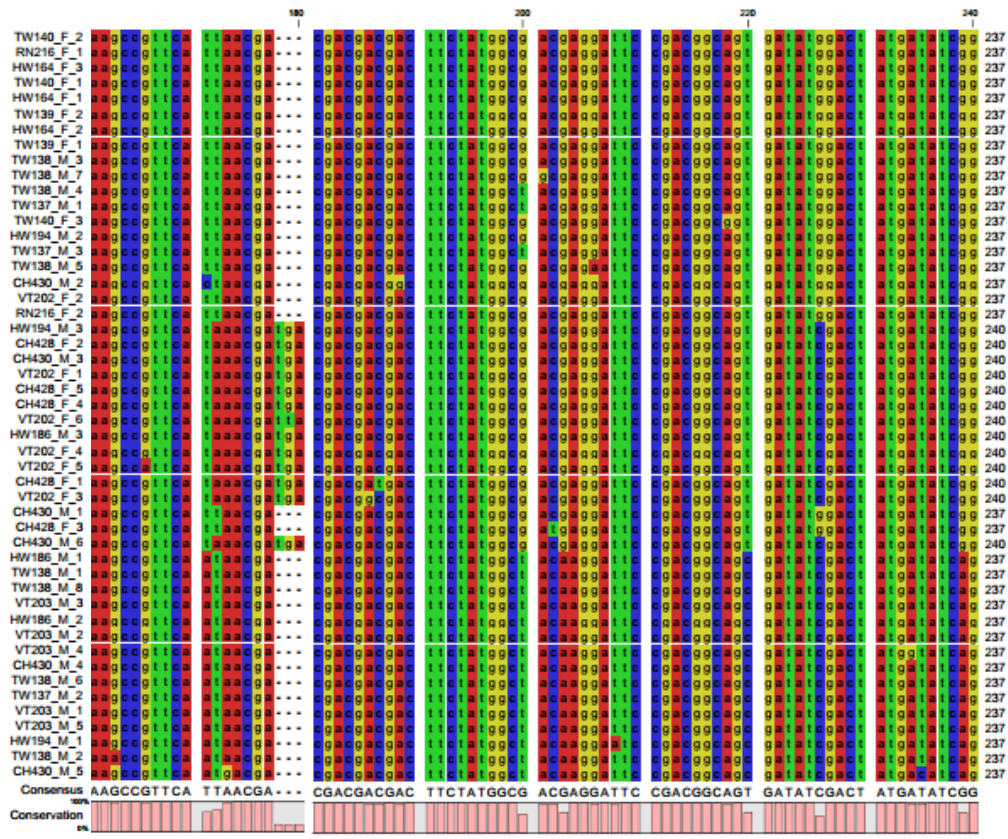
Muestra	Sexo	Localidad	Colonias seleccionadas
BQUCH0137	M	Taiwán	3
BQUCH0138	M	Taiwán	8
BQUCH0139	F	Taiwán	2
BQUCH0140	F	Taiwán	3
BQUCH0202	F	Vietnam	6
BQUCH0203	M	Vietnam	5
BQUCH0428	F	China	5
BQUCH0430	M	China	6
BQUCH0164	F	Hawái	3
BQUCH0186	M	Hawái	3
BQUCH0194	M	Hawái	3
BQUCH0216	F	Rapa Nui	2

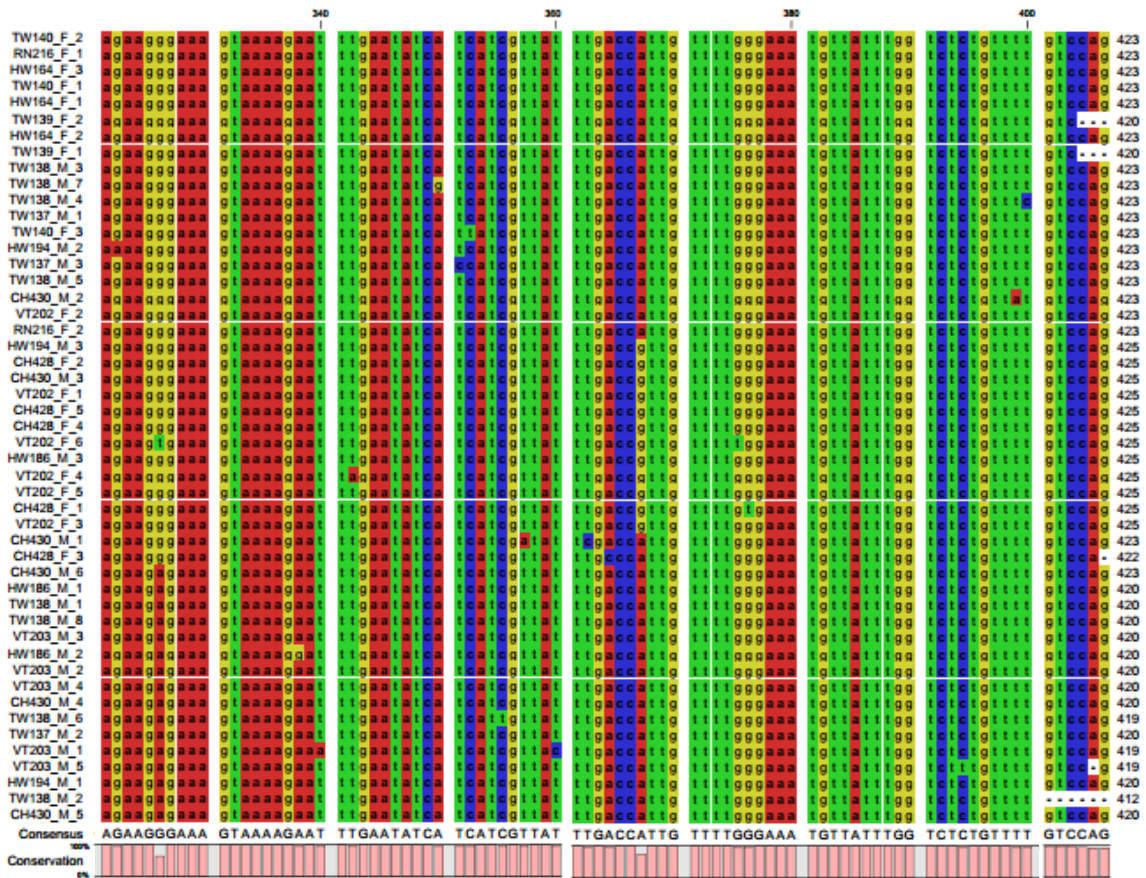
Una vez purificados los plasmidios según lo descrito en la **sección 4.2.4**, confirmada la ligación de los productos PCR, y secuenciados los clones, estos se alinearon usando el software CLC Sequence Viewer 7.5 en el marco del objetivo específico 2. El alineamiento múltiple de todas las secuencias se ilustra en la **Figura 12**, observándose heterogeneidad en cuanto a polimorfismos (SNP e indels) y largo del producto de amplificación clonado, incluso entre las secuencias de una misma muestra. Con el fin de identificar las distintas secuencias de cada muestra se realizaron alineamientos múltiples con cada una de ellas (**Figura 13**). También se realizaron alineamientos de las secuencias masculinas y femeninas, los que se muestran en **Figura 14 a) y b)** respectivamente.

En la **Tabla 4** se indica el número de secuencias distintas encontradas por individuo. Luego se calculó el porcentaje de identidad entre ellas (**Tabla 5**).



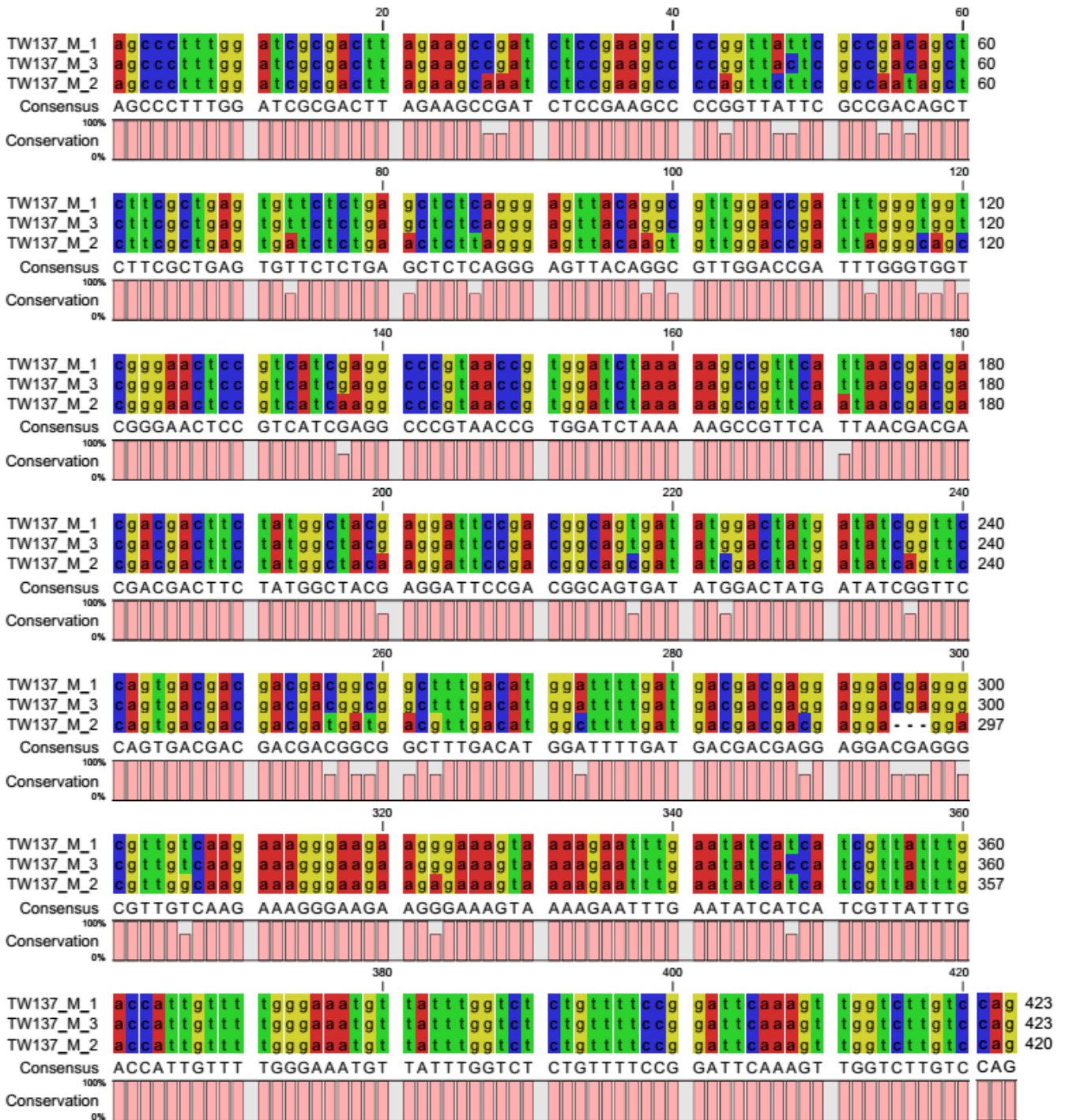




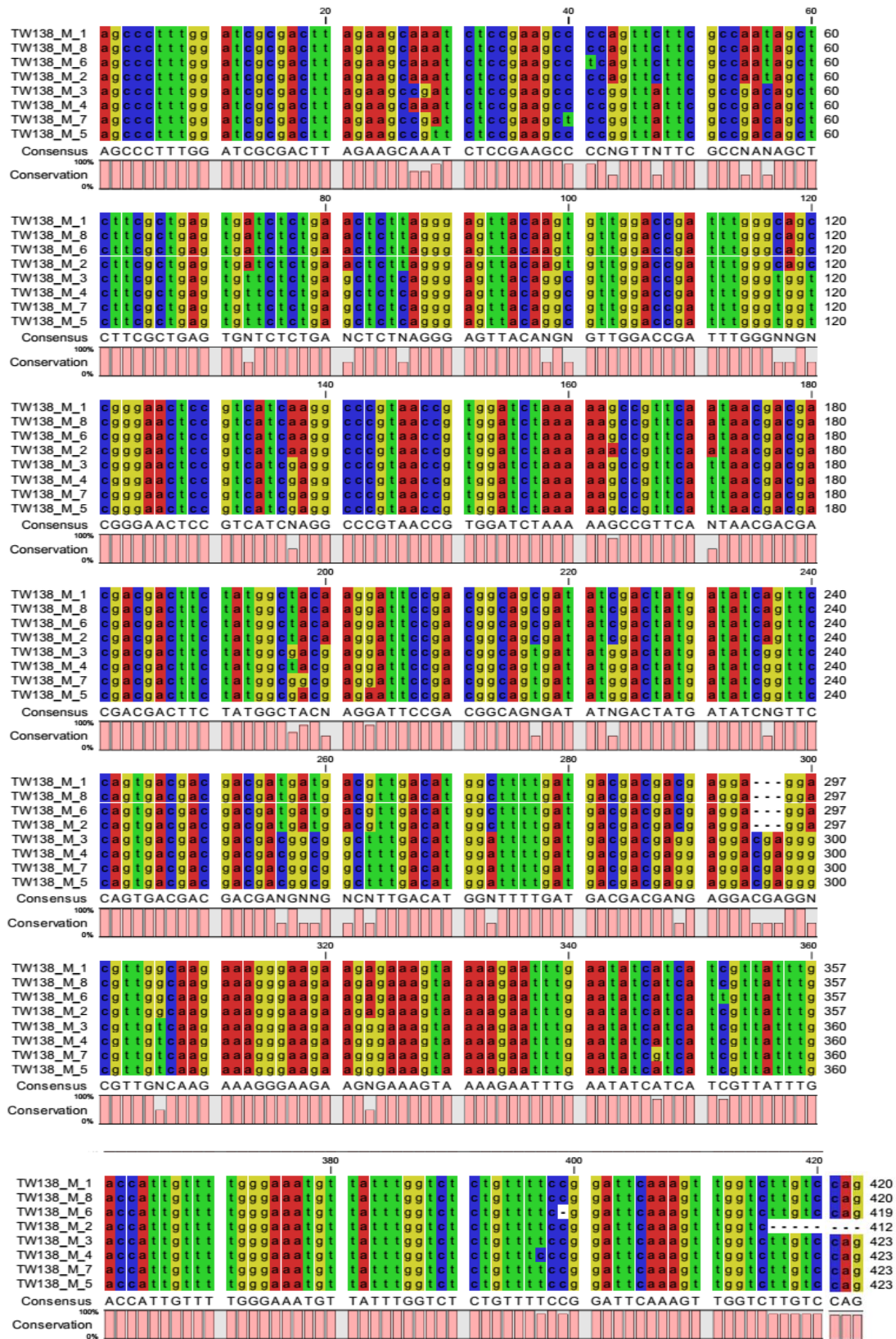


**Figura 12: Alineamiento múltiple de las secuencias del marcador de sexo de *B. papyrifera*.** Se observan las zonas de SNPs, de inserciones/delecciones y la diferencia de longitudes, en los 12 individuos del estudio. **CLC Sequence Viewer 7.5.** El código de nombre alude a las iniciales del lugar de origen de la muestra, el número en el banco genómico, el sexo de la muestra y el número de colonia de donde fue tomada Ej: CH430\_M\_5 es la secuencia del marcador de sexo inserto en el plasmidio de la quinta colonia recombinante originada de la muestra 430 masculina, proveniente de China.

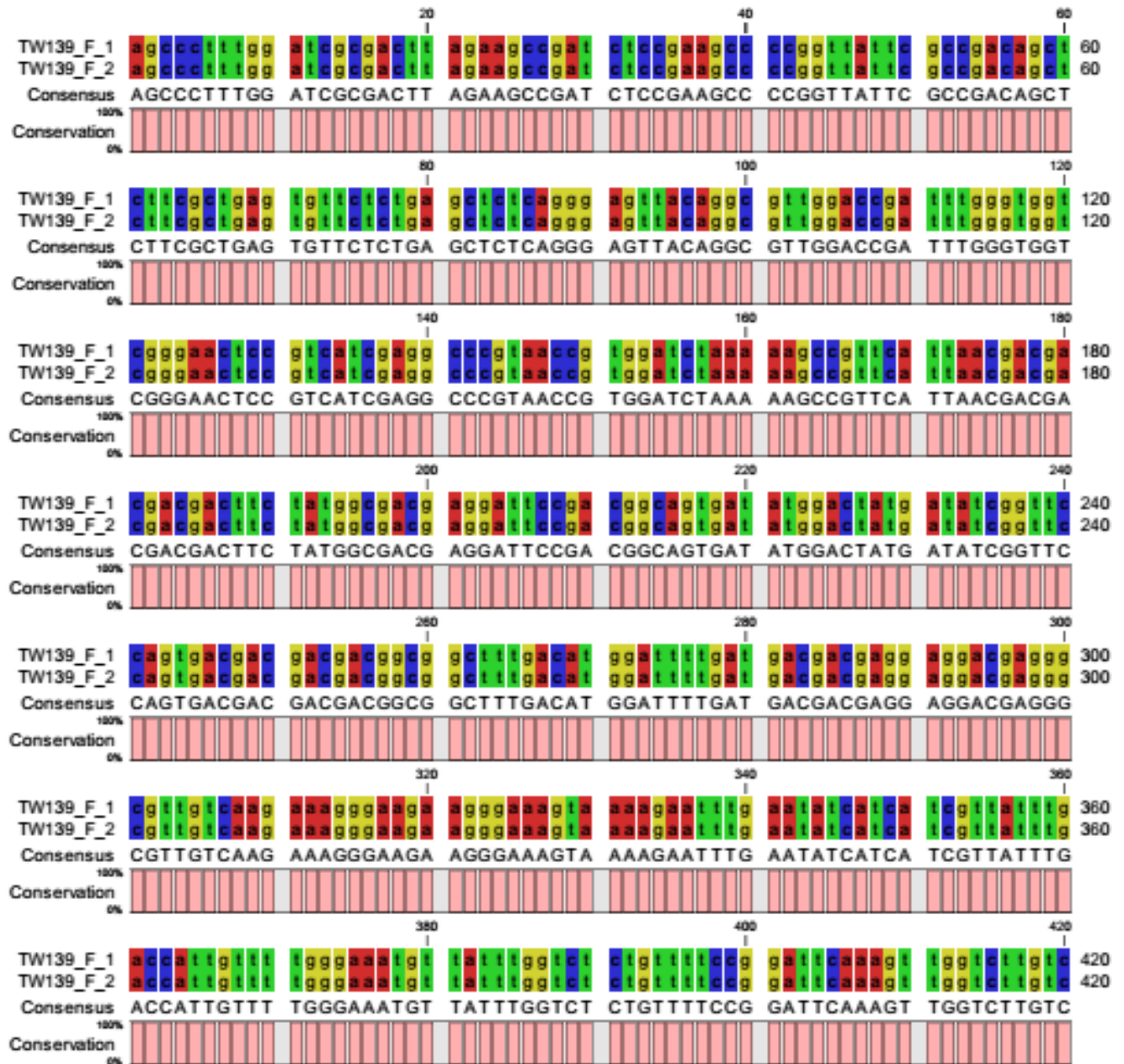
a)



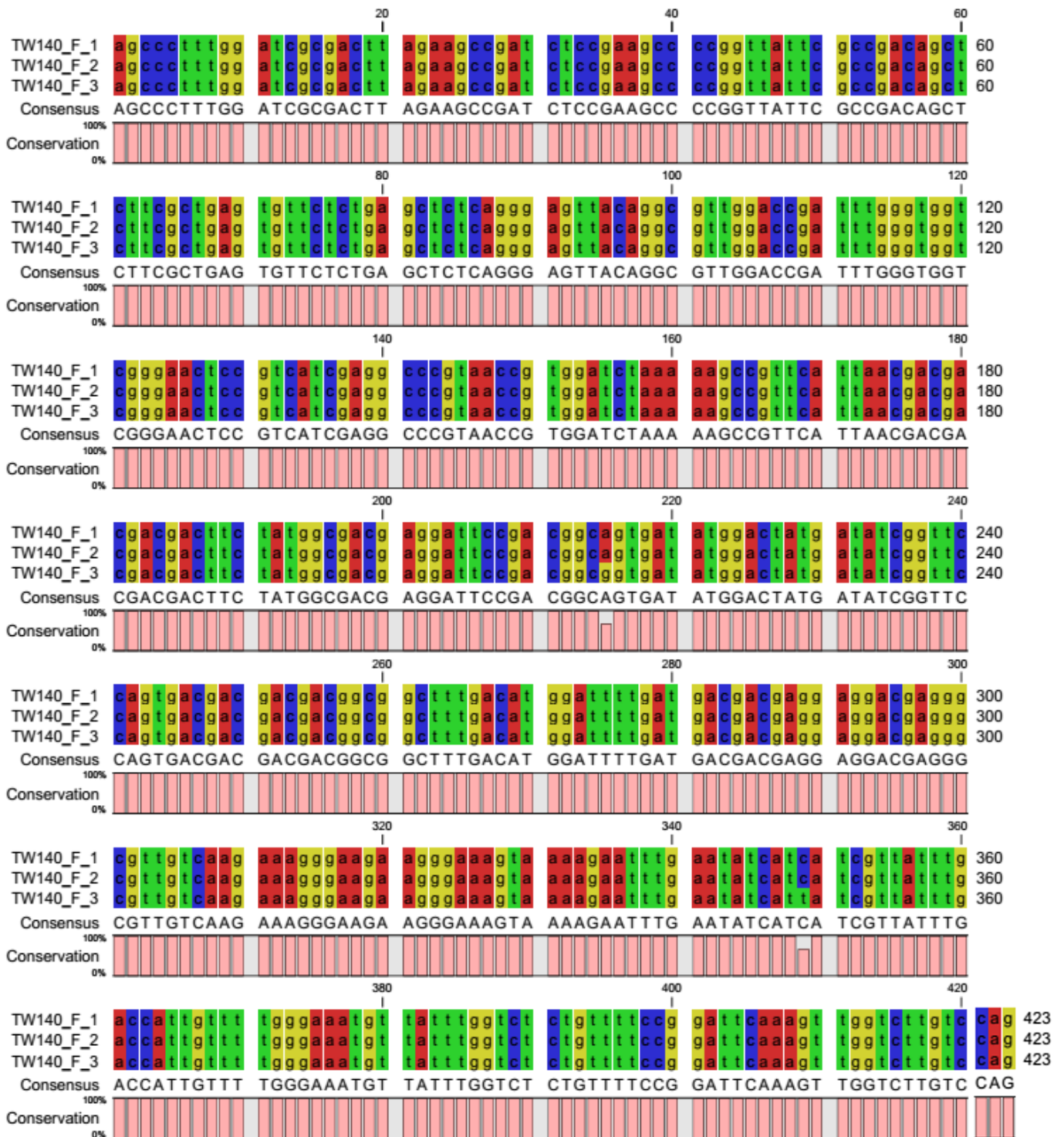
b)



