



BENÉMERITA UNIVERSIDAD AUTÓNOMA DE PUEBLA



INSTITUTO DE CIENCIAS
CENTRO DE INVESTIGACIONES EN CIENCIAS MICROBIOLÓGICAS

POSGRADO EN MICROBIOLOGÍA

"ESTUDIO DE PROTEÍNAS SR PUTATIVAS EN *Ustilago maydis*"

TESIS

PRESENTADA PARA OBTENER EL TÍTULO DE:

MAESTRA EN CIENCIAS (MICROBIOLOGÍA)

PRESENTA:

QFB. MAYRA PATRICIA GASPARIANO CHOLULA

DIRECTORA DE TESIS:

D.C. REBECA D. MARTÍNEZ CONTRERAS

NOVIEMBRE 2016

ESTUDIO DE PROTEÍNAS SR PUTATIVAS EN *Ustilago maydis*



BUAP

Puebla, Pue. a 19 de octubre 2016.

**A LA ACADEMIA DEL POSGRADO
EN MICROBIOLOGÍA
CICM-ICUAP
PRESENTE.**

Por este conducto los abajo firmantes, integrantes del Comité revisor de Tesis de Maestría de la alumna **Mayra Patricia Gaspariano Cholula**, les informamos que hemos revisado el escrito titulado:

"Estudio de proteínas SR putativas en *Ustilago maydis*"

A nuestro juicio, la alumna **Mayra Patricia Gaspariano Cholula** puede proceder a la impresión de la tesis y a la presentación del examen de grado correspondiente.

Sin más que agregar, nos despedimos afectuosamente de ustedes.


Atentamente,
"Pensar Bien Para Vivir Mejor"




D.C. Nora Hilda Rosas Murrieta



D.C. Miguel Castañeda Lucio



D.C. Maricruz Anaya Ruiz



D.C. Nancy Martínez Montiel

Posgrado en Ciencias Microbiológicas
Instituto de Ciencias (ICUAP)

Edificio K11,
Ciudad Universitaria
Col. San Manuel, Puebla, Pue. C.P. 72570
Tel: (221) 229 55 00 Ext. 2632
posgrado.microbiologia@correo.buap.mx

Este trabajo de investigación fue realizado en el laboratorio de Ecología Molecular Microbiana dentro del grupo de investigación Biología Molecular y Celular de Hongos liderado por la D.C. Rebeca Debra Martínez Contreras Profesora e Investigadora Titular del Centro de Investigaciones en Ciencias Microbiológicas del Instituto de Ciencias de la Benemerita Universidad Autónoma de Puebla (CICM-ICUAP-BUAP).

La Q.F.B Mayra Patricia Gaspariano Cholula fue beneficiada con la beca número 619898 otorgada por el Consejo Nacional de Ciencia y Tecnología (CONACyT).

AGRADECIMIENTOS

A la Vicerrectoría de Investigación y Estudios de posgrado (VIEP), por el apoyo a través de programa Becas **Proyecto/VIEP** 2016.

A:

La Dr. Rebeca D. Martínez Contreras por su asesoría y apoyo durante la realización de este proyecto.

La Dr. Nancy Martínez Montiel por el tiempo y consejos que me brindo durante toda mi estancia en el laboratorio.

Mis sinodales y comisión revisora, por todos las observaciones y aportaciones realizadas durante el desarrollo del proyecto.

Mi familia:

A mis padres Germán y Patricia por amarme en primer lugar, porque gracias a ellos y a todo su esfuerzo me han permitido llegar hasta donde estoy el día de hoy; por siempre apoyarme en mis decisiones correctas o incorrectas, por comprenderme y apoyarme a pesar de sus dudas. Gracias por alimentarme y tenerme en su casa, aunque ya no de ni para los chicles.

A mi hermanita Fer por quedarse a mi lado tantas noches en vela y acompañarme a ver películas, series o lo que sea que hiciera ruido por la noche ; por escuchar mis problemas y teorías raras , pero mas que nada por apoyarme y amarme tal como soy.

A mi hermano Daniel por hacerme enojar, llevarme al cine, a comprar tacos y a cazar pokemones, eso me ayudo mucho a distraerme y pensar en otras cosas que no solo fueran respecto a la tesis jeje.

A mis sobrinitos Ximena y Mateo que siempre me hacen el día con sus ocurrencias, aunque en más de una ocasión solo me quitaron el tiempo.

A mis mascotas por estar ahí, aunque en realidad nunca me ayudaron a nada.

Este trabajo es dedicado principalmenete a ustedes.

Los Amo

Mis amigos:

Par este tema los voy a dividir en grupos, con el fin de no causar confusión (en mí).

Laboratorio:

A mi Jefecita y a Nancy, porque su ayuda y apoyo fue más que solo en cuestiones de trabajo sino de manera personal, gracias por todo el tiempo de convivencia, son grandes personas y se me ceren muchas cosas buenas, aunque Nancy tenga problemitas por aquello que se cayó de cabeza, gracias por sus jalones de orejas, porque de esos he aprendido mucho.