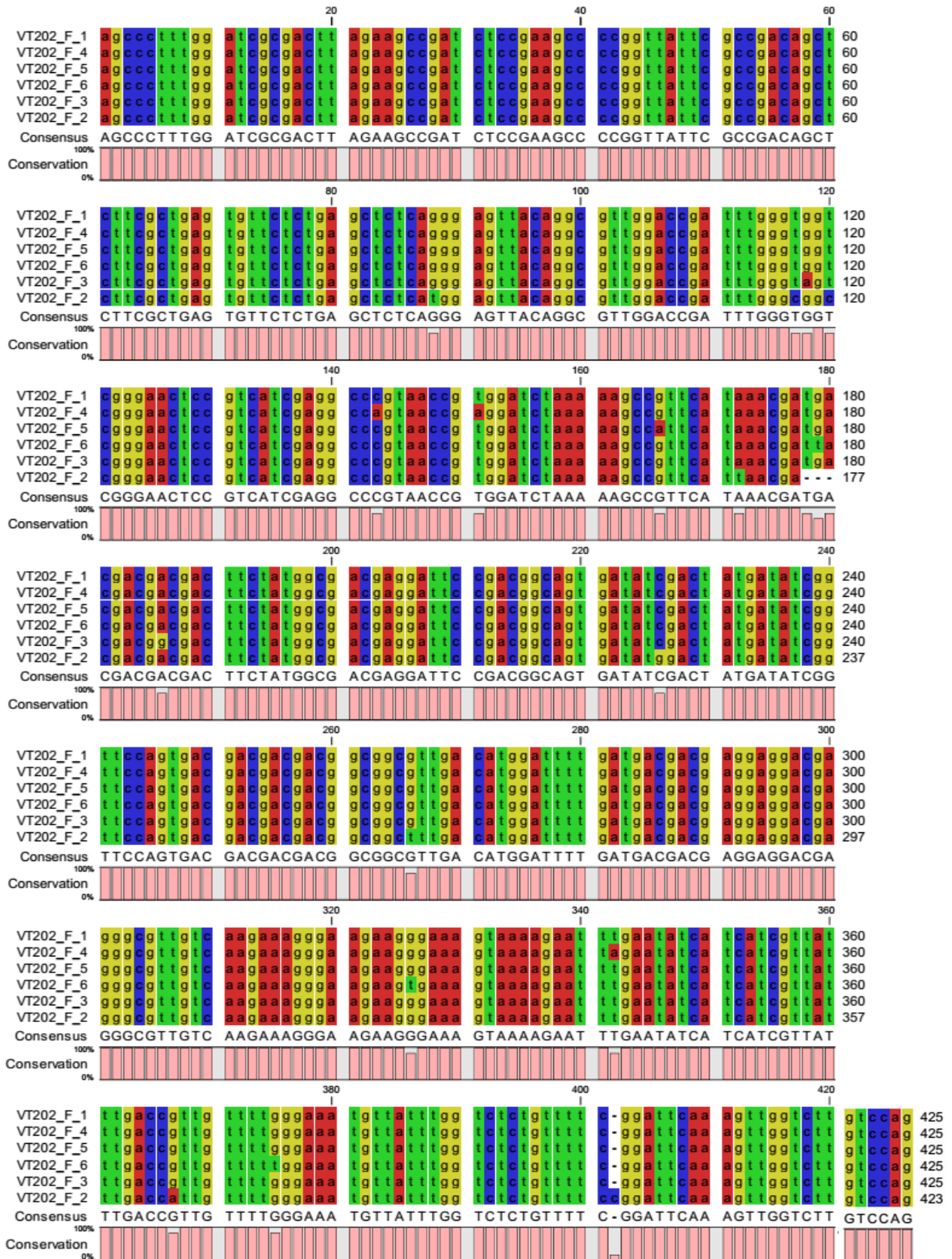
c)



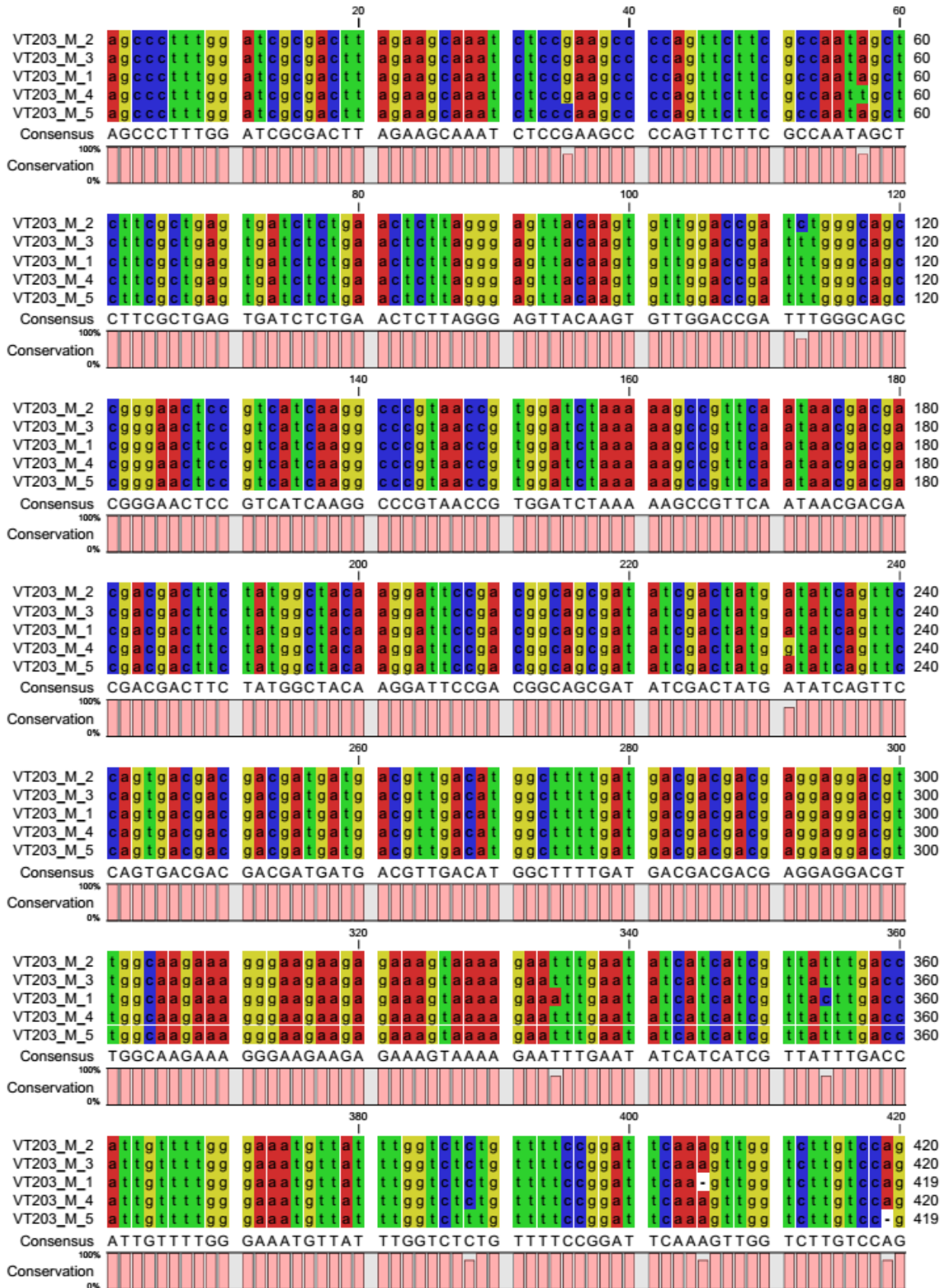
d)



e)

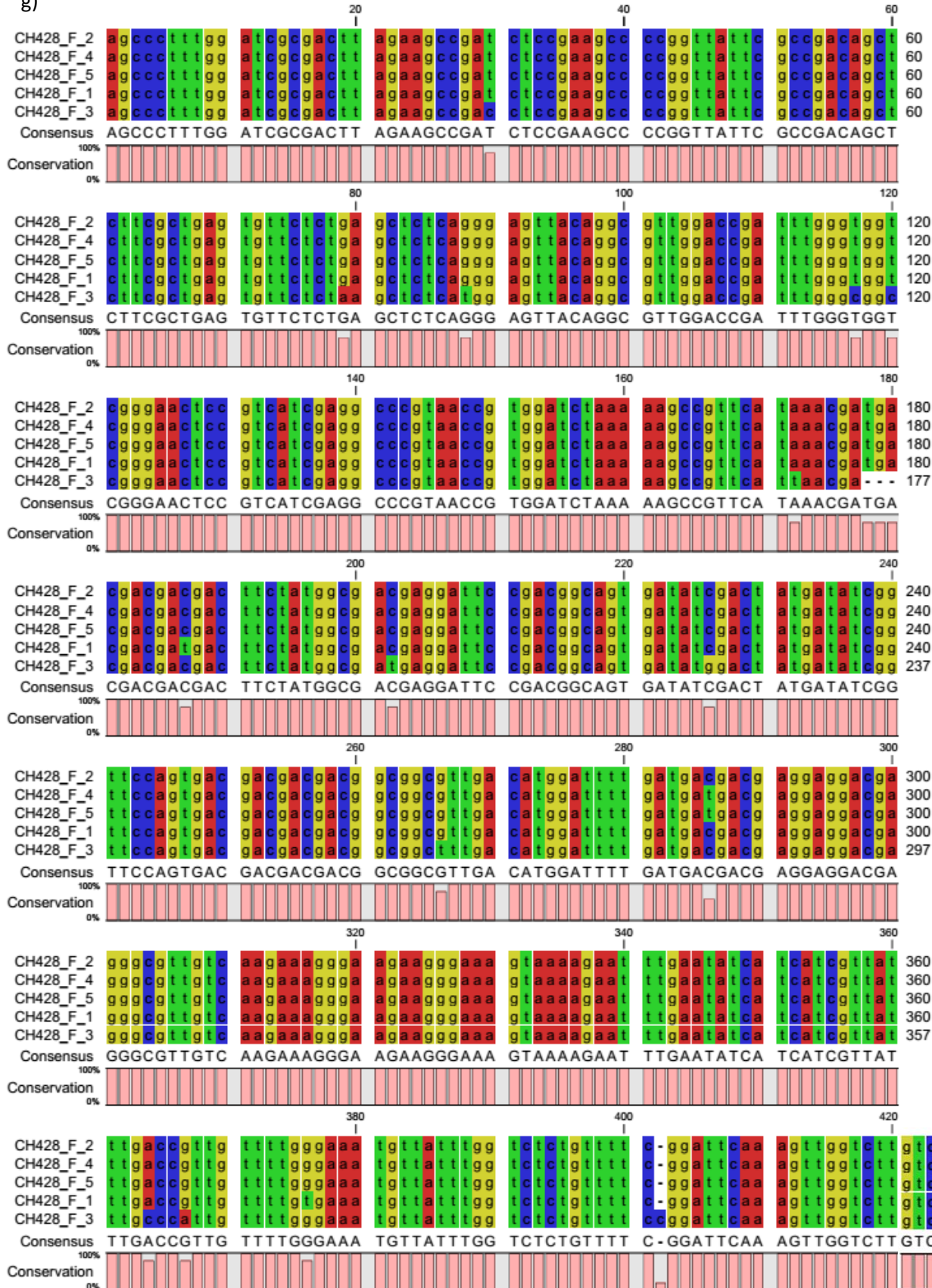


f)

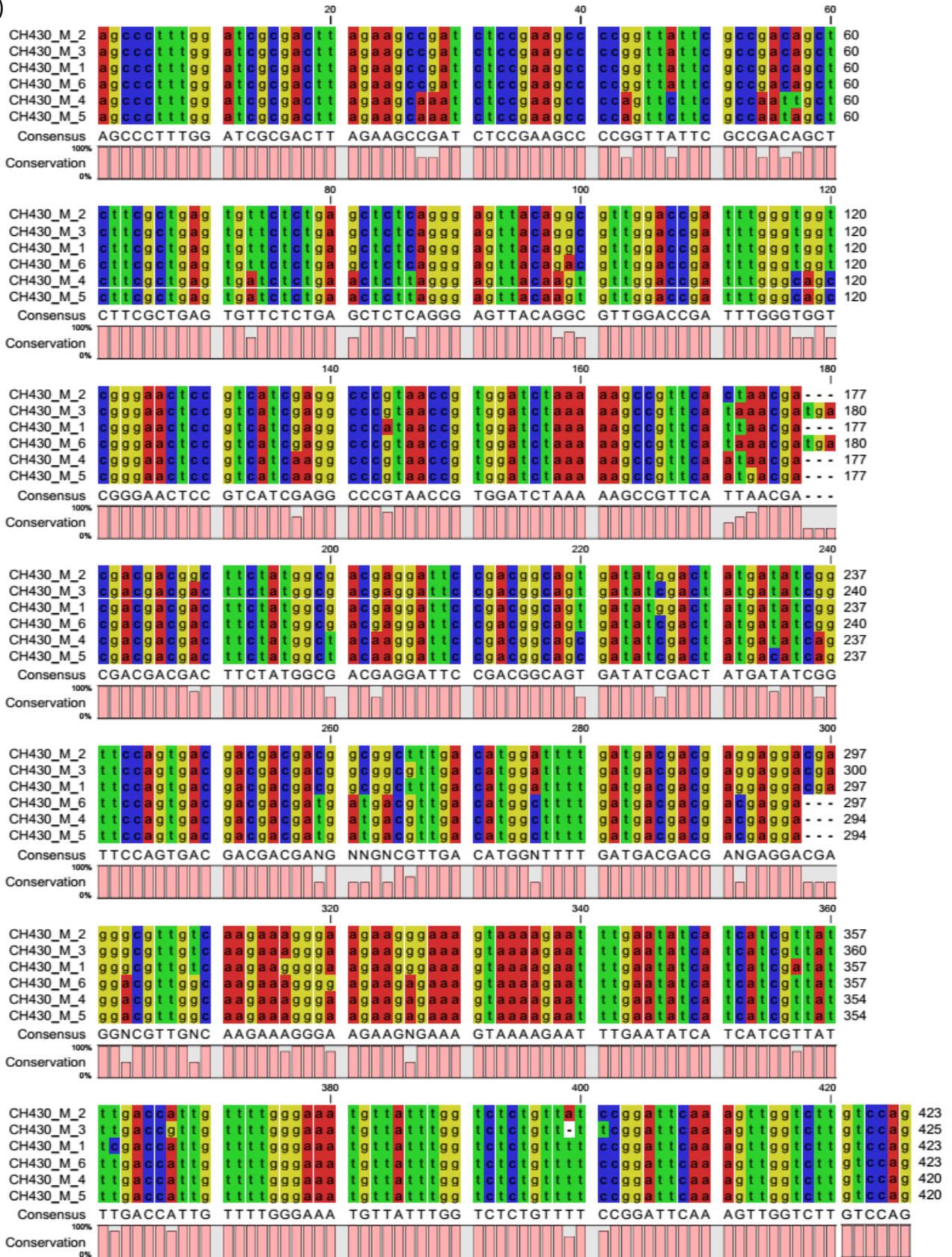




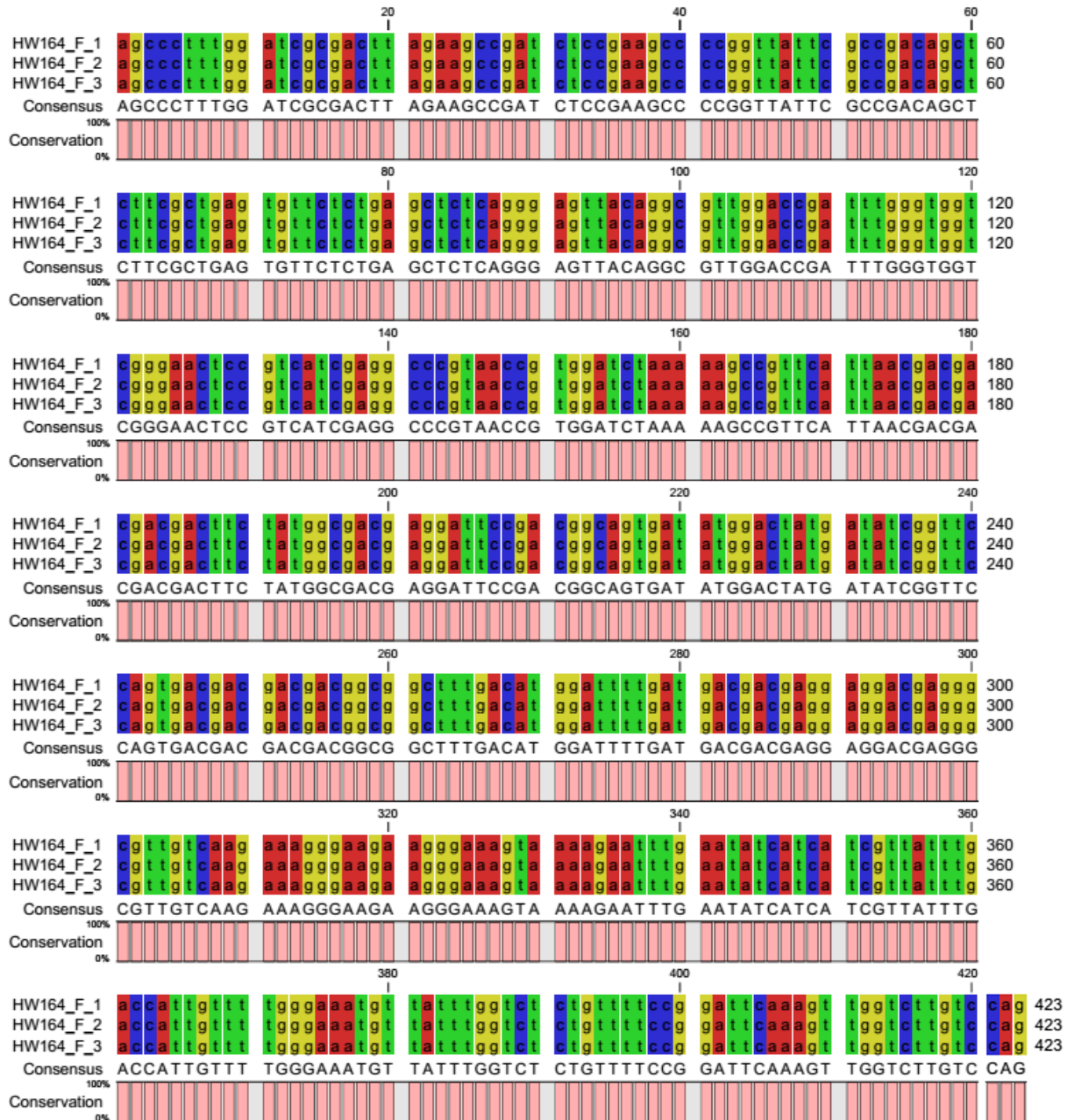
g)



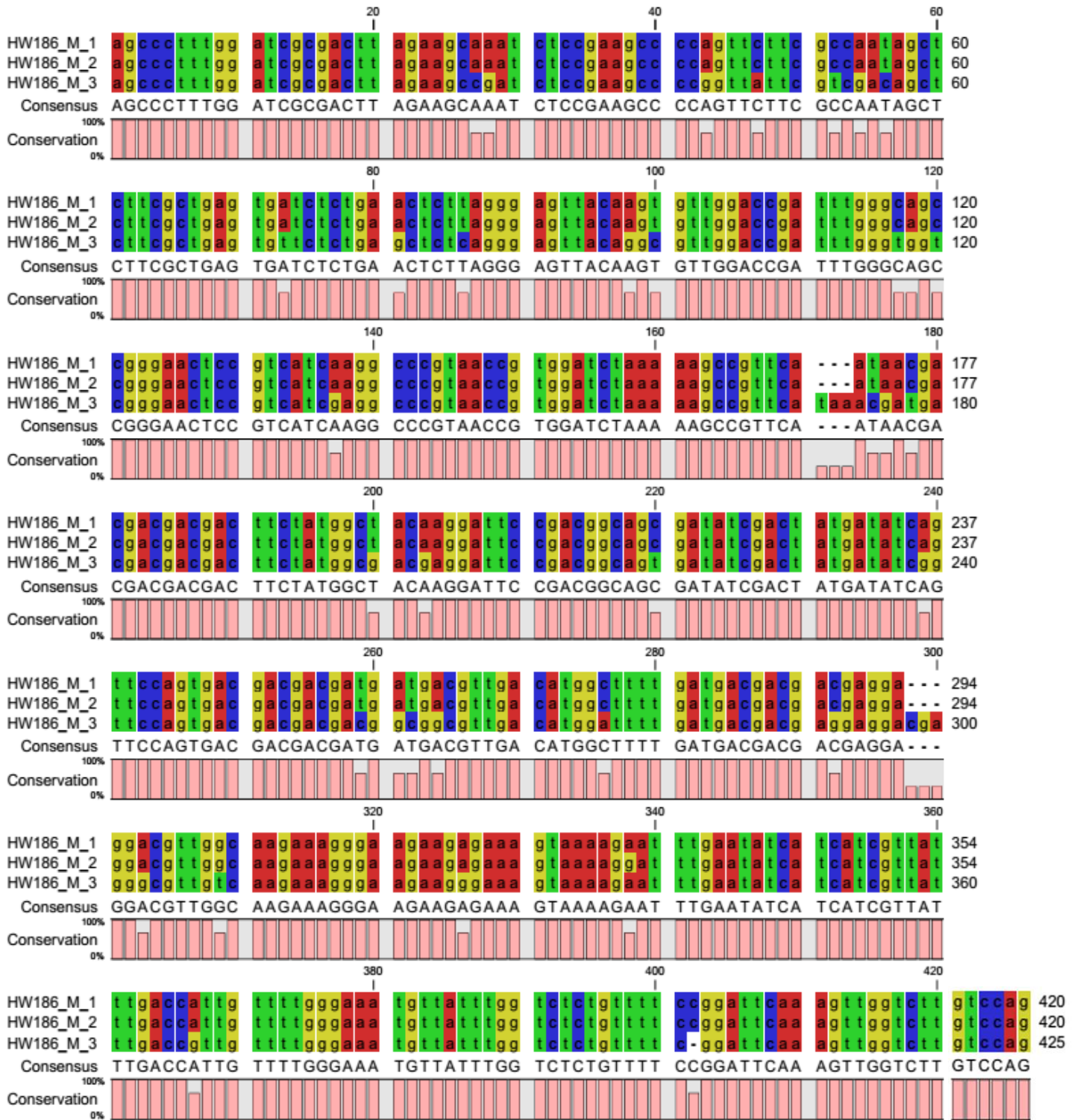
h)