A Mo, por ser mi amiga desde el primer momento en que entré al posgrado, por escucharme, acompañarme e invitarme a comer a tu casa, por hacerme segunda en mis loqueras en el laboratorio, ayudarme en incontables ocasiones y tambien por mantenerme al tanto de las fechas importantes.

A Sonia porque a pesar de tener poco tiempo de conocernos, me ayudaste a resolver mis dudas y me apoyaste de una u otra manera durante todo este tiempo.

A Jenny por ser más Dory que yo, por apoyarme y ser una buena compañera en el laboratorio.

A Gaby, por apoyarnos y ayudarnos a lo largo de todo este tiempo, al parecer coincidimos en más cosas de las que pensé, gracias por escuchar y brindarme tu amistad.

Viejos amigos:

A Alina, Nancy, Fanny, Hilce y Adriana por todas las tardes de café sin y las platicas motivacionales, gracias por todo el apoyo, el cariño que me han brindando y su amistad incondicional.

A Vía, Tavo, Proo y Ricky, por su amistad y apoyo en estos 7 años que llevamos de conocernos, siempre me alentaron a seguir a delante.

Banda gangrena:

Lau, Toño y Pablo por compartir este lapso de tiempo, por ayudarme e invitar las chelas, por las clases que se hacían menos aburridas a su lado y por que me brindaron su amistad.

A todos ustedes Gracias.

INDICE

1. RESUMEN.
2. INTRODUCCIÓN.
 - 2.1 *Splicing* constitutivo y alternativo.
 - 2.2 El *Spliceosoma*.
 - 2.2.1 Promoción del *spliceosoma*.
 - 2.3 Factores centrales del *spliceosoma*.
 - 2.4 Factores auxiliares del *spliceosoma*.
 - 2.4.1 Proteínas SR.
 - 2.4.1.1 Mecanismos de las proteínas SR.
 - 2.5 *Splicing* en hongos.
 - 2.5.1 Proteínas SR en Modelos Fúngicos.
 - 2.6 *Ustilago maydis*.
 - 2.6.1 Ciclo de vida de *Ustilago maydis*.
3. JUSTIFICACIÓN.
4. OBJETIVOS.
 - 4.1 Objetivo General.
 - 4.2 Objetivos Particulares
5. METODOLOGÍA
 - 5.1 Cepas utilizadas, condiciones de cultivo y plásmidos.
 - 5.1.1 *Ustilago maydis* FB1.
 - 5.1.2 *E. coli* DH5 α .
 - 5.1.3 pJet 1.2 (ThermoFisher).
 - 5.1.4 pPROEX HT (c) (ThermoFisher).
 - 5.2 Medios de cultivo.
 - 5.2.1 YEPS (Yeast Extract, Peptone, Sugar).
 - 5.2.2 LB (Luria- Bertani; Sambrook *et al.*, 1989).
 - 5.2.3 SOC.
 - 5.2.4 GYT.
 - 5.3 Análisis bioinformático.
 - 5.3.1 Análisis *in silico* de homólogos para proteínas SR de humanos en el genoma de *Ustilago maydis*.
 - 5.3.2 Análisis *in silico* de proteínas SR putativas en *Ustilago maydis*.
 - 5.3.3 Análisis *in silico* de interacción con otras proteínas.
 - 5.3.4 Búsqueda de sitios blanco.
 - 5.4 Técnicas de Biología Molecular.
 - 5.4.1 Obtención de ADN genómico a partir de la cepa FB1 de *Ustilago maydis* (*Kit Wizard® Genomic DNA Purification, Promega*).
 - 5.4.2 Amplificación de fragmentos de ADN mediante PCR (*New England BioLabs*).
 - 5.4.3 Extracción de ADN plásmidico mediante Lisis alcalina (Green &

- Sambrook, 2012).
- 5.4.4 Electroforesis de ADN en geles de agarosa (Green & Sambrook, 2012).
 - 5.4.5 Clonación de los productos de PCR en el vector pJET 1.2 (CloneJET PCR Cloning Kit- Thermo Fisher Scientific).
 - 5.4.6 Subclonación de los productos digeridos con NotI-XbaI en el vector pPROEX-HTc (Addgen).
 - 5.4.7 Purificación de ADN (High Pure PCR Product Purification Kit- Roche).
 - 5.4.8 Análisis de restricción enzimática.
- 5.5 Técnicas de transferencia genética.
- 5.5.1 Preparación de células *E. coli* DH5 α electrocompetentes (Green y Sambrook, 2012).
 - 5.5.2 Electroporación de células *E. coli* DH5 α electrocompetentes (Green y Sambrook, 2012).
 - 5.5.3 Preparación de células *E. coli* DH5 α competentes (Green y Sambrook, 2012).
 - 5.5.4 Transformación de células *E. coli* DH5 α competentes (Green y Sambrook, 2012).
- 5.6 Técnicas de Proteínas recombinantes.
- 5.6.1 Inducción de proteínas recombinantes con IPTG (pPROEX-HT Procariotic Expression System).
 - 5.6.2 Electroforesis de proteínas en gel de poliacrilamida SDS (Miniprotein4 BioRad).
 - 5.6.3 Tinción de Geles de acrilamida con Coomassie.
- 5.7 Western Blot.
- 5.8 Ensayo de retardamiento
- 5.8.1 Marcaje de oligos con biotina
6. RESULTADOS.
- 6.1 Identificación *in silico* de los homólogos para las proteínas SR reportadas para humano, en el genoma de *Ustilago maydis*.
 - 6.2 Identificación de proteínas que contienen dominios RRM y RS característicos de las proteínas SR en la anotación de genoma de *Ustilago maydis*
 - 6.3 Comprobación de los genes codificantes para las proteínas uSRSF1, uRMB39 y uRNPS1 en *Ustilago maydis* y generación de sus respectivas proteínas recombinantes.
 - 6.3.1 Extracción de ADN y amplificación de fragmentos de ADN por medio de PCR.
 - 6.3.2 Clonación en el vector pJET 1.2.
 - 6.3.3 Subclonación en el vector pPROEX HTc.