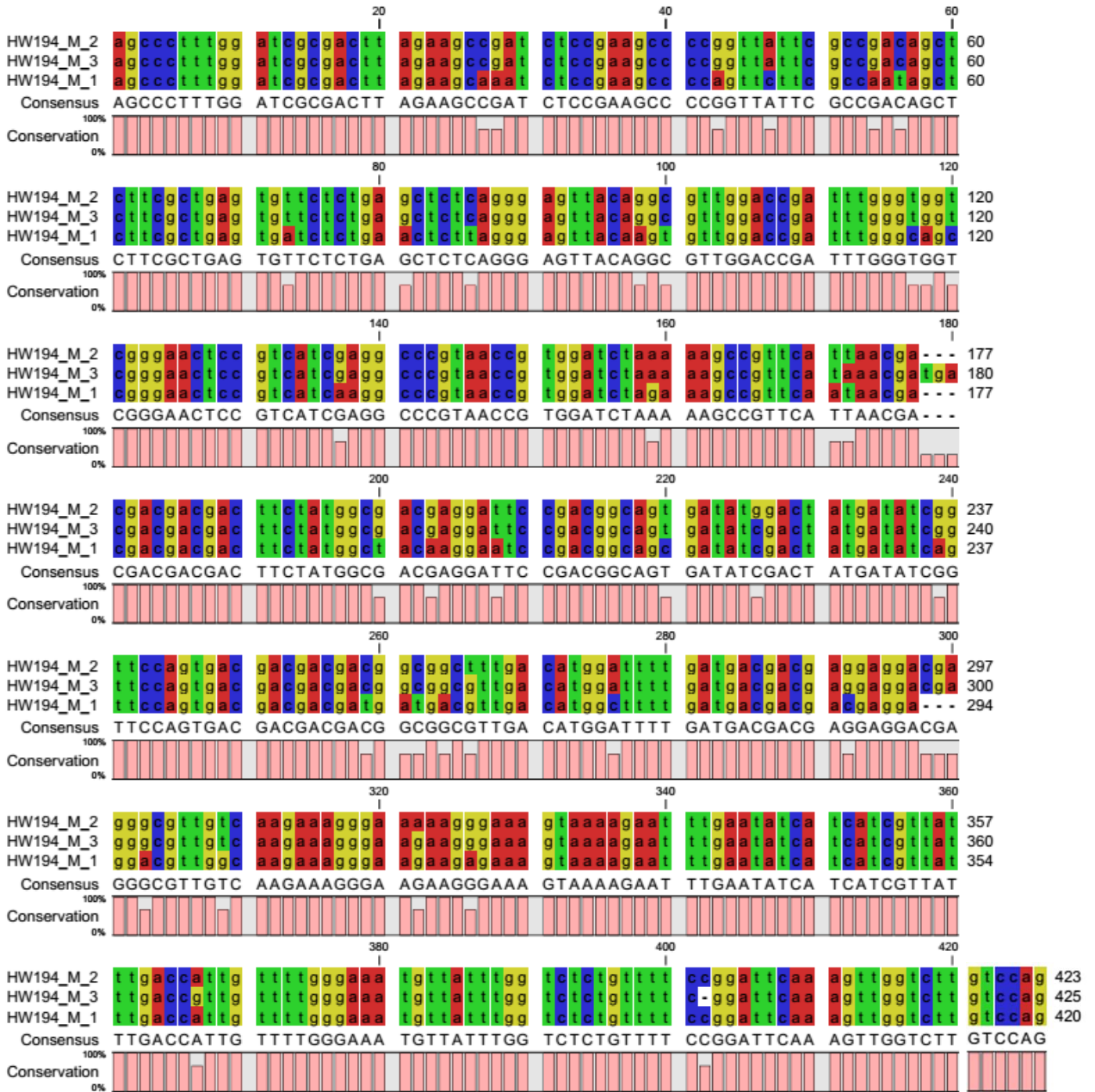
i)



j)



k)



l)

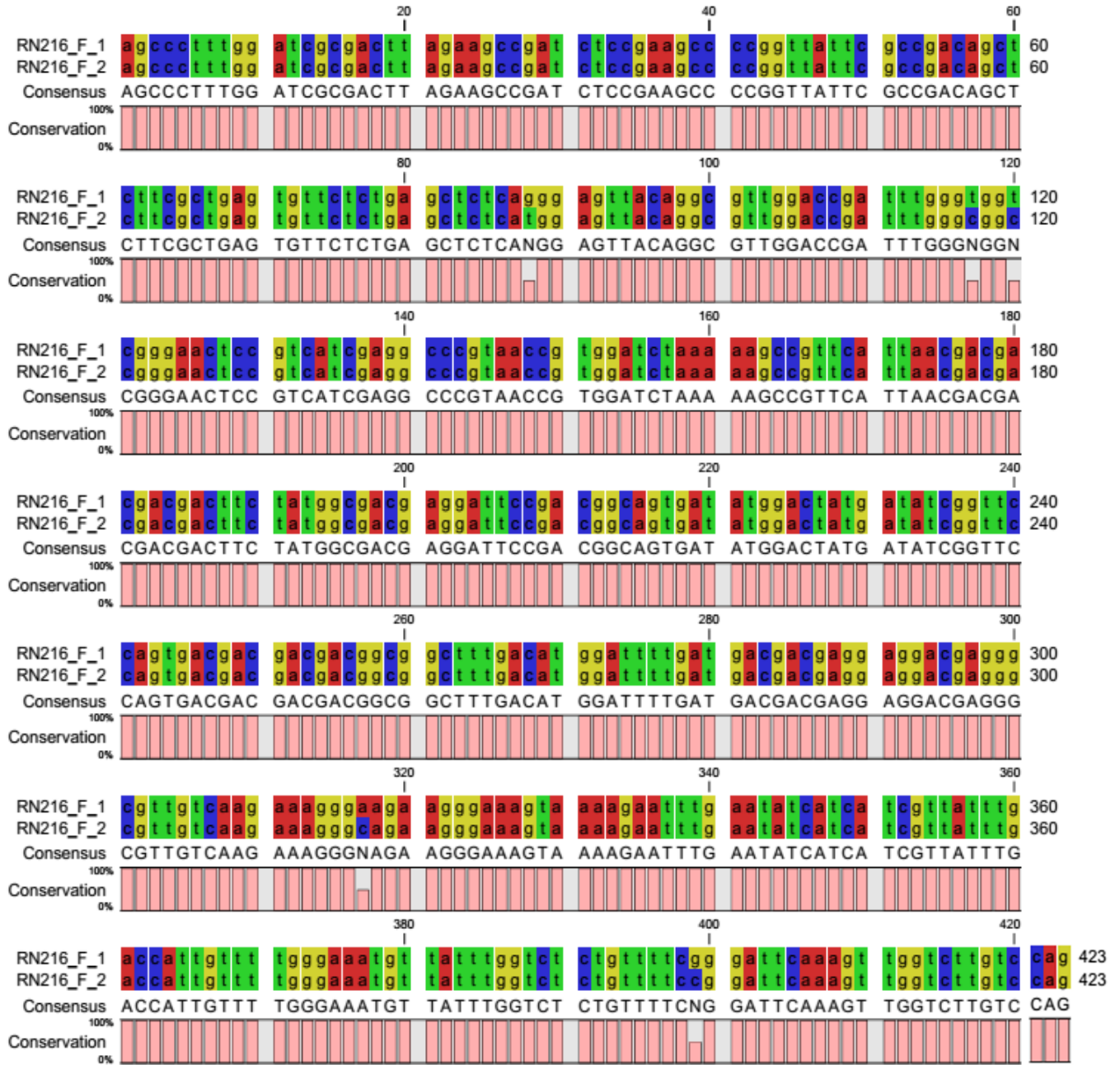


Figura 13: Alineamiento múltiple de secuencias del marcador de sexo de *B. papyrifera* por individuo.

a) Muestra 137, b) Muestra 138, c) Muestra 139, d) Muestra 140, e) Muestra 202, f) Muestra 203, g) Muestra 428, h) Muestra 430, i) Muestra 164, j) Muestra 186, k) Muestra 194, l) Muestra 216. **Software CLC Sequence Viewer.7.5.** El código de nombres está explicado en la **Figura 12.**

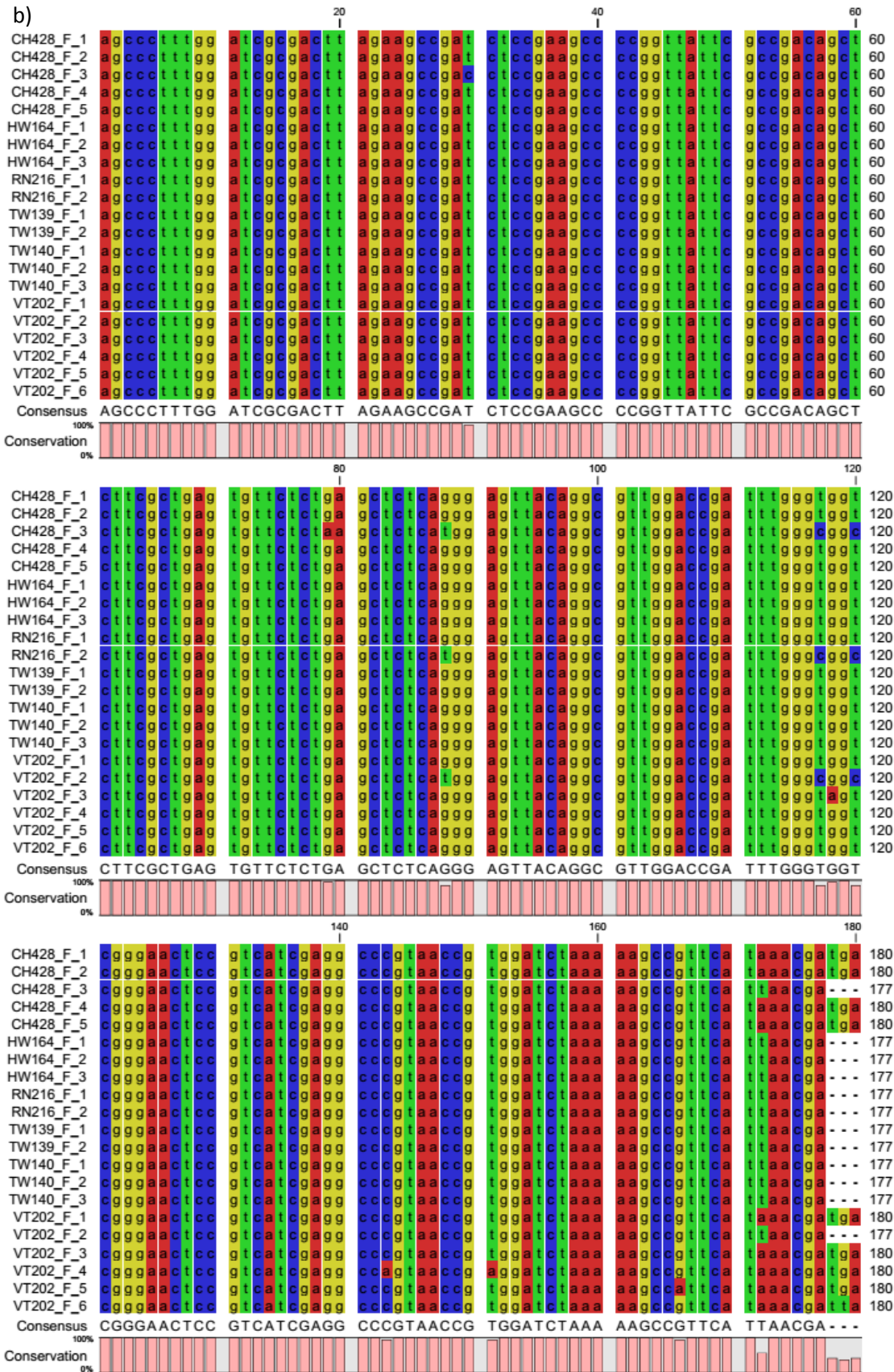


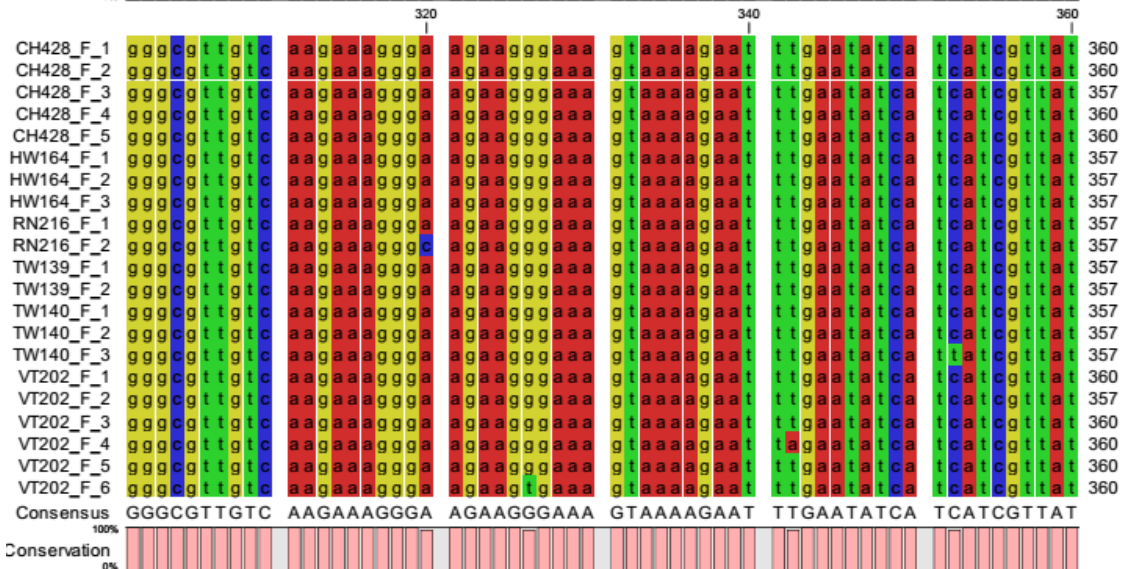
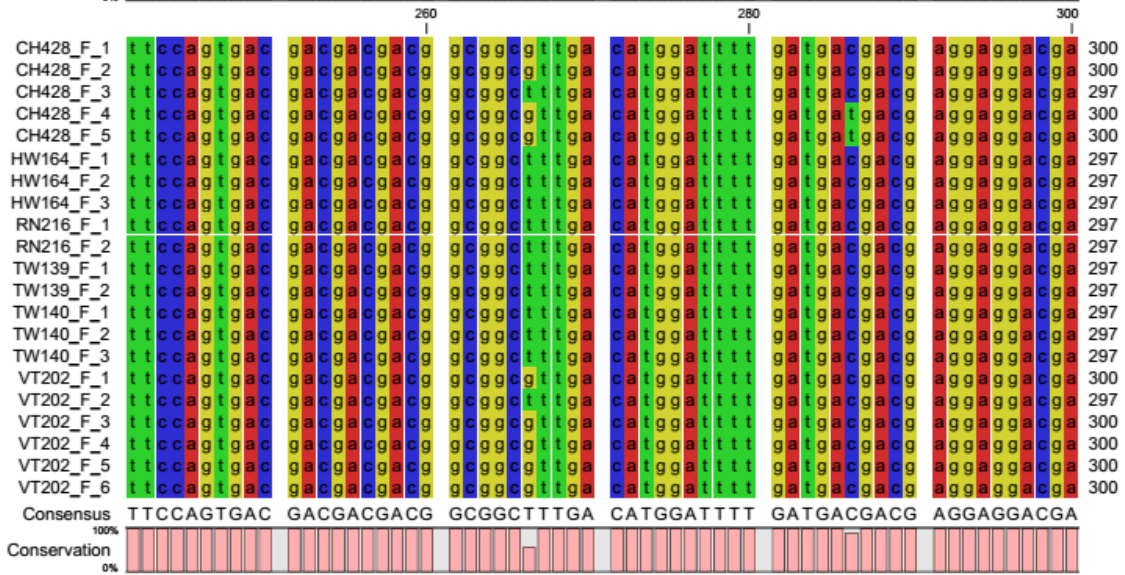
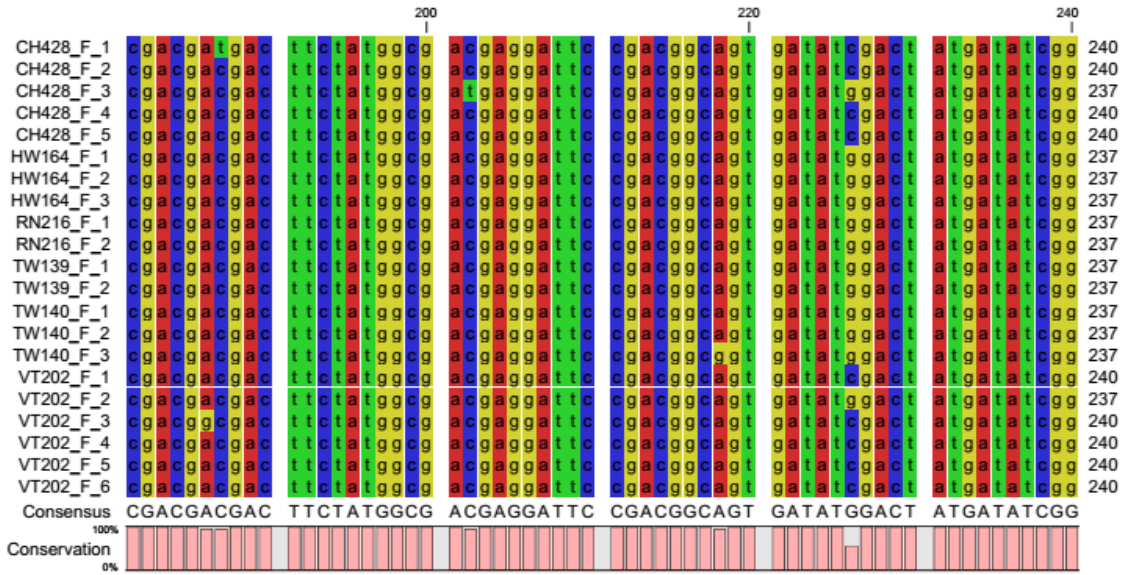












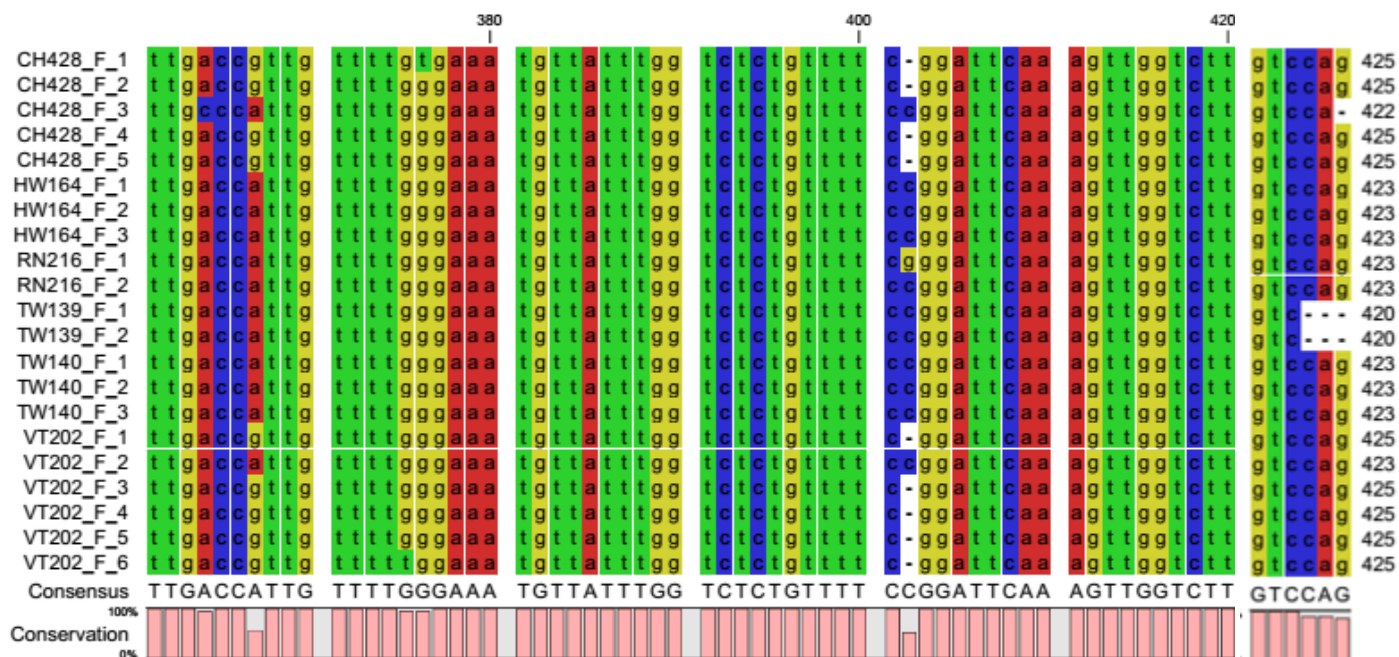


Figura 14: Alineamiento múltiple de secuencias del marcador de sexo de *B. papyrifera*, agrupadas según sexo del individuo. a) Secuencias masculinas, b) Secuencias femeninas. El código de nombres está explicado en la Figura 12.

Tabla 4: Secuencias clonadas por individuo y número de secuencias diferentes identificadas.

Muestra	Sexo	Localidad	Secuencias alineadas	Secuencias diferentes
BQUCH0 137	M	Taiwán	3	3
BQUCH0138	M	Taiwán	8	7
BQUCH0139	F	Taiwán	2	1
BQUCH0140	F	Taiwán	3	2
BQUCH0202	F	Vietnam	6	5
BQUCH0203	M	Vietnam	5	5
BQUCH0428	F	China	5	4
BQUCH0430	M	China	6	6
BQUCH0164	F	Hawái	3	1
BQUCH0186	M	Hawái	3	3
BQUCH0194	M	Hawái	3	3
BQUCH0216	F	Rapa Nui	2	2

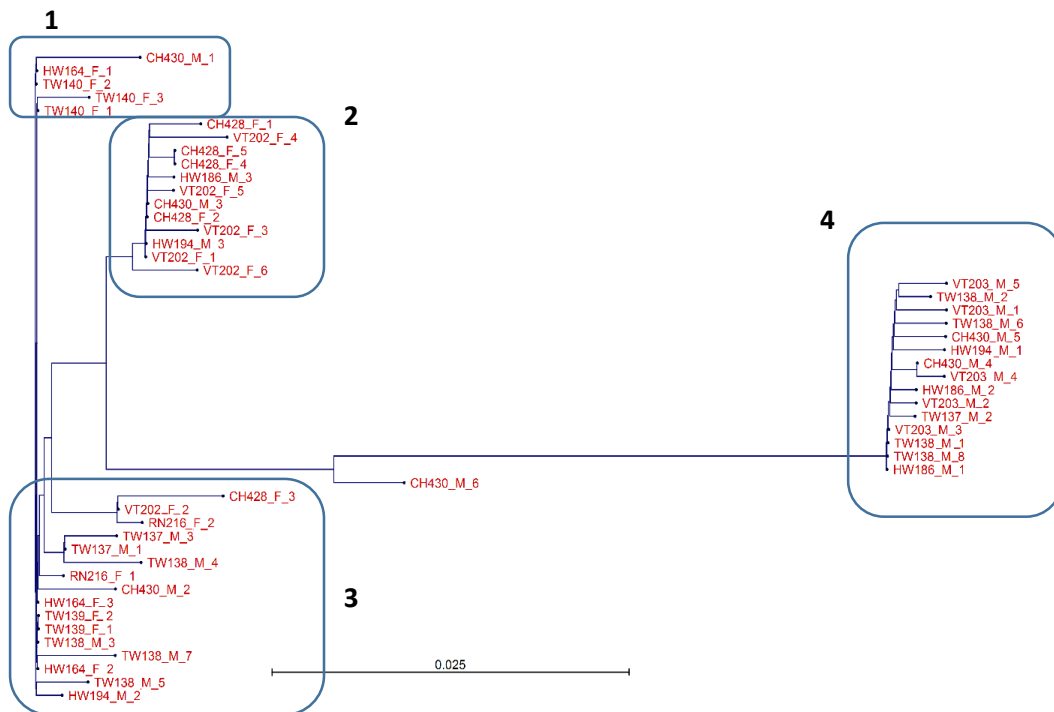
**Tabla 5: Porcentaje de identidad de las secuencias por individuo de *B. papyrifera*.**

Muestra	Sexo	Porcentaje de identidad
BQUCH0137	M	91.5
BQUCH0138	M	87.7
BQUCH0139	F	100
BQUCH0140	F	99.5
BQUCH0202	F	95.5
BQUCH0203	M	97.6
BQUCH0428	F	95.5
BQUCH0430	M	88
BQUCH0164	F	100
BQUCH0186	M	90.6
BQUCH0194	M	88.9
BQUCH0216	F	98.8

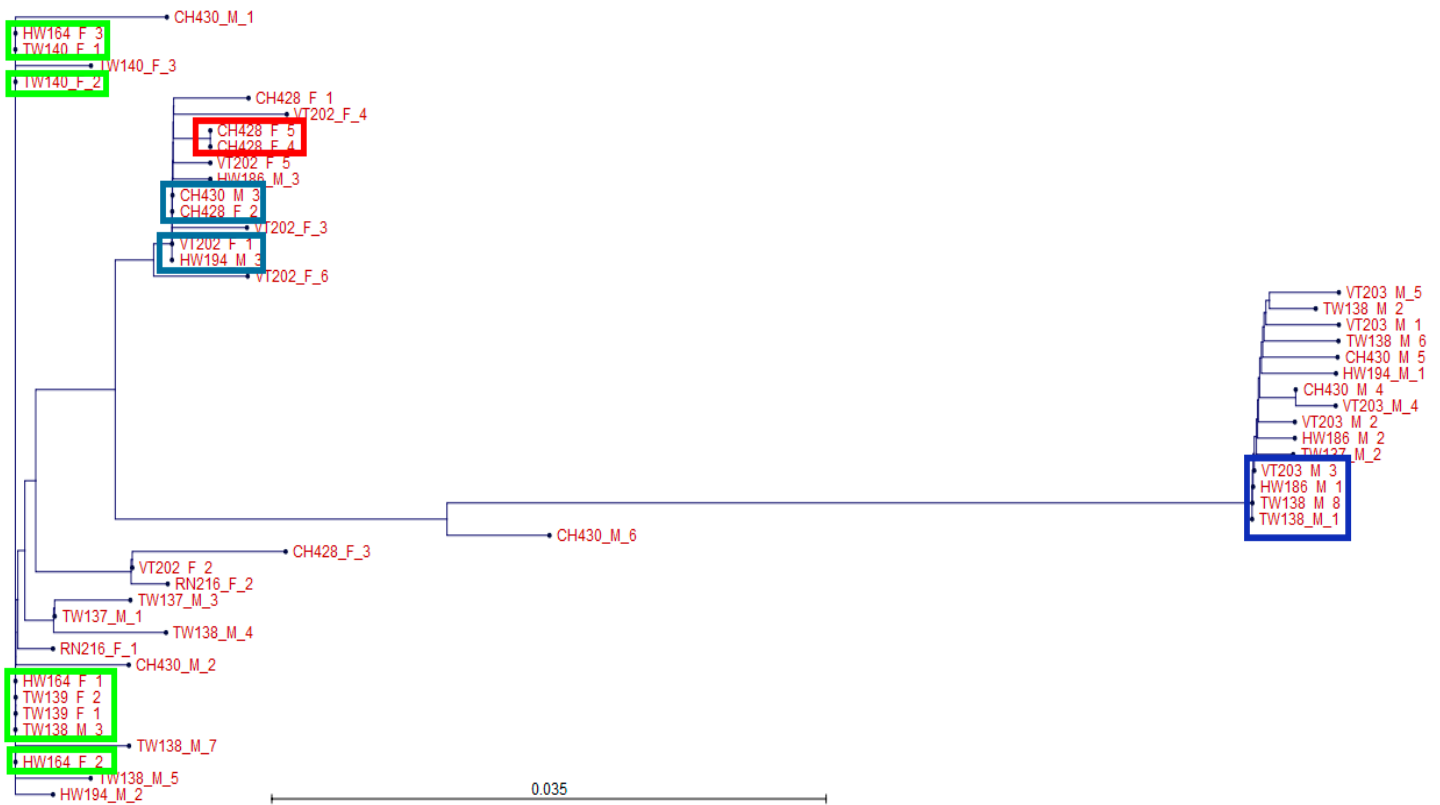
Las secuencias masculinas presentaron una identidad promedio de 90,7% entre las secuencias encontradas en los 6 individuos masculinos, con un rango de 87,7% a 97,6% mientras que las femeninas presentaron una identidad promedio de 98,2%, con un rango de 95,5% a 100%, lo que indica que las secuencias masculinas presentan mayor diversidad genética que las femeninas.

De un total de 12 individuos se clonaron y secuenciaron 21 secuencias provenientes de seis individuos femeninos y 28 secuencias provenientes de seis plantas masculinas. Entre las 49 secuencias totales se identificaron 35 secuencias distintas, de las cuales 15 derivaron de plantas femeninas y 20 de plantas masculinas. Se encontraron secuencias compartidas entre individuos de sexo masculino correspondientes a BQUCH138\_1, BQUCH138\_8, BQUCH186\_1 y BQUCH203\_3; las secuencias BQUCH138\_3, BQUCH139\_1, BQUCH140\_1, BQUCH140\_2, BQUCH164\_1, BQUCH164\_2 y BQUCH164\_3 son iguales entre sí y provienen de individuos de ambos sexos; las secuencias BQUCH194\_3, BQUCH202\_1, BQUCH428\_2 y BQUCH430\_3 también son iguales entre sí y pertenecen también a individuos de ambos sexos; por último las secuencias BQUCH428\_4 y BQUCH428\_5 son idénticas y provienen de una planta femenina. Estos resultados explican la imposibilidad de obtener secuencias de calidad a partir de productos de PCR en individuos masculinos. Como se observa en las **Figuras 12 y 13**, existen diferencias entre las secuencias del marcador de sexo entre individuos masculinos y femeninos de *B. papyrifera*. Considerando estas diferencias, es posible especular que el marcador de sexo podría estar implicado en la expresión génica diferencial de la planta, es decir, posiblemente corresponda a algún tipo de secuencia regulatoria de la expresión génica que tenga relación con el desarrollo de órganos sexuales en la planta, tales como la flor masculina y la femenina en individuos distintos.

Con el fin de analizar gráficamente las diferencias entre las secuencias de las muestras se construyó un dendrograma con todas las secuencias usando el algoritmo Kimura 80/Neighbor Joining [Hillis y cols., 1994; Huelsenbeck, 1994] en CLC Sequence Viewer 7.5 (Figura 15). En este agrupamiento se observan cuatro grupos o clados principales. Los tres primeros grupos son heterogéneos y están compuestos por 34 secuencias, tanto femeninas como masculinas, mientras que el tercer grupo está compuesto por 15 secuencias, pertenecientes a las muestras BQUCH0137, BQUCH0138, BQUCH0194, BQUCH0186, BQUCH0203 y BQUCH0430, todos individuos masculinos. De estas cinco muestras masculinas, cuatro también poseen secuencias en los otros grupos (BQUCH0137, BQUCH0138, BQUCH0194, BQUCH0186 y BQUCH0430), lo que indica que los individuos masculinos poseen versiones distintas del marcador de sexo, una de las cuales agrupa con secuencias femeninas. En un nuevo dendrograma se localizaron las secuencias idénticas, destacándose que son secuencias que se comparten entre individuos de distinto sexo y localidad geográfica (Figura 16).



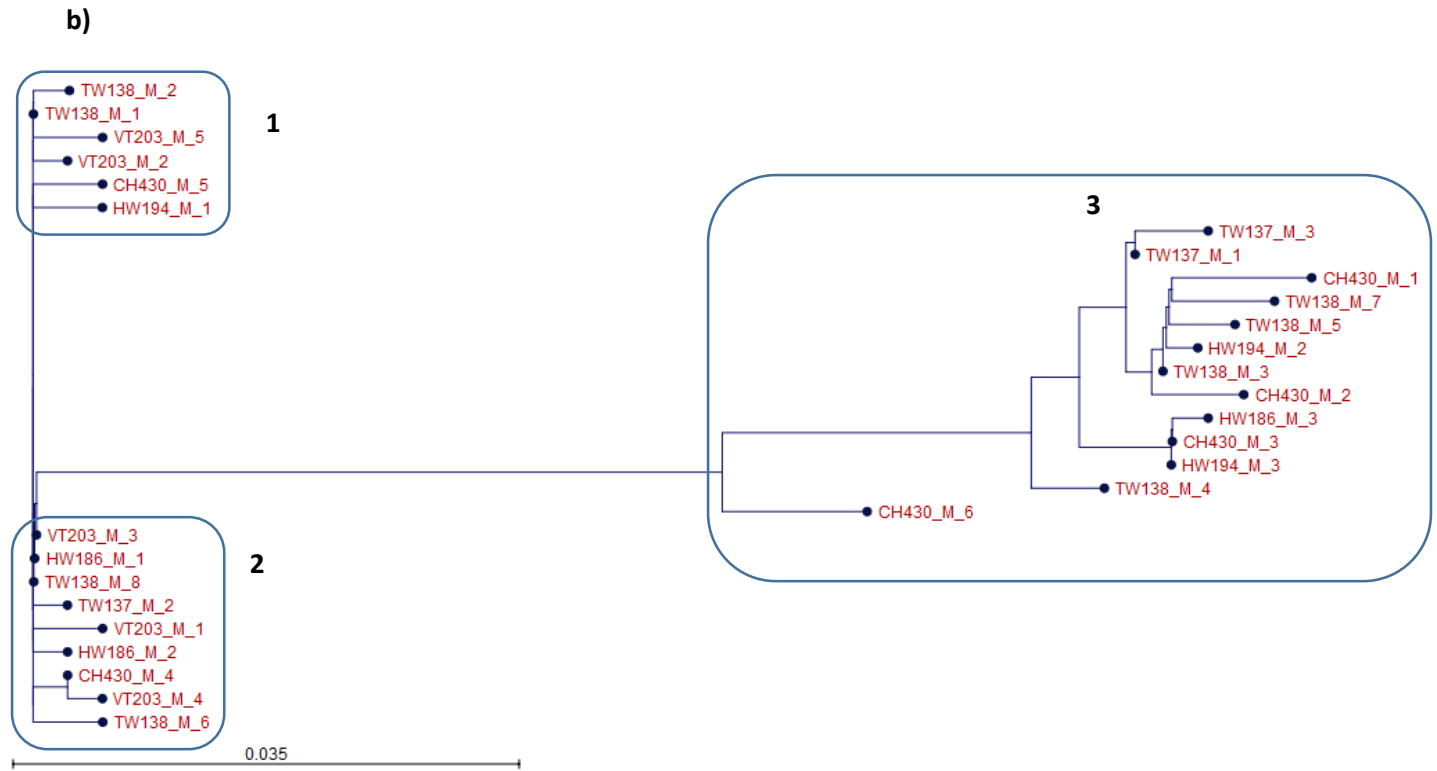
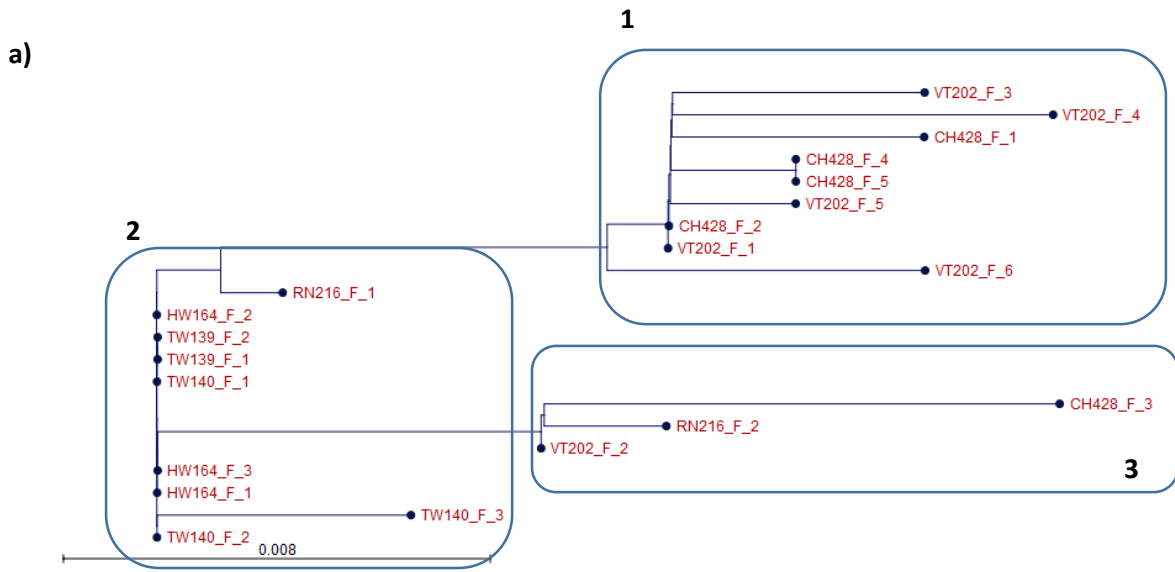
**Figura 15: Dendrograma Kimura 80/Neighbor Joining de las secuencias del marcador de sexo de *B. papyrifera*.** Se observa la formación de cuatro grupos o clados. Los tres primeros (1,2 y 3, a la izquierda), son heterogéneos en el agrupamiento de secuencias del marcador de sexo de individuos masculinos y femeninos. El grupo que forma un clado separado (grupo 4, a la derecha) está compuesto por 15 secuencias pertenecientes únicamente a 5 individuos masculinos, los cuales también poseen secuencias en los otros clados. CLC sequence viewer 7.5.



**Figura 16: Dendrograma Kimura 80/Neighbor Joining con grupos destacados de secuencias idénticas del marcador de sexo de *B. papyrifera*.** Se destacan en verde, azul claro, rojo y azul oscuro los grupos de secuencias idénticas y que por tanto coagrupan entre sí.

Luego se construyeron dendrogramas separados con las secuencias correspondientes a individuos femeninos (**Figura 17 a**) y masculinos (**Figura 17 b**) individuos por sexo. Las secuencias femeninas forman tres clados separados, al igual que las secuencias masculinas, pero con una distancia genética menor, lo que es concordante con el mayor porcentaje de identidad calculado para estas secuencias.





**Figura 17: Dendrograma Kimura 80/Neighbor Joining de las secuencias separadas por sexo.**

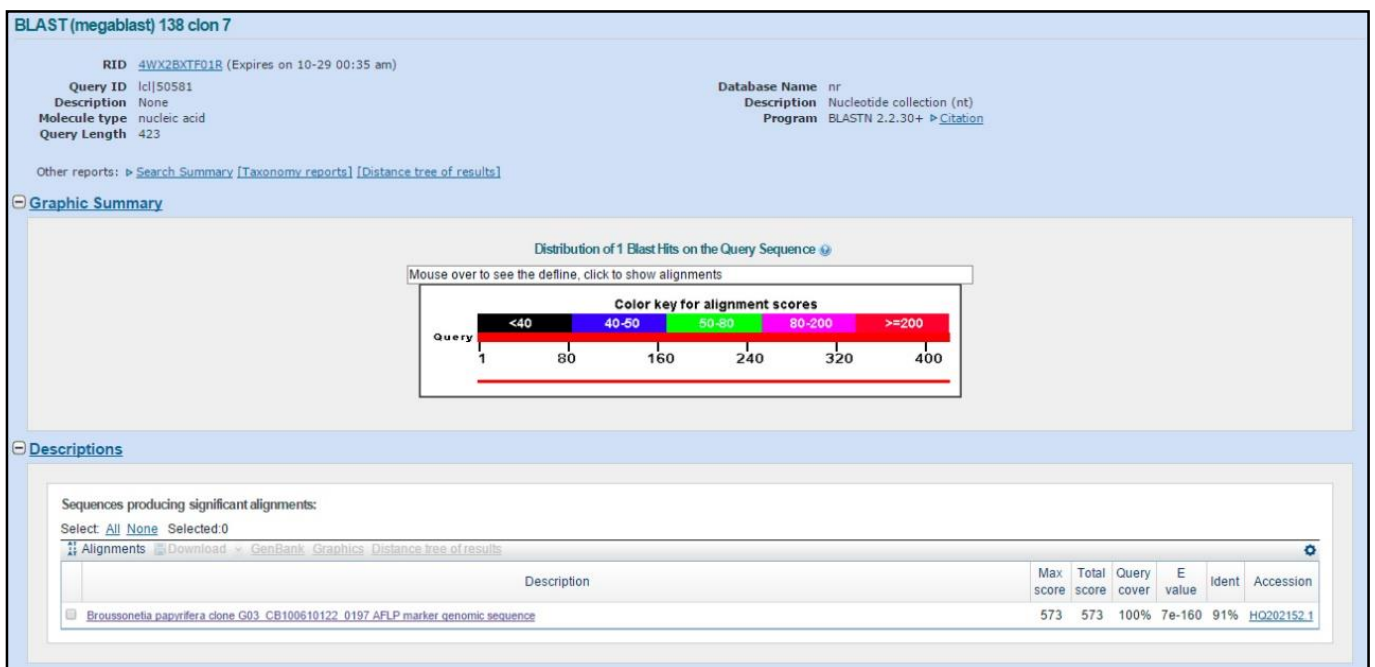
**a) Secuencias femeninas.** Se observa una distribución de tres clados. **b) Secuencias masculinas.** Se observan tres clados. Las secuencias correspondientes a los clados 1 y 2 son las mismas secuencias del clado 4 de la **Figura 15**. La escala de **Figura 17 a)** es menor que **Figura 17 b)**. CLC sequence viewer 7.5

## 5.3 Búsqueda de características de elemento regulatorio de la expresión génica.

### 5.3.1 Búsqueda de homología con secuencias depositadas en GenBank.

Para evaluar si la secuencia del marcador de sexo corresponde a una región regulatoria no codificante, se realizó una búsqueda de elementos regulatorios funcionales. Para comenzar se realizó un alineamiento de la secuencia del marcador de sexo de morera de papel utilizando BLAST usando la base de datos de secuencias nucleotídicas de GenBank, para encontrar una región análoga en otras especies. El resultado de este análisis se muestra en la **Figura 18**.

El resultado obtenido en la **Figura 18** indica que la secuencia 7 de la muestra 138 solo coincide con la secuencia HQ202152.1 y por tanto no se encuentra otra secuencia depositada en las bases de datos de GenBank que tenga similitud con el marcador de sexo de la morera de papel. Esta base de datos incluye a secuencias de distintos organismos modelo. Se verificó este resultado utilizando distintos algoritmos de alineamiento en BLAST. Por esta razón no es posible realizar una búsqueda directa mediante comparación con secuencias depositadas en GenBank.



**Figura 18: Búsqueda de secuencias homólogas usando BLAST.**

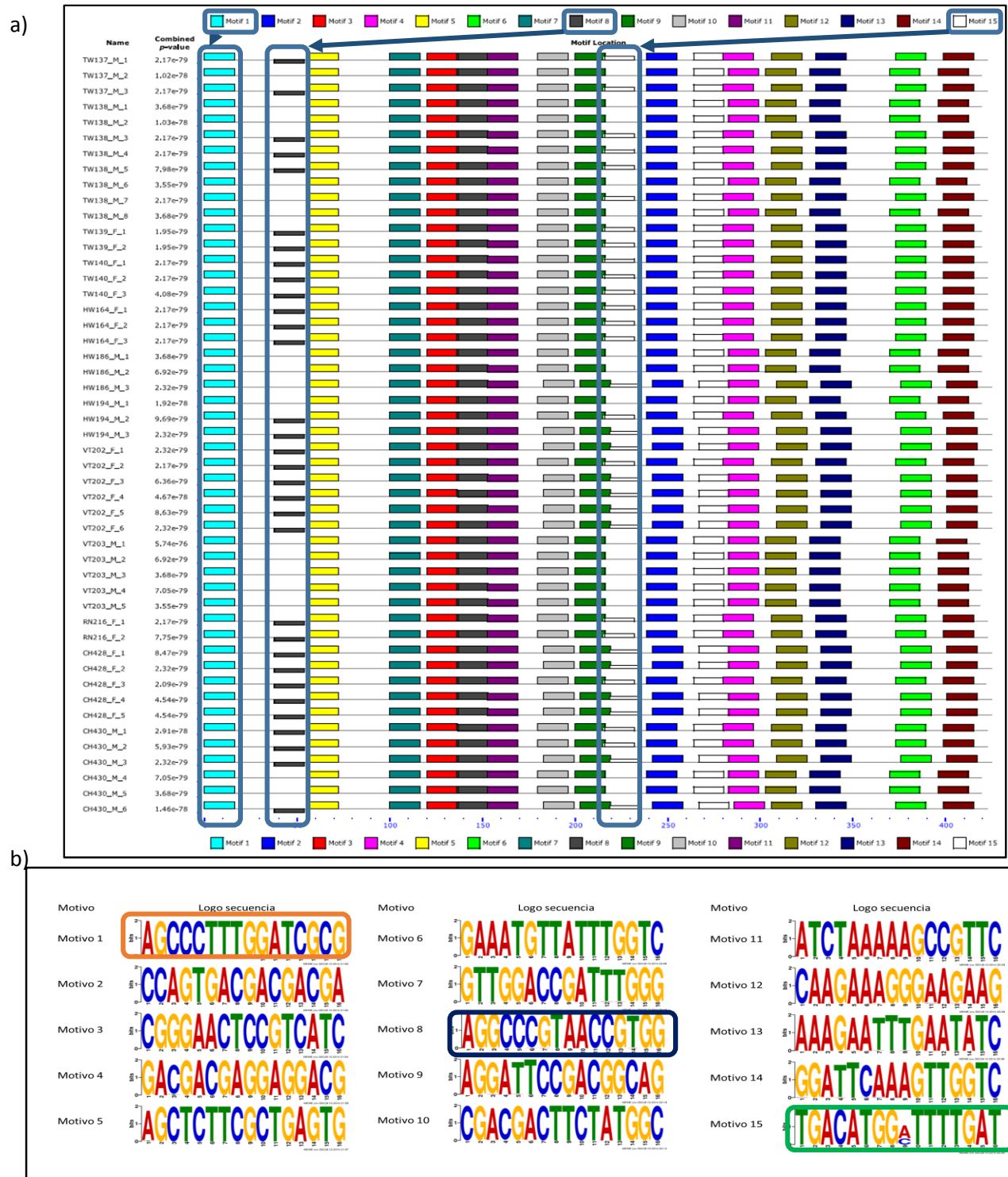
La secuencia de la muestra 138/secuencia 7 se alineó usando BLASTN/MEGABLAST, obteniéndose como homóloga solo a la secuencia HQ202152.1 depositada en GenBank por Wang y cols. (2012).

### 5.3.2 Búsqueda analítica de motivos putativos de unión de Factores de Transcripción utilizando la suite MEME.

#### 5.3.2.1 Utilización de MEME para buscar motivos en la secuencia del marcador de sexo.

En el caso de que el marcador de sexo de *B. papyrifera* participara en la regulación de la expresión génica, debería poseer sitios de unión para factores de transcripción [Maston y cols., 2006]. Dado que no existe una secuencia como parámetro que sirva de modelo (por no tener similitud significativa con otras secuencias), se utilizó la aproximación de la búsqueda analítica de motivos empleada por Lobos y colaboradores [2011] (Objetivo específico 3, análisis no convencional).

En ese estudio se utilizó MEME para buscar motivos con alguna significancia biológica, utilizando distintos parámetros y disponiendo solo de las secuencias de regiones río arriba del sitio de inicio de la transcripción de una familia de genes. Una vez identificados los motivos, mediante MEME se buscaron el (los) gene(s) que en sus regiones río arriba incluyeran cada uno de los motivos identificados en organismos modelo. A partir de esa colección de genes se establecieron redes de regulación transcripcional. En esta herramienta los parámetros se pueden variar hasta encontrar aquellos que permiten visualizar las diferencias de las secuencias en forma de que exhiban un distinto patrón de motivos, por tanto dos conjuntos distintos de parámetros no permiten obtener los mismos patrones. Basado en este trabajo, se realizó la búsqueda de motivos en la secuencia del marcador de sexo de *B. papyrifera* usando los parámetros ZOOPS (*Zero Or One Per Sequence*), ventana (ancho de motivo) de 3 a 16 pares de bases y un máximo de 15 motivos en las secuencias. El resultado obtenido en este caso se esquematiza en la **Figura 19 a**). Los “logos” (descripción gráfica de la conservación de secuencias en el alineamiento) se detallan en la **Figura 19 b**). De esta manera se encontraron 15 motivos de los cuales dos, el motivo 8 y el motivo 15, están ausentes en una segunda posición solamente en algunas de las secuencias masculinas. Es importante notar que las secuencias en las que no se repiten estos dos motivos son las mismas que forman un grupo aparte (grupo o clado 4) en el dendrograma de la **Figura 15**. Todos los motivos presentan un elevado grado de conservación en cuanto a los nucleótidos que los constituyen, según el grado de conservación representado en los *logos* y los valores p que oscilan entre  $1,95 \cdot 10^{-79}$  y  $5,74 \cdot 10^{-76}$  (valores teóricamente cercanos a cero y que implican baja posibilidad de un resultado falso positivo) para la combinación de motivos en cada una de las secuencias.



**Figura 19: Motivos encontrados en la secuencia del marcador de sexo de *B. papyrifera* usando MEME.**

a) Diagrama de motivos encontrados usando los parámetros ZOOPS, ventana 3-16 pb y un máximo de 15 motivos. Se muestran los motivos en cada una de las secuencias siguiendo el código de colores mostrado en la parte superior de la imagen. Se destacan las segundas posiciones variables para los motivos 8 (gris) y 15 (blanco), y el motivo 1 (más conservado), b) Logos de cada uno de los motivos, donde aparecen destacados los motivos 1 (cuadro naranja), 8 (cuadro azul oscuro) y 15 (cuadro verde).

### 5.3.2.2 Utilización de MAST para la identificación de genes que comparten los motivos encontrados en el marcador de sexo de *B. papyrifera*.

Continuando con el objetivo específico 3, una vez obtenidos los motivos en MEME, se utilizó la herramienta MAST (*Motif Alignment and Search Tool*) para identificar genes de *A. thaliana* y de *S. cerevisiae* cuyas regiones río arriba poseen los mismos motivos identificados en el marcador de sexo de *B. papyrifera*, tanto individualmente como agrupados. Esta búsqueda está acotada por las características de las bases de datos a un largo máximo de 1200 pb de región promotora: 1000 pb río arriba y 200 pb río abajo del sitio de inicio de la transcripción. En la **Tabla 6 a)** se indican los genes de *A. thaliana* con cada uno de los motivos, en **6 b)** aparecen los genes que poseen una combinación de motivos. En la **Tabla 7 a)** se indican los genes de *S. cerevisiae* con cada uno de los motivos, en **7 b)** los genes que poseen una combinación de motivos. La búsqueda fue acotada con el fin de mostrar genes con un valor e máximo igual a 10 en orden creciente (mientras más cercano a cero, la posibilidad de que el resultado sea un falso positivo es menor). Las listas fueron truncadas a máximo de 10 genes por cada motivo.

Se buscó la anotación *Gene Ontology* (GO) en la base de datos de TAIR (The *Arabidopsis* Information Resource, [www.arabidopsis.org](http://www.arabidopsis.org)) para cada uno de los genes de *A. thaliana* cuyos promotores poseen los motivos 1, 8 y 15 del marcador de sexo de *B. papyrifera* identificados con MEME. Se incluyeron los genes cuyos promotores poseen el motivo 1 (**Tabla 8**) por ser el que posee mejor score, por tanto es el mejor conservado entre los motivos identificados. Los genes del motivo 8 (**Tabla 9**) y 15 (**Tabla 10**) se incluyeron por ser estos los motivos que no se repiten en las secuencias masculinas y que están ausentes en las secuencias que forman un clado separado en el dendrograma. Además se muestran los genes que poseen mejor puntuación con combinación de motivos (**Tabla 11**), por ser estos los que poseen regiones río arriba con más motivos de los encontrados en el marcador de sexo de *B. papyrifera*.

**Tabla 6: Genes de *A. thaliana* que en sus promotores poseen los motivos encontrados en el marcador de sexo de *B. papyrifera*.**

a)

Genes de <i>Arabidopsis thaliana</i> que poseen los motivos 1-5									
Motivo 1 (+) (-)	Valor e	Motivo 2 (+) (-)	Valor e	Motivo 3 (+) (-)	Valor e	Motivo 4 (+) (-)	Valor e	Motivo 5 (+) (-)	Valor e
AT3G13445 TBP1	3.9	AT1G26355 SP1L1	0.0011	AT1G15440 PWP2	1.3	AT3G10400 AT3G10400	0.031	AT4G36630 EMB2754	2.2
AT4G09220 AT4G09220	4	AT1G65541 AT1G65541	0.057	AT1G10570 OTS2	3.8	AT1G72175 AT1G72175	0.041	AT4G32260 AT4G32260	3.7
AT4G09270 AT4G09270	4	AT4G03590 AT4G03590	0.12	AT1G03600 PSB27	4	AT5G08550 JLP1	0.065	AT4G28420 AT4G28420	6.1
		AT3G26900 SKL1	0.13	AT1G33830 AT1G33830	4	AT1G43600 AT1G43600	0.076	AT3G22920 AT3G22920	6.3
		AT4G14510 CFM3B	0.14	AT1G33910 AT1G33910	4	AT1G43610 AT1G43610	0.076	AT4G37925 NDH-M	6.7
		AT4G23750 CRF2	0.14	AT1G60220 ULP1D	4	AT1G67840 CSK	0.076	AT1G03350 AT1G03350	6.7
		AT4G38470 AT4G38470	0.14	AT4G16210 ECHIA	4	AT2G34120 AT2G34120	0.076	AT1G17410 AT1G17410	7.1
		AT5G04430 BTR1L	0.14			AT1G65280 AT1G65280	0.21		
		AT1G35890 AT1G35890	0.88			AT2G21970 SEP2	0.21		
		AT4G28060 AT4G28060	1.1			AT4G24030 AT4G24030	0.36		
Genes de <i>Arabidopsis thaliana</i> que poseen los motivos 6-10									
Motivo 6 (+) (-)	Valor e	Motivo 7 (+) (-)	Valor e	Motivo 8 (+) (-)	Valor e	Motivo 9 (+) (-)	Valor e	Motivo 10 (+) (-)	Valor e
AT2G03890 PI4K	2.9	AT1G17760 CSTF77	0.63	AT1G01880 AT1G01880	0.16	AT5G66420 AT5G66420	0.51	AT5G25820 AT5G25820	0.51
AT4G02550 AT4G02550	2.9	AT5G58280 AT5G58280	2.9	370200003 AT1G01890	0.41	AT5G51510 AT5G51510	0.65	AT5G47720 AT5G47720	0.95
AT4G39630 AT4G39630	2.9	AT3G07430 YLMG1-1	4.8	AT5G54080 HGO	2.4	AT3G58110 AT3G58110	0.7	AT4G24030 AT4G24030	1.2
		AT2G42940 AT2G42940	5.9	AT1G56630 AT1G56630	2.8	AT5G66420 AT5G66420	0.89	AT4G24026 AT4G24026	1.2
		AT1G04930 AT1G04930	7.1	AT3G48290 CYP71A24	2.8	AT1G49850 AT1G49850	1.3	AT1G75450 CKX5	2.2
		AT1G04945 AT1G04945	7.1	AT5G05540 SDN2	4.9	AT3G48590 NF-YC1	2.3	AT5G43245 AT5G43245	3.4
		AT1G04940 TIC20	7.1	AT3G55480 PAT2	5.1	AT3G51260 PAD1	3.2	AT5G15450 CLPB3	5.6
		AT1G66730 LLG6	7.2			AT1G27090 AT1G27090	4.5	AT1G07690 AT1G07690	7.6
		AT1G76040 CPK29	7.2			AT3G01930 AT3G01930	4.5	AT2G47030 VGDH1	7.6
		AT3G07900 AT3G07900	7.2			AT3G02800 AT3G02800	4.6	AT2G47040 VGD1	7.6
Genes de <i>Arabidopsis thaliana</i> que poseen los motivos 11-15									
Motivo 11 (+) (-)	Valor e	Motivo 12 (+) (-)	Valor e	Motivo 13 (+) (-)	Valor e	Motivo 14 (+) (-)	Valor e	Motivo 15 (+) (-)	Valor e
AT1G21920 AT1G21920	7.7	AT1G32000 AT1G32000	0.48	AT1G54760 AGL85	1	AT1G07700 AT1G07700	1	AT1G21080 AT1G21080	2.9
		AT5G03440 AT5G03440	0.64	AT1G01320 AT1G01320	10	AT5G47730 AT5G47730	1	AT3G23610 DSPTP1	2.9
		3702000517 AT5G03445	0.64	AT1G47380 AT1G47380	10	AT5G40690 AT5G40690	1.6	AT3G26600 ARO4	4.3
		AT2G38040 CAC3	0.71	AT2G04034 AT2G04034	10	AT2G40610 EXPA8	2	AT5G05960 AT5G05960	7.8
		AT1G53620 AT1G53620	0.93	AT2G10014 AT2G10014	10	AT3G06960 PDE320	2	AT1G63280 AT1G63280	8.4
		AT1G58025 AT1G58025	0.93	AT3G09300 ORP3B	10	AT1G65120 AT1G65120	4.2	AT3G60790 AT3G60790	8.4
		AT1G69230 SP1L2	0.93			AT1G13340 AT1G13340	4.8	AT1G28135 AT1G28135	8.4
		AT2G21910 CYP96A5	0.03			AT5G01930 AT5G01930	5.1	AT1G54940 PGSIP4	8.4
		AT2G35510 SR01	0.93			AT3G19850 AT3G19850	5.1	AT4G18220 AT4G18220	8.4
		AT4G09690 AT4G09690	0.93			AT2G26200 AT2G26200	6.6	AT5G09330 INAC082	8.4

b)

Genes de <i>Arabidopsis thaliana</i> con combinación de motivos	
Todos los motivos (+) (-)	Valor e
AT1G49620 ICK5	0.0024
AT3G10400 AT3G10400	0.057
AT4G14510 CFM3B	0.098
AT3G01170 AT3G01170	0.12
AT4G24030 AT4G24030	0.13
AT4G24026 AT4G24026	0.13
AT5G04010 AT5G04010	0.17
AT5G44574 AT5G44574	0.17
AT4G38470 AT4G38470	0.23
AT3G01160 AT3G01160	0.32

a) Lista de genes de *A. thaliana* que en sus promotores poseen los motivos 1 al 15 encontrados en el marcador de sexo de *B. papyrifera*. b) Lista de genes de *A. thaliana* que poseen en sus regiones río arriba una combinación de motivos encontrados en el marcador de sexo de *B. papyrifera*. La lista está compuesta por genes con un valor e menor o igual a 10 y y/o truncada a un máximo de 10 genes.

**Tabla 7: Lista de genes de *S. cerevisiae* que en sus promotores poseen los motivos encontrados en el marcador de sexo de *B. papyrifera*.**

a)

Genes de <i>Saccharomyces cerevisiae</i> que poseen los motivos 1-5									
Motivo 1 (+) (-)	valor e	Motivo 2 (+) (-)	valor e	Motivo 3 (+) (-)	valor e	Motivo 4 (+) (-)	valor e	Motivo 5 (+) (-)	valor e
YHR169W YHR169W	0.43	YJR093C YJR093C	0.49	YBR091C YBR091C	5.9	YGL244W YGL244W	0.027	YLR099W-A YLR099W-A	0.8
YKL201C YKL201C	0.88	YML092C YML092C	0.51	YPR118W YPR118W	6.6	YLR055C YLR055C	0.05	YNR022C YNR022C	1.1
YDL100C YDL100C	6.4	YEL035C YEL035C	4.7	YLR417W YLR417W	8.9	YGL023C YGL023C	0.28	YMR286W YMR286W	1.1
YHL020C YHL020C	6.8	YMR160W YMR160W	4.8			YBR231C YBR231C	0.3	YIL056W YIL056W	1.6
YDR430C YDR430C	7.4	YMR156C YMR156C	6.6			YOR141C YOR141C	0.89	YHR155W YHR155W	4.5
YDR432W YDR432W	7.4	YGL008C YGL008C	6.8			YOR023C YOR023C	0.94		
YOR006C YOR006C	7.5	YKL027W YKL027W	8			YGR197C YGR197C	1		
YDR414C YDR414C	8	YBL047C YBL047C	8.4			YPL036W YPL036W	1.5		
YMR190C YMR190C	8.7	YLR153C YLR153C	8.8			YDR171W YDR171W	1.6		
		YBR271W YBR271W	9.5			YAL011W YAL011W	4		
Genes de <i>Saccharomyces cerevisiae</i> que poseen los motivos 6-10									
Motivo 6 (+) (-)	valor e	Motivo 7 (+) (-)	valor e	Motivo 8 (+) (-)	valor e	Motivo 9 (+) (-)	valor e	Motivo 10 (+) (-)	valor e
YPL161C YPL161C	0.34	YLR213C YLR213C	5.9	YKL198C YKL198C	1.8	YJL219W YJL219W	1.7	YPR058W YPR058W	0.65
YLR045C YLR045C	3.8	YML043C YML043C	8.5	YPL145C YPL145C	4.8	YOL156W YOL156W	1.7	YHR169W YHR169W	1.4
YGL100W YGL100W	4.5	YML042W YML042W	8.5	YPL144W YPL144W	4.8			YOR097C YOR097C	4.4
YLR004C YLR004C	5.7	YNR016C YNR016C	8.5	YCL038C YCL038C	5.2			YDR084C YDR084C	4.9
YLR005W YLR005W	5.7	YNR017W YNR017W	8.5	YJL029C YJL029C	8			YCR024C-A YCR024C-A	5.9
YPL191C YPL191C	6.4			YJL028W YJL028W	8			YMR177W YMR177W	8.1
YLR228C YLR228C	9.2			YAR014C YAR014C	8.5			YBL113C YBL113C	8.2
								YCR020C YCR020C	8.7
Genes de <i>Saccharomyces cerevisiae</i> que poseen los motivos 11-15									
Motivo 11 (+) (-)	valor e	Motivo 12 (+) (-)	valor e	Motivo 13 (+) (-)	valor e	Motivo 14 (+) (-)	valor e	Motivo 15 (+) (-)	valor e
YLR248W YLR248W	0.048	YOR252W YOR252W	0.22	YBL036C YBL036C	0.44	YLR351C YLR351C	0.32	YPL096W YPL096W	2.9
YGR078C YGR078C	4.5	YDL024C YDL024C	0.68	YAL051W YAL051W	2.4	YKL189W YKL189W	0.68	YKL022C YKL022C	3.3
YER105C YER105C	4.7	YCR004C YCR004C	0.81	YMR243C YMR243C	3.6	YBR210W YBR210W	1.3	YIR026C YIR026C	3.7
YER106W YER106W	4.7	YHR124W YHR124W	1.2	YMR244W YMR244W	3.6	YDR233C YDR233C	1.7	YOR293C-A YOR293C-A	4.4
YGL133W YGL133W	5.8	YPL122C YPL122C	1.4	YPL249C-A YPL249C-A	4.1	YDR234W YDR234W	1.7	YGL254W YGL254W	4.5
YMR137C YMR137C	7.7	YBL091C YBL091C	1.4	YER186C YER186C	5.6	YOR165W YOR165W	3.7	YMR238W YMR238W	4.8
YMR138W YMR138W	7.7	YLR117C YLR117C	1.5	YGL098W YGL098W	7.6	YLR047C YLR047C	4.4	YPR117W YPR117W	5.7
		YIL016W YIL016W	2.5	YPL101W YPL101W	7.7	YLR048W YLR048W	4.4	YBR057C YBR057C	6.6
		YML064C YML064C	2.9	YIL169C YIL169C	9.3	YBR073W YBR073W	6.2	YLR307C-A YLR307C-A	7.8
		YML063W YML063W	2.9			YLR120C YLR120C	6.3	YLR308W YLR308W	7.8

b)

Genes de <i>Saccharomyces cerevisiae</i> con combinación de motivos	
Todos los motivos (+) (-)	Valor e
YGL023C YGL023C	0.42
YHR169W YHR169W	0.43
YLR055C YLR055C	0.7
YBR019C YBR019C	1.9
YBR020W YBR020W	2.2
YNL103W YNL103W	3.7
YER116C YER116C	4.7
YHR061C YHR061C	6.1
YPL249C-A YPL249C-A	6.3
YLR248W YLR248W	6.9

a) Lista de genes de *S. cerevisiae* que en sus promotores poseen los motivos 1 al 15 encontrados en el marcador de sexo de *B. papyrifera*. b) Lista de genes de *S. cerevisiae* que poseen combinación de motivos encontrados en marcador de sexo de *B. papyrifera*. La lista está compuesta por genes con un valor e menor o igual a 10 y/o truncada a un máximo de 10 genes.

**Tabla 8: Genes de *A. thaliana* que poseen el motivo 1 del marcador de sexo de *B. papyrifera* en sus promotores.**

Gen	Descripción de la proteína	Expresión	Anotación GO	
AT3G13445 TBP1	TBP ( <i>TATA binding protein</i> ) se asocia con TAF(II) ( <i>TBP-associated factors</i> ) para formar el complejo general de transcripción TFIID.	Ubicua	Proceso biológico	Regulación de la transcripción DNA dependiente, iniciación de la transcripción desde promotores RNA pol II.
			Componente celular	Núcleo
			Función molecular	Unión a DNA
AT4G09220 AT4G09220	Proteína desconocida		Proceso biológico	Desconocido
			Componente celular	Desconocido
			Función molecular	Desconocido
AT4G09270 AT4G09270	Proteína desconocida		Proceso biológico	Desconocido
			Componente celular	Desconocido
			Función molecular	Desconocido

En la lista se muestran los genes de *A. thaliana* que poseen el motivo 1 en su región río arriba. En dos de los tres genes, no se ha descrito una proteína y por tanto se desconoce su función, mientras que el único gen descrito codifica para una proteína TBP (*TATA binding protein*) que actúa como factor de transcripción.

**Tabla 9: Genes de *A. thaliana* que poseen el motivo 8 del marcador de sexo de *B. papyrifera* en sus promotores.**

Gen	Descripción de la proteína	Expresión	Anotación GO	
AT1G01880 AT1G01880	Miembro de la familia de nucleasas Rad2/XPG de clase 4 que procesan las <i>Holliday Junctions</i> de manera análoga a las resolvasas HJ de fagos, arqueas y bacterias.	Células guardianas del estoma.	Proceso biológico	Reparación de DNA.
			Componente celular	Núcleo.
			Función molecular	Actividad endodeoxirribonucleasa de <i>crossover junction</i> .
3702000003 AT1G01890	tRNA-lys.	Ubicua	Proceso biológico	Elongación de la traducción.
			Componente celular	Desconocido.
			Función molecular	Actividad de adaptador triplete-aminoácido
AT5G54080 HGO	1,2-dioxygenase que puede convertir homogentisato to malilacetoacetato.	Ubicua	Proceso biológico	Proceso catabólico de L-fenilalanina, proceso catabólico de clorofila, beta-oxidación de ácidos grasos.



			Componente celular	Citoplasma, citosol.
			Función molecular	Actividad 1,2-dioxigenasa de homogentisato.
<b>AT1G56630 AT1G56630</b>	Proteína de la superfamilia de alfa/beta hidrolasas.	Desconocida	Proceso biológico	Proceso metabólico de lípidos.
			Componente celular	Mitocondria.
			Función molecular	Actividad triglicérido lipasa.
<b>AT3G48290 CYP71A24</b>	Citocromo p450 putativo.	Células guardianas del estoma	Proceso biológico	Proceso de oxidación reducción.
			Componente celular	Unión a oxígeno.
			Función molecular	Actividad de transportador de electrones.
<b>AT5G05540 SDN2</b>	Nucleasa 2 de RNAs pequeños.	Ubicua	Proceso biológico	Desconocido.
			Componente celular	Intracelular, núcleo.
			Función molecular	Actividad exonucleasa, unión a ácidos nucleicos.
<b>AT3G55480 PAT2</b>	Media la biogénesis y función de vacuolas líticas, que puede complementar parcialmente la levadura mutante correspondiente	Ubicua	Proceso biológico	Transporte de Golgi a vacuola, organización de vacuolas líticas, regulación del pH intracelular.
			Componente celular	Aparato de Golgi, núcleo, membrana.
			Función molecular	Actividad transportadora de proteínas.

En la lista se muestran los genes de *A. thaliana* que poseen el motivo 8 en su región río arriba. Los genes son diversos en cuanto a función. Los genes con mejor puntuación codifican para resolvasas de *Holliday junctions*, para tRNA de lisina, para proteínas de vías metabólicas, citocromo p450, nucleasas de RNAs pequeños y proteínas mediadoras de formación de vacuolas líticas. Muchos de estos genes parecen ser importantes en la sobrevivencia de *A. thaliana*, como el tRNA de lisina y los participantes de vías metabólicas.

**Tabla 10: Genes de *A. thaliana* que poseen el motivo 15 del marcador de sexo de *B. papyrifera* en sus promotores.**

Gen	Descripción de la proteína	Expresión	Anotación GO	
<b>AT1G21080 AT1G21080</b>	Proteína que contiene dominio N-terminal DNAJ de shock térmico.	Ubicua	Proceso biológico	Plegamiento de proteínas.
			Componente celular	Aparato de Golgi, citoplasma.

			Función molecular	Unión de proteínas de <i>shock</i> térmico, unión de proteínas desplegadas.
AT3G23610 DSPTP1	Codifica una proteína fosfatasa con especificidad dual cuya actividad es regulada por unión a CaM	Ubicua	Proceso biológico	Inactivación de MAPK, defosforilación, proceso biosintético de fosfatidilinositol, respuesta celular a frío.
			Componente celular	Cloroplasto.
			Función molecular	Unión a calmodulina, actividad MAPK fosfatasa.
AT3G26600 ARO4	Proteína de repetición de armadillo. Se expresa en tejidos vegetativos, anteras y óvulos.	No se expresa en polen	Proceso biológico	Desconocido.
			Componente celular	Plasmodesma.
			Función molecular	Desconocido.
AT5G05960 AT5G05960	Proteína miembro de la superfamilia 2S albúmina bifuncional inhibidora/transportadora de lípidos almacenaje de proteína en semilla	Ubicua	Proceso biológico	Transporte de lípidos.
			Componente celular	Región extracelular.
			Función molecular	Unión a lípidos.
AT1G63280 AT1G63280	Proteína de la familia de las inhibidoras de serina proteasa (SERPIN)	Desconocido	Proceso biológico	Desconocido.
			Componente celular	Citoplasma.
			Función molecular	Actividad inhibidora endopeptidasa tipo serina.
AT3G60790 AT3G60790	Proteína de la familia de las F-box	Desconocido	Proceso biológico	Desconocido.
			Componente celular	Cloroplasto.
			Función molecular	Desconocido.
AT1G28135 AT1G28135	Proteína desconocida	Desconocido	Proceso biológico	Desconocido.
			Componente celular	Desconocido.
			Función molecular	Desconocido.
AT1G54940 PGSIP4	Codifica una xilano glucuronosiltransferasa	Hoja, flor, pétalo, células espermáticas, raíz, sépalo	Proceso biológico	Proceso biosintético.
			Componente celular	Aparato de Golgi.
			Función molecular	Actividad glucuronosiltransferasa, actividad transferasa.
AT4G18220 AT4G18220	Proteína de la superfamilia de los transportadores de drogas y metabolitos	Células guardianas del estoma	Proceso biológico	Transportador de compuestos que contienen nucleobase.
			Componente celular	Desconocido.
			Función molecular	Actividad de transportador nucleobase púricas transmembrana
AT5G09330 NAC082		Ubicua	Proceso biológico	Desarrollo multicelular, actividad de factor de

	Proteína 82 que contiene dominio NAC			transcripción dependiente DNA
			Componente celular	Núcleo
			Función molecular	Unión a DNA, actividad de factor de transcripción secuencia específico.

Los genes de *A. thaliana* que poseen el motivo 15 en sus regiones río arriba codifican proteínas con gran diversidad de funciones. Se muestran proteínas de shock térmico, proteínas fosfatasa y de señalización y proteínas no descritas. Uno de los genes de la lista, AT3G26600|ARO4, se expresa en forma ubicua, a excepción del polen, lo que puede resultar importante considerando que se está tratando de establecer relaciones entre genes de *A. thaliana* y *B. papyrifera* que posean similitud en sus regiones río arriba que puedan estar involucrados en desarrollo de fenotipo diferenciado sexualmente.

**Tabla 11: Genes de *A. thaliana* que poseen una combinación de los motivos encontrados en el marcador de sexo de *B. papyrifera* en sus promotores.**

Gen	Descripción de la proteína	Expresión	Anotación GO	
AT1G49620 ICK5	CKI Inhibidor de quinasa dependiente de ciclina (CDK)	Ubicua, no se expresa en <i>shoot apical meristem</i>	Proceso biológico	Endorreduplicación de DNA, arresto del ciclo celular, progresión de línea estomatal.
			Componente celular	Núcleo de células germinativas masculinas.
			Función molecular	Unión a ciclina, unión a proteína, actividad de inhibidor a serina/treonina quinasa dependiente de ciclina.
AT3G10400 AT3G10400	Codifica una proteína U11/U12-31K tipo U12 espliceosomal.  Implicada en <i>splicing</i> U12 de intrones vía actividad chaperona de RNA y que afecta el desarrollo de la planta.	Ubicua	Proceso biológico	Proceso de desarrollo, <i>splicing</i> de RNA.
			Componente celular	Núcleo.
			Función molecular	Unión de ácidos nucleicos, unión de nucleótidos, unión a ión zinc.
AT4G14510 CFM3B	Codifica una proteína con dominio CRM, homóloga de CDM3a, la que posee función in el <i>splicing</i> en intrones del grupo IIB en cloroplastos.	Ubicua	Proceso biológico	Desarrollo de semilla.
			Componente celular	Cloroplasto.
			Función molecular	Unión de RNA.
AT3G01170 AT3G01170	Proteína de la superfamilia de proteína ribosomal L34e	Ubicua	Proceso biológico	Traducción.
			Componente celular	Intracelular, ribosoma.
			Función molecular	Constituyente estructural del ribosoma.
AT4G24030 AT4G24030	Proteína desconocida	Desconocido	Proceso biológico	Desconocido.

			Componente celular	Desconocido.
			Función molecular	Desconocido.
<b>AT4G24026 AT4G24026</b>	Proteína desconocida	Células guardianas del estoma	Proceso biológico	Desconocido
			Componente celular	Núcleo
			Función molecular	Desconocido
<b>AT5G04010 AT5G04010</b>	Proteína de la familia F-box	Desconocido	Proceso biológico	Desconocido.
			Componente celular	Desconocido.
			Función molecular	Desconocido.
<b>AT5G44574 AT5G44574</b>	Proteína desconocida	Células guardianas del estoma	Proceso biológico	Desconocido.
			Componente celular	Mitocondria.
			Función molecular	Desconocido.
<b>AT4G38470 AT4G38470</b>	Proteína de la familia de ACT-like con actividad proteína serina/treonina/tirosina quinasa.	Ubicua	Proceso biológico	Organización de filamentos de actina, organización de cloroplasto, procesos metabólicos, fosforilación de proteínas.
			Componente celular	Citosol, membrana plasmática.
			Función molecular	Unión de ATP, unión de aminoácidos, actividad proteína quinasa, actividad transferasa.
<b>AT3G01160 AT3G01160</b>	Proteína desconocida		Proceso biológico	Desconocido.
			Componente celular	Citosol, núcleo.
			Función molecular	Desconocido.

Los genes de *A. thaliana* que contienen una combinación de los motivos encontrados en el marcador de sexo de *B. papyrifera* codifican para proteínas que están involucradas en la detención del ciclo celular y en el desarrollo de la planta por poseer actividad de *splicing*, por lo que estos genes podrían modular el desarrollo de un fenotipo característico de la planta. Además aparecen genes que codifican para probables proteínas fosfatasa y otros genes cuyos productos son aún desconocidos.

Utilizando la base de datos Yeastmine de SGD (*Saccharomyces* genome database, <http://yeastmine.yeastgenome.org/yeastmine/begin.do>) se buscó la anotación GO (proceso biológico, componente celular y función molecular) de cada uno los genes de *S. cerevisiae* que poseen los motivos identificados en el marcador de sexo de *B. papyrifera* en sus regiones río arriba. Al igual que para los genes de *A. thaliana* se muestra la lista de genes que poseen el motivo 1 (**Tabla 12**), por ser el que posee mejor *score* en la búsqueda. La lista de genes que poseen motivos 8 y 15, que difieren en las secuencias

masculinas, se muestran en las **Tablas 13 y 14**, respectivamente. Los genes que poseen mejor puntuación con combinación de motivos se indican en la **Tabla 15**.

**Tabla 12: Genes de *S. cerevisiae* que poseen el motivo 1 del marcador de sexo de *B. papyrifera* en sus promotores.**

Gen	Descripción de la proteína	Anotación GO	
YHR169W YHR169W	RNA helicasa, ATPasa putativa de la familia de las DEAD-box	Proceso biológico	Ruptura endonucleolítica para generar extremo 5' maduro en SSU-rRNA desde (SSU-rRNA 5.8S rRNA, LSU-rRNA), biogénesis de ribosoma.
		Componente celular	Prerrribosoma 90S, nucléolo.
		Función molecular	Actividad helicasa, unión a ATP, unión a nucleótido, unión a RNA.
YKL201C YKL201C	Regulador putativo positivo de la manosilfosfato transferasa Mnn6p	Proceso biológico	Glicosilación de proteínas N-unidas y O-unidas, respuesta a estrés osmótico.
		Componente celular	Componente integral de membrana.
		Función molecular	Activador enzimático.
YDL100C YDL100C	Factor de intercambio de guanina para Gpa1p	Proceso biológico	Inserción de proteínas dentro de la membrana del ER, proceso metabólico, direccionamiento postraducciona de proteínas a la membrana.
		Componente celular	Complejo GET, citoplasma, retículo endoplásmico.
		Función molecular	Actividad hidrolasa, unión a ión metálico.
YHL020C YHL020C	Regulador transcripcional de una variedad de genes	Proceso biológico	Proceso biosintético de fosfolípido, regulación positiva de la transcripción en promotor RNA pol II.
		Componente celular	Membrana nuclear, núcleo, nucleoplasma.
		Función molecular	Unión a DNA
YDR430C YDR430C	Metaloproteasa lisina-específica de la familia de pitrilisina	Proceso biológico	Proteólisis, actividad peptidasa.
		Componente celular	Espacio intermembrana de la mitocondria.
		Función molecular	Actividad metalopeptidasa, actividad hidrolasa, unión a ión metálico.
YDR432W YDR432W	Proteína de unión a RNA, promueve elongación, regula terminación	Proceso biológico	Exportación de mRNA desde el núcleo, regulador de la traducción.
		Componente celular	Complejo ribonucleoproteína, citoplasma, nucléolo.
		Función molecular	Unión a RNA, unión a poli(A).
YOR006C YOR006C	Proteína requerida para el procesamiento del pre rRNA 20S	Proceso biológico	Maduración de SSU-rRNA, procesamiento de rRNA, biogénesis de ribosoma.
		Componente celular	Citoplasma, núcleo.
		Función molecular	Desconocido.

YDR414C YDR414C	Proteína predicha de membrana requerida para la retención luminal de proteínas en el ER	Proceso biológico	Glicosilación de proteínas, y transporte, retención de proteínas en el lumen de ER.
		Componente celular	Membrana, retículo endoplásmico.
		Función molecular	Desconocido.
YMR190C YMR190C	DNA helicasa nucleolar de la familia RecQ	Proceso biológico	Segregación meiótica de cromosomas, respuesta ante estímulo de daño a DNA, actividad helicasa.
		Componente celular	Complejo helicasa RecQ-Topo III
		Función molecular	Actividad hidrolasa, unión a ATP, unión a nucleótido.

Los genes de *S. cerevisiae* que poseen el motivo 1 codifican para proteínas con funciones diversas, tal como en el caso de los de *A. thaliana*. Se describen proteínas con función de helicasas de RNA y DNA, proteínas involucradas en la estabilidad y estabilización de RNA, de procesamiento de rRNAs, de retención de proteínas en el retículo endoplásmico, entre otras.

**Tabla 13: Genes de *S. cerevisiae* que poseen el motivo 8 del marcador de sexo de *B. papyrifera* en sus promotores.**

Gen	Descripción de la proteína	Anotación GO	
YKL198C YKL198C	Proteína serina/treonina quinasa putativa	Proceso biológico	Actividad quinasa, transporte de poliamina, fosforilación de proteínas.
		Componente celular	Desconocido
		Función molecular	Actividad quinasa, actividad transferasa, unión de ATP, actividad proteína quinasa.
YPL145C YPL145C	Uno de los 7 miembros de la familia de proteínas de unión a oxisterol en levaduras	Proceso biológico	Proceso biosintético de esteroide, mantención de la polaridad celular, endocitosis.
		Componente celular	Citoplasma, membrana de Golgi.
		Función molecular	Unión a lípido, unión a fosfatidilinositol 4-fosfato, actividad de transportador de esteroles.
YPL144W YPL144W	Componente de chaperona heterodimérica Poc4p-Irc25p	Proceso biológico	Ensamblaje de complejo proteico mediado por chaperona.
		Componente celular	Citoplasma.
		Función molecular	Desconocido.
YCL038C YCL038C	Proteína integral de membrana vacuolar requerida para el eflujo de aminoácidos	Proceso biológico	Autofagia, transporte, transporte de aminoácidos.
		Componente celular	Vacuola, membrana de vacuola.
		Función molecular	Desconocido.
YJL029C YJL029C		Proceso biológico	Transporte, Transporte Golgi a vacuola, transporte retrógrado: endosoma a Golgi.

	Componente del complejo GARP (Golgi-associated retrograde protein)	Componente celular	Membrana de endosoma, aparato de Golgi, complejo GARP.
		Función molecular	Desconocido.
<b>YJL028W YJL028W</b>	Proteína de función desconocida	Proceso biológico	Desconocido.
		Componente celular	Componente integral de membrana.
		Función molecular	Desconocido.
<b>YAR014C YAR014C</b>	Proteína implicada en la selección de sitio de brote ( <i>bud-site selection</i> )	Proceso biológico	Regulación negativa de la polimerización de filamento de actina, organización de citoesqueleto, regulación de localización de proteína.
		Componente celular	Punta de brote celular, citoplasma, núcleo.
		Función molecular	Actividad de regulador de proteína fosfatasa tipo 1.

Los genes de *S. cerevisiae* que poseen el motivo 8 en sus regiones río arriba, codifican para genes de variadas funciones que incluyen vías metabólicas, proteínas quinasa y fosfatasa, proteínas de membrana y proteínas de función desconocida.

**Tabla 14: Genes de *S. cerevisiae* que poseen el motivo 15 del marcador de sexo de *B. papyrifera* en sus promotores.**

Gen	Descripción de la proteína	Anotación GO	
<b>YPL096W YPL096W</b>	Péptido N-glicanasa conservada requerida durante la degradación de proteínas en proteasoma	Proceso biológico	Proceso catabólico de proteínas mal plegadas o incompletamente sintetizadas, deglicosilación de proteínas.
		Componente celular	Citoplasma, mitocondria, núcleo.
		Función molecular	Actividad hidrolasa, unión a ión metálico, unión a DNA dañado.
<b>YKL022C YKL022C</b>	Subunidad del complejo promotor de anafase/ciclosoma	Proceso biológico	Ubiquitinación de proteínas, regulación negativa del inicio de la replicación DNA dependiente.
		Componente celular	Complejo promotor de anafase, núcleo, citosol.
		Función molecular	Desconocido.
<b>YIR026C YIR026C</b>	Proteína fosfatasa, implicada en crecimiento vegetativo a bajas temperaturas	Proceso biológico	Defosforilación de proteínas, división nuclear meiótica, ensamblaje de la subunidad ribosomal grande.
		Componente celular	Citoplasma, núcleo.
		Función molecular	Actividad hidrolasa, actividad fosfoproteína fosfatasa, actividad proteína tirosina fosfatasa.

<b>YOR293C-A YOR293C-A</b>	Proteína putativa de función desconocida	Proceso biológico	Desconocido.
		Componente celular	Desconocido.
		Función molecular	Desconocido.
<b>YGL254W YGL254W</b>	Factor de transcripción implicado en el metabolismo de sulfito	Proceso biológico	Regulación positiva del transporte de sulfito por regulación positiva desde promotor RNA pol II.
		Componente celular	Núcleo.
		Función molecular	Unión a DNA secuencia-específica al <i>core</i> del promotor de RNA pol II, unión a ión metálico.
<b>YMR238W YMR238W</b>	Manosidasa putativa implicada en crecimiento filamentosos	Proceso biológico	Actividad hidrolasa, actuando en enlaces glicosil, proceso catabólico de carbohidratos, biogénesis de pared celular tipo fungal.
		Componente celular	Membrana plasmática, retículo endoplásmico.
		Función molecular	Actividad manano endo-1-6 alfa manosidasa.
<b>YPR117W YPR117W</b>	Proteína putativa de función desconocida	Proceso biológico	Desconocido.
		Componente celular	Componente integral de membrana
		Función molecular	Desconocido
<b>YBR057C YBR057C</b>	Proteína esencial para la replicación meiótica de DNA y esporulación	Proceso biológico	Esporulación resultando en una espora celular, división nuclear meiótica, metilación de mRNA.
		Componente celular	Citoplasma, núcleo, nucléolo.
		Función molecular	Unión a proteína.
<b>YLR307C-A YLR307C-A</b>	Proteína putativa de función desconocida	Proceso biológico	Desconocido.
		Componente celular	Desconocido.
		Función molecular	Desconocido.
<b>YLR308W YLR308W</b>	Quitina deacetilasa, requerida para la rigidez apropiada en la pared de la ascospora	Proceso biológico	Proceso catabólico de polisacárido, esporulación resultando en la formación de espora celular, proceso catabólico de quitina.
		Componente celular	Pared de quitosán de la pared de la espora.
		Función molecular	Actividad hidrolasa, actividad quitina deacetilasa, actividad catalítica.

Los genes de *S. cerevisiae* que poseen el motivo 15 en sus regiones río arriba codifican para proteínas con funciones como marcaje de proteínas mal plegadas, proteínas con función desconocida y proteínas fosfatasa. Además se indican proteínas con función de factor de transcripción y otras esenciales en el proceso de replicación meiótica del DNA y participación en el complejo promotor de anafase que determina el avance en el ciclo celular.



**Tabla 15: Genes de *S. cerevisiae* que poseen una combinación de los motivos encontrados en el marcador de sexo de *B. papyrifera* en sus promotores.**

Gen	Descripción de la proteína	Anotación GO	
YGL023C YGL023C	Proteína de función desconocida	Proceso biológico	Desconocido
		Componente celular	Vacuola, membrana, mitocondria
		Función molecular	Desconocida
YHR169W YHR169W	RNA helicasa, ATPasa putativa de la familia de las DEAD-box	Proceso biológico	Ruptura endonucleolítica para generar extremo 5' madur en SSU-rRNA desde (SSU-rRNA 5.8S rRNA, LSU-rRNA), biogénesis de ribosoma
		Componente celular	90S preribosoma, nucleolo
		Función molecular	Actividad helicasa, unión a ATP, unión a nucleótido, unión a RNA.
YLR055C YLR055C	Subunidad del complejo regulatoria de transcripción SAGA	Proceso biológico	Acetilación de histonas, regulación de la transcripción, DNA dependiente, modificación de cromatina.
		Componente celular	Complejo SAGA, núcleo.
		Función molecular	Actividad de cofactor de transcripción, unión de proteína clase TBP.
YBR019C YBR019C	UDP-glucosa-4-epimerasa	Proceso biológico	Proceso metabólico, actividad isomerasa, proceso catabólico de galactosa.
		Componente celular	Citosol.
		Función molecular	Unión a coenzima, actividad aldosa 1-epimerasa.
YBR020W YBR020W	Galactoquinasa	Proceso biológico	Proceso catabólico de galactosa vía UDP-galactosa, proceso metabólico de galactosa.
		Componente celular	Citoplasma.
		Función molecular	Actividad quinasa, actividad galactoquinasa, unión a ATP.
YNL103W YNL103W	Activador transcripcional leucina- <i>zipper</i>	Proceso biológico	Regulación positiva de la transcripción desde promotor RNA pol II, respuesta a ión cadmio, proceso biosintético de metionina.
		Componente celular	Núcleo.
		Función molecular	Actividad de coactivador transcripcional, unión a DNA.
YER116C YER116C	Subunidad del complejo Slx5-Slx8 SUMO-dirigida ubiquitina ligasa	Proceso biológico	Respuesta a estímulo de daño a DNA, sumoilación de proteínas, ubiquitinación de proteínas.
		Componente celular	Citoplasma, núcleo.
		Función molecular	Unión a ión zinc, unión a ión metálico.
YHR061C YHR061C	Proteína implicada en la iniciación del <i>budding</i> y polarización celular	Proceso biológico	Organización del anillo de septina, establecimiento de polaridad celular, regulación de salida desde mitosis.

		Componente celular	Citoesqueleto, citoplasma, punta de proyección de <i>mating</i> .
		Función molecular	Unión de GTPasa Rho.
YPL249C-A YPL249C-A	Proteína L36B de la subunidad ribosomal 60S	Proceso biológico	Traducción, traducción citoplasmática.
		Componente celular	Ribosoma, complejo ribonucleoproteico.
		Función molecular	Unión a RNA, constituyente estructural del ribosoma.
YLR248W YLR248W	Proteína serina/treonina quinasa implicada en respuesta a estrés oxidativo y osmótico	Proceso biológico	Actividad quinasa, regulación de la meiosis, fosforilación, respuesta celular a estrés oxidativo.
		Componente celular	Citoplasma.
		Función molecular	Actividad quinasa, actividad transferasa, unión a calmodulina, unión a ATP, actividad proteína quinasa.

Los genes de *S. cerevisiae* que poseen una combinación de los motivos encontrados en el marcador de sexo de *B. papyrifera* codifican para proteínas con funciones como RNA helicasa que está implicada en el proceso de formación del ribosoma, miembros del complejo SAGA regulatorio de la transcripción, proteínas del ribosoma, factores de transcripción y proteínas que participan en vías metabólicas. Se repite el caso de los genes de *A. thaliana* cuyos promotores poseen distintas combinaciones de motivos encontrados en el marcador de sexo de *B. papyrifera*. Estos genes codifican proteínas esenciales en la regulación de la transcripción, del ciclo celular y en el desarrollo y supervivencia de la célula.

### 5.3.2.3 Búsqueda de factores de transcripción de *S. cerevisiae* que se unen a los motivos identificados en el marcador de sexo de *B. papyrifera*

Continuando con el objetivo 3, luego de la búsqueda de la funcionalidad de genes de *A. thaliana* y *S. cerevisiae* utilizando TAIR y Yeastmine de *Saccharomyces* Genome Database, respectivamente se utilizó el sitio YEASTRACT (<http://www.yeasttract.com>) para encontrar factores de transcripción de levadura que se unieran a los motivos 1, 8 y 15 identificados en el marcador de sexo de *B. papyrifera*. Para la búsqueda se permitieron 0, 1 y 2 sustituciones nucleotídicas en los sitios de unión registrados en bases de datos (evidencia experimental), lo que permite conocer si el motivo se conserva entre *S. cerevisiae* y *B. papyrifera*. Se muestran los factores de transcripción y sus sitios de unión en el motivo 1 (**Tabla 16**), del motivo 8 (**Tabla 18**) y del motivo 15 (**Tabla 20**) por ser estos los motivos destacados y con los que se hizo todo el análisis previo.

**Tabla 16: Factores de transcripción de *S. cerevisiae* y sitio de unión en el motivo 1 del marcador de sexo de *B. papyrifera*.**

0 sustitución		No se encontraron resultados					
1 sustitución	Sitio de unión	TF	Sitio de unión	TF	Sitio de unión	TF	
	TTGGRG	Adr1p	CCYWWWNRRG	Mcm1p	TCGAG	Xbp1p	
	YTGAT	Ash1p	ATGGAT	Mot3p			
	NCCDTYNVNCCGN	Cat8p Sip4p	TAGGAT	Mot3p			
	YCCNYTNRCCGN	Cat8p Sip4p	TMCCTT	Mot3p			
	GGAWG	Gcr1p	CCCCT	Msn2p Msn4p Nrg1p Rph1p			
	CCCCT	Gis1p YER130C	CCCTC	Nrg1p			
DCCYWWWNRRG	Mcm1p	SNCCG	Stb5p				
2 sustituciones	Sitio de unión	TF	Sitio de unión	TF	Sitio de unión	TF	
	TNCGTNNNNNTGAT	Abf1p	CGGNNNNNNNCCGG	Gsm1p	CCCCT	Msn2p Msn4p Nrg1p Rph1p	
	TTGGRG	Adr1p	CCGNNNNNNNCCG	Gsm1p	CCCTC	Nrg1p	
	YTGAT	Ash1p	GCTGG	Hac1p	GAGGG	Nrg1p	
	ATCAR	Ash1p	TTTTCHCG	Ime1p	CACGTK	Pho4p	
	TGACTC	Bas1p Gcn4p	TCCRNYGGA	Lys14p	MACGTG	Pho4p	
	NCCDTYNVNCCGN	Cat8p Sip4p	TCCRNYGGA	Lys14p	CTGGCA	Rim101p	
	NCGGNBNRAHGGN	Cat8p Sip4p	TTGCKCR	Mac1p	CGGANNA	Rgt1p	
	YCCNYTNRCCGN	Cat8p Sip4p	DCCYWWWNRRG	Mcm1p	TNNTCCG	Rgt1p	
	GAGCCC	Crz1p	CYNNWWWRGGH	Mcm1p	GTCAC	Rtg1p Rtg3p	
	GGGCTC	Crz1p	CCYWWWNRRG	Mcm1p	GGTAC	Rtg1p Rtg3p	
	HTHNGCTGD	Cup2p	CYNNWWWRGG	Mcm1p	GTACC	Rtg1p Rtg3p	
	HCAGCNDAD	Cup2p	CTGTGGC	Met31p Met32p	GGCYGGC	Skn7p	
	TRTKRY	Fkh1p Fkh2p	TACCTW	Mot3p	GGCTGGC	Skn7p	
	TGAATCA	Gcn4p	CCTCT	Mot3p	CGGNS	Stb5p	
	CTCC	Gcr1p	ATGGAT	Mot3p	SNCCG	Stb5p	
	GGAAG	Gcr1p	ATCCAT	Mot3p	CGGCTC	Stp1p Stp2p	
	CWTCC	Gcr1p	TAGGAT	Mot3p	GGAATG	Tec1p	
	GGAWG	Gcr1p	TMGGAA	Mot3p	CTCGA	Xbp1p	
	ATCCCTWA	Gis1p	TTCKA	Mot3p	TCGAG	Xbp1p	
CCCCT	Gis1p YER130C	TMCCTT	Mot3p	TTTGMA	Yap1p		
CGGNNNNNNNCCGG	Gsm1p	AAGGWT	Mot3p				
CCGNNNNNNNCCG	Gsm1p	AWCCTT	Mot3p				

**Tabla 17: Función de factores de transcripción de *S. cerevisiae* que se unen al motivo 1 del marcador de sexo de *B. papyrifera*.**

Factor de transcripción	Función	Tipo
<b>Mot3p</b>	Factor de transcripción que actúa globalmente. Modula el <i>mating</i> de levaduras, regula el metabolismo de carbono en respuesta a estrés, inhibe los genes de la vía del ergosterol. Tiene un rol en la mantención de la composición de la pared celular y también en la señalización de feromonas.	Represor, <i>zinc-finger</i> .
<b>Mcm1p</b>	Factor de transcripción que regula expresión de genes de metabolismo de arginina y activa la transcripción de genes de ciclo celular M/G1 y G2/M. También controla la expresión de genes que están involucrados en la formación del complejo de prerreplicación. Tiene un rol en la recombinación y osmotolerancia.	Activador, factor alfa hélice.

**Tabla 18: Factores de transcripción de *S. cerevisiae* y sitio de unión en el motivo 8 del marcador de sexo de *B. papyrifera*.**

0 sustitución	Sitio de unión		TF					
	SNCCG		Stb5p					
1 sustitución	Sitio de unión		TF		Sitio de unión		TF	
	NCCDTYNVNC CGN		Cat8p Sip4p		CCCCT	Msn2p Msn4p Nrg1p Rph1p	GTACC	Rtg1p Rtg3p
	YCCNYTNRRC CGN		Cat8p Sip4p		MACGTG	Pho4p	CGGNS	Stb5p
	CCCCT		Gis1p YER130C		GTCAC	Rtg1p Rtg3p	SNCCG	Stb5p
	AWCCTT		Mot3p		GTGAC	Rtg1p Rtg3p		
2 sustituciones	Sitio de unión		TF		Sitio de unión		TF	
	ATCANNNNNACGNNA		Abf1p		CCGNNNNNNNCCG	Gsm1p	CCCCT	Msn2p Msn4p Nrg1p Rph1p
	GCTGGT		Ace2p Swi5p		CCGNNNNNNNCCG	Gsm1p	AGGGG	Msn2p Msn4p Nrg1p Rph1p
	CYCAA		Adr1p		CAGCGTG	Hac1p	CCCTC	Nrg1p
	YRCACCCR		Aft2p Aft1p		CCAGC	Hac1p	CACGTK	Pho4p
	CGCACCC		Aft2p		GCTGG	Hac1p	MACGTG	Pho4p
	ATCAR		Ash1p		N TTCNNGAAN	Hsf1p	TGTTACCCGT	Reb1p
	NCCDTYNVNC CGN		Cat8p Sip4p		N TTCNNGAAN	Hsf1p	CGGANNA	Rgt1p
	NCGGNBNRAHGGN		Cat8p Sip4p		NGAANNTTCNNGAAN	Hsf1p	TNNTCCG	Rgt1p
	YCCNYTNRRC CGN		Cat8p Sip4p		CCGGNCCGG	Leu3p	GTCAC	Rtg1p Rtg3p
	CACGTGAY		Cbf1p		CCGGNCCGG	Leu3p	GTGAC	Rtg1p Rtg3p
	CAYGTGAY		Cbf1p		ACGCGT	Mbp1p	GGTAC	Rtg1p Rtg3p
	AACGTGA		Cbf1p		ACGCGT	Mbp1p	GTACC	Rtg1p Rtg3p
	GAGCCC		Crz1p		DCCYWWWNNRG	Mcm1p	GGCYGGC	Skn7p
	GGGCTC		Crz1p		CYNNWWWRRGH	Mcm1p	GGCCCAGA	Skn7p
	HCAGCNNDAD		Cup2p		CCYWWWNNRG	Mcm1p	GGCCGGC	Skn7p
	RYMAAYA		Fkh1p Fkh2p		CYNNWWWRRGG	Mcm1p	GGCCAGA	Skn7p
	CTTATC		Gat1p Gln3p Gzf3p		AAACTGTGG	Met31p Met32p	CGGNS	Stb5p
	GCACGTAG		Gcn4p		AAACTGTG	Met31p Met32p	SNCCG	Stb5p
	CACGTG		Gcn4p		CACGTGA	Met4p	CGGCTC	Stp1p Stp2p
	CACGTG		Gcn4p		TACCTW	Mot3p	GAGCCG	Stp1p Stp2p
	GGAAG		Gcr1p		CAGGYA	Mot3p	CNCACCNG	Stp2p
	CWTCC		Gcr1p		TRCCTG	Mot3p	CACGAAA	Swi4p
	GGAWG		Gcr1p		ATCCAT	Mot3p	AAAWNCCGCS	Uga3p
	AGGGG		Gis1p YER130C		TMGGAA	Mot3p	CTCGA	Xbp1p
	CCCCT		Gis1p YER130C		TMCTT	Mot3p		
	CTAATC		Gln3p		AWCCTT	Mot3p		

Se observa que sin sustitución nucleotídica, Stb5p es el único factor de transcripción de *S. cerevisiae* que se une al motivo 8 del marcador de sexo de *B. papyrifera*. Por tanto el motivo 8 es más conservado entre ambas especies que el motivo 1. Por otra parte, con 1 y con 2 sustituciones nucleotídicas, el número de factores de transcripción que se unen al motivo 8, Mot3p es el que aparece más recurrentemente en la lista, al igual que en el motivo 1. La función de estos factores de transcripción se describe en la **Tabla 19**.

**Tabla 19: Función de factores de transcripción de *S. cerevisiae* que se unen al motivo 8 del marcador de sexo de *B. papyrifera*.**

Factor de transcripción	Función	Tipo
<b>Mot3p</b>	Factor de transcripción que actúa globalmente. Modula el <i>mating</i> de levaduras, regula el metabolismo de carbono en respuesta a estrés, inhibe los genes de la vía del ergosterol. Tiene un rol en la mantención de la composición de la pared celular y también en la señalización de feromonas.	Represor, <i>zinc-finger</i> .
<b>Stb5p</b>	Stb5p es un factor de transcripción que activa genes involucrados en la vía de las pentosa-fosfato y producción de NADPH. Además es requerido para la resistencia a estrés provocado por agentes oxidantes. Se une a su propio promotor por tanto se sugiere que podría autorregular su expresión.	Activador y represor. C4 <i>zinc finger</i> .

**Tabla 20: Factores de transcripción de *S. cerevisiae* y sitio de unión en el motivo 15 del marcador de sexo de *B. papyrifera*.**

0 sustitución	Sitio de unión TF					
	YTGAT	Ash1p				
1 sustitución	Sitio de unión TF		Sitio de unión TF		Sitio de unión TF	
	TNNCGTNNNNNTGAT	Abf1p	TRTTKRY	Fkh1p Fkh2p	TCGAG	Xbp1p
	YTGAT	Ash1p	ATGGAT	Mot3p	WCCGYKKWW	Yrr1p
2 sustituciones	AVCGCANWTTTS	Dal81p Dal82p	CGGNS	Stb5p		
	Sitio de unión TF		Sitio de unión TF		Sitio de unión TF	
	TNNCGTNNNNNTGAT	Abf1p	CCGNNNNNNNCCG	Gsm1p	TTCKKA	Mot3p
	TTGGRG	Adr1p	CGGNNNNNNNCCG	Gsm1p	AAGGKA	Mot3p
	YTGAT	Ash1p	CCGNNNNNNNCCG	Gsm1p	TMCCTT	Mot3p
	ATCAR	Ash1p	GCTGG	Hac1p	AAGGWT	Mot3p
	TGACTC	Bas1p Gcn4p	GCTGG	Hac1p	AWCCTT	Mot3p
	NCGGNBNRAHGGN	Cat8p Sip4p	NGAANNTTCN	Hsf1p	AGGGG	Msn2p Msn4p Nrg1p Rph1p
	NCGGYNARNGGR	Cat8p Sip4p	NGAANNTTCN	Hsf1p	GAGGG	Nrg1p
	CAYGTGAY	Cbf1p	NTTCNNGAAN	Hsf1p	CACGTK	Pho4p
	GGGCTC	Crz1p	NTTCNNGAAN	Hsf1p	MACGTG	Pho4p
	AVCGCANWTTTS	Dal81p Dal82p	NGAANNTTCNNGAAN	Hsf1p	CGGANNA	Rgt1p
	TRTTKRY	Fkh1p Fkh2p	NTTCNNGAANNTTCN	Hsf1p	GTCAC	Rtg1p Rtg3p
	GATAAG	Gat1p Gln3p Gzf3p	WYTTCAYRTGS	Ino2p Ino4p	GTGAC	Rtg1p Rtg3p
	CTTATC	Gat1p Gln3p Gzf3p	DCCYWWNNNRG	Mcm1p	GGTAC	Rtg1p Rtg3p
	TGACTMT	Gcn4p	CCYWWNNNRG	Mcm1p	GTACC	Rtg1p Rtg3p
	CACGTG	Gcn4p	TCACGTG	Met4p	TGACGTTT	Sko1p
	CACGTG	Gcn4p	CCCCGCTTTAA	Mig1p Mig2p Mig3p	CGGNS	Stb5p
	CTTCC	Gcr1p	TACCTW	Mot3p	SNCCG	Stb5p
	GGAAG	Gcr1p	CAGGYA	Mot3p	GGAATG	Tec1p
CWTCC	Gcr1p	TRCCTG	Mot3p	RRGAATKY	Tec1p	
GGAWG	Gcr1p	CCTCTT	Mot3p	SGCGGNWTTT	Uga3p	
AGGGG	Gis1p YER130C	ATGGAT	Mot3p	CTCGA	Xbp1p	
CGGNNNNNNNCCG	Gsm1p	TAGGAT	Mot3p			

Con 0 sustitución nucleotídica se observa que solo Ash1p une al motivo 15, lo que sugiere que el sitio de unión de este factor en el motivo 15 es conservado entre *S. cerevisiae* y *B. papyrifera*. Con 1 y 2 sustituciones nucleotídicas aumenta el número de factores de transcripción de levadura que se unen al motivo 15. De todos los factores de transcripción de la lista, se verifica que nuevamente Mot3p es el factor que aparece más recurrentemente, seguido por Ash1p. Ambos factores de transcripción se describen en la **Tabla 21**.

**Tabla 21: Función de factores de transcripción de *S. cerevisiae* que se unen al motivo 15 del marcador de sexo de *B. papyrifera*.**

Factor de transcripción	Función	Tipo
<b>Mot3p</b>	Factor de transcripción que actúa globalmente. Modula el <i>mating</i> de levaduras, regula el metabolismo de carbono en respuesta a estrés, inhibe los genes de la vía del ergosterol. Tiene un rol en la mantención de la composición de la pared celular y también en la señalización de feromonas.	Represor, <i>zinc-finger</i> .
<b>Ash1p</b>	Ash1p es un factor de transcripción represor que está involucrado en la regulación del <i>mating</i> y el ciclo celular. Su objetivo mejor descrito es HO, cuyo producto génico es una endonucleasa relacionada con el <i>mating</i> de levadura. Se requiere para el desarrollo del fenotipo filamentoso de <i>Saccharomyces</i> .	Represor, C4.

Es importante notar que en los 3 motivos analizados (motivos 1, 8 y 15), Mot3p es el factor de transcripción que aparece mencionado más recurrentemente. Esto se debe a que posee un carácter de factor global de transcripción en *S. cerevisiae*. Este factor, junto con Ash1p, tiene un rol en la regulación del fenómeno de *mating* de levaduras, por tanto es interesante el análisis de mecanismos de regulación de expresión génica y desarrollo de fenotipos a partir de la acción de este factor de transcripción. Mcm1p y Ash1p son factores que regulan el avance del ciclo celular. Stb5p es en cambio un factor de transcripción involucrado en la regulación de genes de vías metabólicas. Por tanto se observa que los factores de transcripción que se unen a sitios de los motivos 1, motivo 8 y motivo 15 del marcador de sexo son factores globalmente relacionados con el *mating*, respuesta a estrés, biosíntesis y con la mantención y regulación del ciclo celular.

## 6 Discusión

En esta tesis se abordó el análisis genético-molecular y bioinformático de una región genómica de *B. papyrifera*, descrita como un marcador de sexo, en busca de características de elementos regulatorios de la expresión génica. El estudio de este marcador ayudará a comprender los mecanismos que permiten la diferenciación sexual de esta especie dioica. El estudio comparativo entre las secuencias de esta región proveniente de individuos masculinos y femeninos podría aportar conocimiento desde el punto de vista evolutivo acerca de cómo esta región se diferencia entre individuos de ambos sexos.

### 6.1 Búsqueda de homología con secuencias depositadas en GenBank.

Mediante la utilización de BLAST se comprobó que la secuencia del marcador de sexo de *B. papyrifera* no posee homología con secuencias depositadas en GenBank (NCBI), por lo tanto no es posible realizar una búsqueda directa o canónica por la vía de la comparación con secuencias conocidas en otras especies. Este resultado es interesante, pues la mayoría de las secuencias depositadas en las bases de datos corresponden a secuencias codificantes. Por lo tanto, es posible que este marcador corresponda a una región no codificante y se planteó la interrogante de si este marcador de sexo es una región con características de región regulatoria en *cis*, lo cual sería muy interesante desde el punto de vista de la regulación de la expresión génica. En este caso se esperaría encontrar secuencias del marcador de sexo diferentes en los individuos de distinto sexo y también la presencia de motivos correspondientes a sitios conservados de unión a factores de transcripción. Estos sitios de unión de factores de transcripción son conservados, y están descritos en especies modelo como *S. cerevisiae*. En consecuencia la identificación de motivos conservados en la región del marcador de sexo sugiere que podrían existir factores de transcripción homólogos que se unan a ellos, sugiriendo que esta región es efectivamente una región regulatoria.

Aparte de no encontrar secuencias homólogas utilizando GenBank, el marcador de sexo de *B. papyrifera* parece ser especie-específico, de acuerdo a los resultados de utilizar los partidores diseñados por Peñailillo [2014] en distintas especies de plantas incluso del mismo género. *B. kazinoki*, la otra especie del género *Broussonetia* incluida en este análisis, no es dioica. La única morácea dioica analizada fue *Cannabis sativa*, la que tampoco registró amplificación para el marcador de sexo [Payacán, resultados no publicados]. Estos resultados indican que la secuencia del marcador de sexo ha surgido independientemente en *B. papyrifera* y podría tener relación en cómo se originó o diferenció su estrategia



reproductiva, más si se considera al marcador de sexo como determinante de regulación de expresión génica y que las secuencias masculinas son diferentes a las femeninas.

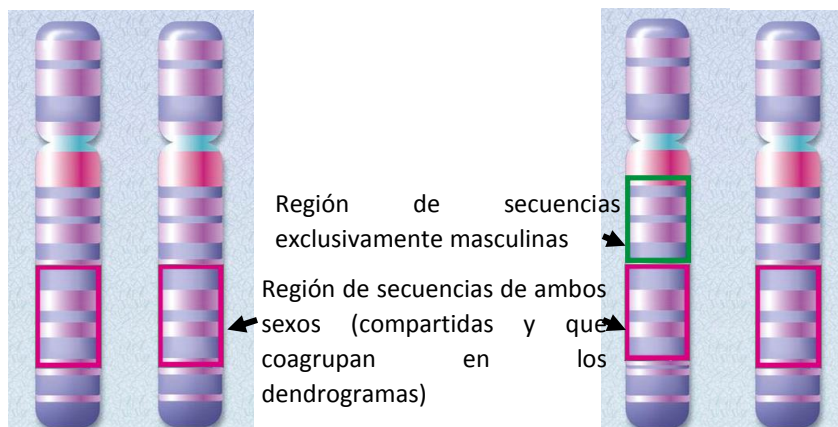
Sin embargo, la amplificación negativa en *C. sativa* cierra la posibilidad de postular al marcador de sexo como un marcador de carácter dioico. Es de interés entonces analizar el resto de especies dioicas del género *Broussonetia* en búsqueda de certificar la especie-especificidad del marcador o de postularlo como un marcador de carácter dioico acotado al género.

## **6.2 Variación en la secuencia del marcador según el sexo del individuo de *B. papyrifera*.**

Se clonaron los productos PCR de la región del marcador de sexo proveniente de 12 individuos de *B. papyrifera* (seis de cada sexo), con el fin de separar las distintas secuencias presentes en un mismo ejemplar. Ya en el contexto del objetivo específico 2, se encontró que existen versiones diferentes de esta región en un mismo individuo (**Figuras 12, 13 y 14**). La magnitud de las diferencias encontradas en las secuencias depende del sexo de la muestra, encontrándose mayor diversidad en las secuencias provenientes de individuos masculinos que de individuos femeninos. Este resultado se cuantificó mediante el alineamiento de las secuencias de cada una de las muestras y el cálculo del porcentaje de identidad (**Tabla 5**). Los dendrogramas obtenidos mostraron que algunas secuencias son compartidas y que presentan una distribución de clados diferente para individuos masculinos y femeninos.

En el dendrograma de la **Figura 15** se observó la formación de un clado de secuencias masculinas alejadas genéticamente del resto de las secuencias (determinado por la separación horizontal del resto de clados). En este grupo se encuentran secuencias de individuos de origen asiático {CH430\_M\_4, CH430\_M\_5 (China), TW137\_M\_2, TW138\_M\_1, TW138\_M\_2, TW138\_M\_6, TW138\_M\_8 (Taiwán), VT203\_M\_1, VT203\_M\_2, VT203\_3, VT203\_M\_4, VT203\_M\_5 (Vietnam)}. Este clado incluye también secuencias de individuos masculinos de morera de papel de Hawái (HW186\_M\_1, HW186\_M\_2, HW194\_M\_1). Este resultado concuerda con las observaciones de González **[2014]** respecto de que en Hawái hay morera de papel que podría ser de reciente introducción de origen asiático. Las poblaciones de *B. papyrifera* en la Polinesia son casi exclusivamente de sexo femenino; **[Seelenfreund y cols., 2010; Peñailillo, 2014]**. Recientemente (durante la redacción de esta memoria), en nuestro laboratorio, el análisis de la única muestra disponible proveniente de la remota isla de Rapa determinó que se trata de un individuo masculino **[Payacán, 2015]**, constituyendo la excepción respecto de la distribución por sexos encontrada en *B. papyrifera* en Polinesia, que excluye al archipiélago de Hawái.

Las diferencias observadas entre las secuencias en cada una de las muestras de *B. papyrifera* tanto en los alineamientos, como en los dendrogramas, revelan que los individuos masculinos poseen una secuencia adicional que está ausente en los individuos femeninos. Frente a este resultado surge la interrogante de cómo se generaron estas secuencias y qué función cumplen. Es posible que el marcador de sexo esté en un solo *locus* o en distintos *loci* en cromosomas específicos. En *B. papyrifera* no se ha descrito la presencia de cromosomas sexuales [Oginuma y Tobe, 1990]. Es importante notar que los cromosomas sexuales de *B. papyrifera*, en caso de poseerlos, serían del tipo homomórficos, pues en el análisis de cariotipo no se ha descrito un par de cromosomas sexuales distinguibles de los cromosomas autosómicos [Oginuma y Tobe, 1990]. Se ha descrito que en regiones que determinan características sexuales o los cromosomas sexuales sufren acumulación de mutaciones en este cromosoma, perdiendo la capacidad de recombinación [Charlesworth, 2002]. Es posible que las secuencias que agrupan en forma separada en los dendrogramas se encuentren en un *locus* específico. En la **Figura 20** se presenta un esquema hipotético de la organización de las distintas secuencias del marcador de sexo en *loci* diferentes en cromosomas sexuales homomórficos.



**Figura 20:** Esquema de los *loci* del marcador de sexo en los cromosomas sexuales de *B. papyrifera*.

- a) Par de cromosomas de individuo femenino. Se ilustra la región donde estarían ubicados los *loci* del marcador de sexo (región que es compartida con los cromosomas masculinos).
- b) Par de cromosomas de individuo masculino. Se ilustra la región donde estarían ubicados los *loci* del marcador de sexo, indicando la región exclusiva del cromosoma masculino, que podría haberse originado por pérdida de recombinación.

Una explicación alternativa es que los *loci* de las secuencias presentes exclusivamente en individuos masculinos estén sufriendo los efectos genómicos de la domesticación de *B. papyrifera*, tal como se ha descrito para *Prunus persica* [Verde y cols., 2013; Zheng y cols., 2014; Matthews,

**comunicación personal]**, donde, como producto del cultivo y la domesticación transgeneracional, se observa acumulación de regiones variables en los cromosomas relacionadas con funciones determinadas. Estas mutaciones generarían diferencias con respecto del mismo *locus* en variedades no domesticadas. Es posible que un fenómeno de esta naturaleza pudiese estar ocurriendo con *B. papyrifera* en su hábitat nativo.

### **6.3 Utilización de MEME para buscar motivos en la secuencia del marcador de sexo.**

Como se describió en esta tesis, el marcador de sexo de *B. papyrifera* no posee secuencias homólogas conocidas, por lo que se realizó una búsqueda alternativa de motivos con posible significancia biológica. La búsqueda de motivos utilizando MEME correspondió a una búsqueda no canónica (Objetivo específico 3) y predictiva de sitios putativos de unión a factores de transcripción en la que se identifican los motivos en función tanto de los parámetros que el usuario predetermina (ancho de motivo o ventana, y máximo de motivos a encontrar dentro de la secuencia), como del conjunto de secuencias entregadas.

Usando los parámetros descritos se identificaron 15 motivos con altos valores de coincidencia (valor *e*). El análisis posterior incluyó al motivo 1 por ser el mejor conservado. Adicionalmente, los motivos 8 y 15 no se repiten en algunas secuencias, encontrándose esta organización o composición de motivos exclusivamente en individuos masculinos (**Figura 19 a**). Esto se puede atribuir a diferencias de la secuencias en las regiones donde se ubican estos motivos. El posible origen de estas diferencias se discutió brevemente en la sección anterior, que corresponderían a una acumulación de mutaciones atribuidas a un proceso de domesticación y aislamiento geográfico.

### **6.4 Factores de transcripción de *S. cerevisiae* que se unen a lo motivos encontrados por MEME en la secuencia del marcador de sexo de *B. papyrifera*.**

Mediante la suite YEASTRACT y sus herramientas se llevó a cabo la búsqueda de factores de transcripción de *S. cerevisiae* que se unen específicamente a las secuencias de DNA homólogas a las de motivos encontrados en el marcador de sexo de *B. papyrifera*. El análisis de los motivos en busca de TFs específicos de *S. cerevisiae* se realizó permitiendo 0, 1 y 2 sustituciones nucleotídicas en los motivos estudiados. Este estudio indicó que los TFs Mot3p y Mcm1p son los que se unirían con mayor frecuencia al motivo 1, Mot3p y Stb5p al motivo 8, y Mot3p y Ash1p al motivo 15. Está muy bien descrito que estos

TFs regulan la expresión de genes participantes en diversas funciones, tales como modificación de ácidos nucleicos (endonucleasas), factores de transcripción, o participantes de vías metabólicas, entre muchas otras.

Mot3p ha sido descrito como un factor que actúa globalmente modulando el fenómeno de *mating* en levaduras, regulando el metabolismo de carbono en respuesta a estrés osmótico e inhibiendo los genes de la vía del ergosterol, entre otras actividades conocidas. Además, Mot3p tiene un rol en la mantención de la composición de la pared celular y también en la señalización de feromonas [Grishin y cols., 1998]. Su función en la respuesta a estrés osmótico es mediante la inhibición de la expresión de genes implicados en la biosíntesis de ergosterol, el análogo del colesterol en *S. cerevisiae*, el cual es esencial en la mantención de la fluidez y permeabilidad de la membrana en esta levadura [Montañés y cols., 2011]. Mot3p participa disminuyendo la señalización por estimulación de feromonas, probablemente mediante la inducción de genes represores [Grishin y cols., 1998]. La respuesta a las feromonas es clave en el *mating* de *S. cerevisiae*, pues las células responden o eligen a las células con mayor producción de feromonas para la conjugación [Jackson y Hartwell, 1990]. Es interesante la similitud del proceso de *mating* con lo que sucede en especies dioicas, pues uno de los factores no genéticos que promueve el desarrollo o prevalencia de los órganos florales de un sexo por sobre el otro, son las hormonas vegetales, particularmente las citoquininas y las giberelinas. Las giberelinas son hormonas vegetales o fitohormonas importantes, pues están estrechamente involucradas en el proceso de reproducción sexual de las plantas [Khryanin, 2002]. Es notable encontrar una conservación de la regulación del proceso de diferenciación sexual entre levaduras y plantas, a pesar de que la evolución de las estrategias reproductivas de las plantas es un proceso independiente de la familia a la que pertenece la especie [Ainsworth, 2000].

El marcador de sexo de *B. papyrifera* es una secuencia muy corta (aprox. 425 pb), en comparación con el tamaño aproximado de su genoma y con el tamaño promedio de las regiones regulatorias de la expresión génica en plantas [Maston y cols., 2006]. El estudio bioinformático predictivo realizado en esta memoria, es claramente insuficiente para entender el rol preciso que cumple la región del marcador de sexo en la regulación de la expresión génica en la determinación sexual en esta planta. Así la búsqueda de funciones regulatorias en que participan los factores de transcripción de levaduras es útil para inferir y extrapolar su funcionamiento para el caso de *B. papyrifera*, pues las bases de datos de *S. cerevisiae* usadas en esta memoria (YEASTRACT y YEASTMINE) son construidas en base a datos obtenidos experimentalmente, al contrario de lo que sucede en las bases de datos de plantas, que están apoyadas en resultados predictivos y todavía carentes de sustento experimental. Los análisis predictivos realizados

en MAST se basan en la utilización de bases de datos con regiones promotoras acotadas a un largo de no más de 1000 pb localizadas en la vecindad del sitio de inicio de transcripción, y tanto en *A. thaliana* como en *S. cerevisiae* existe una diversidad de genes que poseen los motivos identificados en el marcador de sexo de *B. papyrifera*, por tanto esto plantea la posibilidad de que esta secuencia corresponda a una región promotora central o central. Por el alcance de las bases de datos, en donde no se incluyen los elementos regulatorios distales, no se puede concluir ni descartar que la secuencia corresponda a este tipo de secuencia. Por lo tanto, el hecho de que los motivos, los factores de transcripción y las redes regulatorias sean conservadas, permitió sostener los resultados y conclusiones de esta memoria.

El aporte de esta memoria es analizar la secuencia del marcador de sexo de *B. papyrifera*, descubriendo que un producto PCR de 425 pb representa múltiples secuencias distribuidas de forma desconocida dentro de su genoma. Además se determinó que estas secuencias son distintas en individuos femeninos y masculinos. El análisis predictivo realizado con la suite MEME, sienta las bases para futuros estudios que tengan como objetivo dilucidar la funcionalidad del marcador de sexo como una región regulatoria de la expresión génica.

## **6.5 Proyecciones.**

Para comprender la función del marcador de sexo será necesario estudiar esta región en su contexto en el genoma de *B. papyrifera*. El análisis predictivo realizado en esta memoria con herramientas bioinformáticas, debe ser comprobado de forma experimental en el futuro utilizando técnicas que verifiquen la unión de factores de transcripción al marcador de sexo.

## 7 Conclusiones

- Los resultados obtenidos en la amplificación del marcador en otra especie sugieren que el marcador de sexo de *B. papyrifera* es especie-específico.
- Se identificó más de una secuencia en la región del marcador de sexo en cada individuo de *B. papyrifera* y algunas de estas secuencias son compartidas.
- Las secuencias femeninas presentan menor diversidad genética que las secuencias provenientes de individuos masculinos. Adicionalmente, algunas secuencias del marcador de sexo son exclusivamente masculinas, las cuales forman un clado separado.
- Los motivos encontrados en el marcador de sexo de *B. papyrifera* mediante un análisis no canónico están presentes en los promotores de genes de *A. thaliana* y de *S. cerevisiae*. Estos genes tienen participación en rutas metabólicas, en regulación de expresión génica, entre otras funciones. Los motivos son conservados y permiten proponer una función hipotética para estos motivos en *B. papyrifera*.
- El patrón de motivos en la secuencia del marcador de sexo de *B. papyrifera* varía entre las distintas secuencias. Si estos motivos representan sitios de unión de factores de transcripción en un contexto de elementos regulatorio en *cis*, estas diferencias podrían ser determinantes en la regulación diferencial de la expresión génica en individuos de distinto sexo.
- Esta tesis muestra una forma indirecta de buscar funcionalidad de elementos regulatorios de la expresión génica en una secuencia que posee la característica de ser una secuencia marcadora de sexo, de no poseer homólogos en otras especies, que se presenta en más de una copia en el genoma de *B. papyrifera*, y cuyas secuencias masculinas presentan composición de motivos distinta y de mayor diversidad genética que las secuencias femeninas.

## 8 Bibliografía

**Abukabar, M.** (2013). The Effectiveness of *Broussonetia* Mat in Slope Stabilization and Soil Fertility Improvement. *School of Graduate Studies, Kwame Nkrumah University of Science and Technology*.

**Ainsworth, C.** (2000). Boys and girls come out to play: the molecular biology of dioecious plants. *Annals of Botany*, 86, 211–221.

**Bachtrog, D., Mank, J., Peichel, C., Kirkpatrick, M., Otto, S., Ashman, T., Hahn M., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, Lau., Valenzuela, N., Vamosi, J., The Tree of Sex Consortium.** (2014). Sex determination: Why so many ways of doing it? *PLOS Biology*, 12 (7), 1-13.

**Bailey, T. & Elkan, C.** (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Technical Report, UCSD, Department of Computer Science and Engineering. CS94-351.

**Bailey, T. L., Williams, N., Mischak, C. & Li, W.** (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research*, 34 (2), W369-W373.

**Barker, C.** (2002). Plate 432. *Broussonetia papyrifera*. *Curtis's Botanical Magazine*, 19, 8–18.

**Blackwood, E. & Kadonaga, J.** (1998). Going the distance: a current view of enhancer action. *Science*, 281, 60-63.

**Charlesworth, D.** (2002). Plant sex determination. *Heredity*, 88, 94-101.

**Clark, T.** (1965). Plant fibers in the paper industry. *Economic Botany*, 19, 394–405.

**Clement, W. & Weiblen, G.** (2009). Morphological evolution in the mulberry family (Moraceae). *Systematic Botany*, 34(3), 530-552.

**Ewins, R.** (1987). Barkcloth and the origins of paper, in *Conference Papers of the First National Paper Conference*. Hobart: Papermakers of Australia, 1–4.

**González, J.** (2014). Análisis genético de la morera de papel, una planta de importancia cultural en Asia y la Polinesia, mediante el uso de marcadores moleculares anónimos ISSR. Tesis de pregrado. *Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile*.

**Grishin, A., Rothenberg, M., Downs, M., Blumer, K.** (1998). Mot3, a Zn finger transcription factor that modulates gene expression and attenuates mating pheromone signaling in *Saccharomyces cerevisiae*. *Genetics*, 149, 879-892.

**Guo, F., Feng, L., Huang, C., Ding, H., Zhang, X., Wang, Z. & Li, Y.** (2013). Prenylflavone derivatives from *Broussonetia papyrifera*, inhibit the growth of breast cancer cells *in vitro* and *in vivo*. *Phytochemistry Letters*, 6, 331–336.

**Guo, M., Wang, M., Deng, H., Zhang, X. & Wang, Z.** (2013). A novel anticancer agent Brousoflavonol B downregulates estrogen receptor (ER)- $\alpha$  expression and inhibits growth of ER-negative breast cancer MDA-MB-231 cells. *European journal of Pharmacology*, 714, 56–64.

**Hillis, D., Huelsenbeck, J. & Cunningham, C.** (1994). Application and accuracy of molecular phylogenies. *Science-AAAS-Weekly Paper Edition-including Guide to Scientific Information*, 264(5159), 671-676.

**Haudry, A., Platts, A., Vello, E., Hoen, D., Leclercq, M., Williamson, R., Forczek, E., Joly-Lopez, Z., Steffen, J., Hazzouri, K., Dewar K., Stinchcombe, J., Schoen, D., Wang, X., Schmutz, J., Town, C., Edger, P., Pires, J., Schumaker, K., Jarvis, D., Mándaková, T., Lysak, M., van der Berg, E., Schranz, M., Harrison, P., Moses, A., Bureau, T., Wright, S. & Blanchette, M.** (2013). An atlas of over 90.000 conserved non-coding sequences provides insight into crucifer regulatory regions. *Nature Genetics*, 45, 891-898.

**Huelsenbeck, J.** (1995). Performance of phylogenetic methods in simulation. *Systematic Biology*, 44(1), 17-48.



**Jackson, C. & Hartwell, L.** (1990). Courtship in *S. cerevisiae*: Both cell types choose mating partners by responding to the strongest pheromone signal. *Cell*, 63 (5), 1039-1051.

**Kejnovsky, E., Hobza, R., Cermak, T., Kubat, Z. & Vyskot, B.** (2009). The role of repetitive DNA in structure and evolution of sex chromosomes in plants. *Heredity*, 102(6), 533-541.

**Khryanin, V.** (2002). Role of phytohormones in sex differentiation in plants. *Russian Journal of Plant Physiology*, 49(4), 545-551.

**Kwak, W., Moon, T., Lin, C, Rhyn, H., Jung, H., Lee, E., Kwon, D., Son, K., Kim, H. P., Kang, S., Murakami, M., Kudo, I. & Chang, H.** (2003). Papyriflavonol A from *Broussonetia papyrifera* inhibits the passive cutaneous anaphylaxis reaction and has a secretory phospholipase A2-inhibitory activity. *Biological and Pharmaceutical Bulletin*, 26, 299–302.

**Lobos, S., Polanco, R., Tello, M., Cullen, D., Seelenfreund, D. & Vicuña, R.** (2011). Understanding LiP Promoters from *Phanerochaete chrysosporium*: A Bioinformatic Analysis. *Series: Book Chapter*.

**Lodhi, M., Ye, G., Weeden, N. & Reisch, B.** (1994). A simple and efficient method for DNA extraction from grapevine cultivars and *Vitis* species. *Plant Molecular Biology Reporter*, 12(1), 6-13.

**Manoj, P. & Ravichandran, N.** (2008). Development of sex specific molecular markers in dioecious *Piper longum*.L plants by differential display. *Journal of Theoretical and Applied Information Technology*, 4(5), 459-465

**Martinez, F., Pascual-Ahuir, A., Proft, M.** (2011). Repression of ergosterol biosynthesis is essential for stress resistance and is mediated by the Hog1 MAP kinase and the Mot3p and Rox1 transcription factors. *Molecular Microbiology*, 79(4), 1008-1023.

**Maston, G. A., Evans S. K., & Green, M. R.** (2006). Transcriptional regulatory elements in the human genome. *Annual. Reviews: Genomics and Human Genetics*, 7, 29-59.

**Matsumoto, J., Fujimoto, T., Takino, C., Saitoh, M., Hano, Y., Fukai, T. & Nomura, T. (1985).** Components of *Broussonetia papyrifera* (L.) Vent. I. Structures of two new isoprenylated flavonols and two chalcone derivatives. *Chemical & Pharmaceutical Bulletin*, 33, 3250–3256.

**Matthews, P. (1996).** Ethnobotany, and the origins of *Broussonetia papyrifera* in Polynesia: an essay on tapa Prehistory. *New Zealand Journal of Archaeology*, 117-132.

**Meudt, H. & Clarke, A. (2007).** Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends in Plant Science*, 12, 106–117.

**Milewicz, M. & Sawicki, J. (2012).** Mechanisms of Sex Determination in Plants. *Casopis slezského zemskeho muzea (A)*, 61(2), 123-129.

**Ogbourne, S. & Antalis, T. (1998).** Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochemical Journal*, 331, 1-14.

**Oginuma, K. & Tobe, H. (1995).** Karyomorphology of some Moraceae and Cecropiaceae (Urticales). *Journal of Plant Research*, 108(3), 313-326

**Oyanedel, N. (2010).** Análisis de la variabilidad genética de ejemplares de la especie vegetal *Broussonetia papyrifera* en la Polinesia: Evaluación mediante AFLP de un nuevo trazador de rutas migratorias humanas. Tesis de Magister. *Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile.*

**Parrish, T., Koelewijn H. & van Dijk P. (2004).** Identification of a male-specific AFLP marker in a functionally dioecious fig, *Ficus fulva* Reinw. ex Bl. (Moraceae). *Sexual Plant Reproduction*. 17(1), 17-22.

**Payacán, C.(2015).** Caracterización genética a parti de ADN antiguo de muestras de herbarios de *Broussonetia papyrifera* (L.) L'Herit. Ex. Vent. del Pacífico mediante marcadores moleculares. Tesis de Magister. *Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile.*

**Peñailillo, J.** (2014). Estudio de la diversidad genética de la morera de papel (*Broussonetia papyrifera* (L.) L'Herit.Ex Vent.) en el Pacífico, mediante un marcador molecular de sexo y microsatélites. Tesis de Magister. *Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile.*

**Prats, A., Roy, C., Wang, P., Erard, M., Housset, V., Gabus, C. & Darlix, J.** (1990). cis elements and trans-acting factors involved in dimer formation of murine leukemia virus RNA. *Journal of Virology*, 64(2), 774-783.

**Renner, S. & Ricklefs, R.** (1995). Dioecy and its correlates in the flowering plants. *American Journal of Botany*, 82(5), 596-606.

**Rustici, G. & Gardner, K.** (2006). Transcriptional Regulatory Networks from Yeasts to Humans. *Microarrays and Transcription Networks*, Chapter 3, 25-41.

**Sakai, R. & Aerts, J.** (2014). Sequence diversity diagram for comparative analysis of multiple sequence alignments. *BMC proceedings*, 8(2), S9.

**Schiffman, M, Bauer, H., Lorincz, A., Manos, M., Byrne, J., Glass, A. & Howley, P.** (1991). Comparison of Southern blot hybridization and polymerase chain reaction methods for the detection of human papillomavirus DNA. *Journal of Clinical Microbiology*, 29(3), 573-577.

**Schlötterer, C.** (2004). The evolution of molecular markers—just a matter of fashion? *Nature Reviews Genetics*, 5, 63–69.

**Seelenfreund D., Clarke A., Oyanedel N., Piña R., Lobos S., Matisoo-Smith, E. & Seelenfreund A.** (2010). Paper mulberry (*Broussonetia papyrifera*) as a commensal model for human mobility in Oceania: anthropological, botanical and genetics considerations. *New Zealand Journal of Botany*, 48, 231-247

**Sohn, H., Kwon, C. & Son, K.** (2010). Fungicidal effect of prenylated flavonol, Papyriflavonol A, isolated from *Broussonetia papyrifera* (L.) Vent. against *Candida albicans*. *Journal of Microbiology and Biotechnology*, 20, 1397–1402.

**Stehlik, I. & Blattner, F.** (2004). Sex-specific SCAR markers in the dioecious plant *Rumex nivalis* (Polygonaceae) and implications for the evolution of sex chromosomes. *Theoretical and Applied Genetics*, 108(2), 238-242.

**Stepanova, M., Tiazhelova, T., Skoblov, M. & Baranova, A.** (2005). A comparative analysis of relative occurrence of transcription factor binding sites in vertebrate genomes and gene promoter areas. *Bioinformatics*, 21 (9), 1789-1796.

**Tanaka, E., Bailey, T. L. & Keich, U.** (2014). Improving MEME via a two-tiered significance analysis. *Bioinformatics*, 30(14), 1965-1974.

**Verde, I., Abbott, A., Scalabrin, S., Jung, S., Shu, S., Marroni, F., Zhebentyayeva, T., Dettori, M., Grimwood, J., Cattonaro, F., Zuccolo, A., Rossini, L., Jenkins, J., Vendramin, E., Meisel, L., Decroocq, V., Sosinski, B., Prochnik, S., Mitros, T., Policriti, A., Cipriani, G., Dondini, L., Ficklin, S., Goodstein, D., Xuan, p., Del Fabro, C., Aramini, V., Copetti, D., Gonzalez, S., Horner, D., Falchi, R., Lucas, S., Mica, E., Maldonado, J., Lazzari, B., Bielenberg, D., Pirona, R., Miculan, M., Barakat, A., Testolin, R., Stella, A., Tartarini, S., Tonutti, P., Arús, P., Orellana, A., Wells, C., Main, D., Vizzotto, G., Silva, H., Salamini, F., Schmutz, J., Morgante, D. & Rokhsar, D.** (2013). The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature Genetics*, 45, 487-494.

**Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M. & Zabeau, M.** (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, 23(21), 4407-4414.

**Vyskot, B. & Hobza, R.** (2004). Gender in plants: sex chromosomes are emerging from the fog. *Trends in Genetics*, 20(9), 432-438.

**Wang, L., Changbo, D., Liu, D. & Liu, Q.** (2012). Identification of a male-specific amplified fragment length polymorphism (AFLP) marker in *Broussonetia papyrifera*. *African Journal of Biotechnology*, 11(33), 8196-8201.

**Wu, Y., Liu, C., Li, P., Wang, J., Xing, D. & Wang, B.** (2009). Photosynthetic characteristics involved in adaptability to Karst soil and alien invasion of paper mulberry (*Broussonetia papyrifera* (L.) Vent.) in comparison with mulberry (*Morus alba* L.). *Photosynthetica*, 47(1), 155-160.

**Zhang, M., Fang, Y., Liang, Z. & Huang, L.** (2012). Enhanced expression of vacuolar H<sup>+</sup>-ATPase subunit E in the roots is associated with the adaptation of *Broussonetia papyrifera* to salt stress. *PLOS one*, 7(10), e48183.

**Zheng, Y., Crawford, G. & Chen, X.** (2014). Archaeological evidence for Peach (*Prunus persica*) cultivation and domestication in China. *PLOS one*, 9 (9), e106595.