- 6.3.4 Inducción de proteínas recombinantes (IPTG).
- 6.4 Identificación de secuencias de regulación putativas de las proteínas SR en el genoma de *Ustilago maydis*.
- 6.5 Análisis de interacción de uRNPS1 *in silico*.
- 7. DISCUSIÓN DE RESULTADOS.
 - 7.1 Identificación *in silico* de los homólogos para las proteínas SR reportadas para humano, en el genoma de *Ustilago maydis*.
 - 7.2 Identificación de proteínas que contienen dominios RRM y RS característicos de las proteínas SR en la anotación de genoma de *Ustilago maydis*.
 - 7.3 Comprobación de la presencia de los genes codificantes para las proteínas SR: SRSF1, RMB39 y RNPS1 en *Ustilago maydis* y generación de la proteína recombinante uRNPS1.
 - 7.4 Búsqueda de las secuencias de regulación para proteínas SR en el genoma de *Ustilago maydis*.
 - 7.5 Interacción de uRNPS1 con otras proteínas.
- 8. CONCLUSIONES
- 9. PERSPECTIVAS.
- 10. BIBLIOGRAFÍA.

INDICE DE FIGURAS.

Figura 1. Principales eventos de *splicing* alternativo.

Figura 2. Formación del *spliceosoma*.

Figura 3. Elementos en *cis* reguladores del *splicing*.

Figura 4. Composición de proteínas y estructura secundaria de snRNA de los principales snRNPs de humanos.

Figura 5. Dominios estructurales y funcionales característicos de las proteínas SR.

Figura 6. Funciones de proteínas SR en el proceso de *splicing*.

Figura 7. Ciclo de vida de *Ustilago maydis*.

Figura 8. Comparación de secuencias de proteínas predichas en *Ustilago maydis*, *Saccharomyces cerevisiae* y *Homo sapiens*.

Figura 9. Productos amplificados por PCR.

Figura 10. Digestión de la construcción pJET-uSRSF1 con la enzima BglI.

Figura 11 Digestión de la construcción pJET-uRMB39 con la enzima Styl.

Figura 12. Digestión de la construcción pJET-uRNPS1 con la enzima HindIII

Figura 13 Análisis de restricción de pPROc-uSRSF1 con la enzima SacI.

Figura 14 Análisis de restricción de pPROc-uRBM39 con la enzima Sall.

Figura 15 Análisis de restricción de pPROc-uRNPS1 con la enzima BamHI.

Figura 16. Identificación en SDS PAGE de la proteína recombinante uSRSF1.

Figura 17. Identificación en SDS PAGE de la proteína recombinante uRBM39.

Figura 18. Identificación en SDS PAGE de la proteína recombinante uRNPS1.

Figura 19. Western blot de la proteína recombinante uRNPS1.

Figura 20. Frecuencia de genes que muestran secuencias de pegado para RNPS1 en el genoma de *Ustilago maydis*.

Figura 21. Interacción de RNPS1 con diferentes proteínas.

INDICE DE TABLAS

Tabla 1. Nomenclatura actual de las proteínas SR

Tabla 2. Diversidad de funciones de proteínas SR.

Tabla 3. Oligonucleótidos empleados para obtener las proteínas recombinantes de *U. maydis*.

Tabla 4. Condiciones para las reacciones de PCR

Tabla 5. Componentes y cantidades para las reacciones de ligación.

Tabla 6. Componentes y cantidades para las reacciones de ligación

Tabla 7. Proteínas SR humanas y sus características generales.

Tabla 8. Secuencias homólogas para proteínas SR de humanos en *Ustilago maydis*.

Tabla 9. Proteínas con dominios RRM y RS identificadas en la base de datos de *Ustilago maydis*.

Tabla 10. Proteínas SR putativas en *Ustilago maydis*.

Tabla 11. Sitios de unión reportados para SRSF1 y RNPS1.

Tabla 12. Características de las proteínas que interactúan con uRNPS1.

11. RESUMEN.

Las proteínas SR participan en el *splicing* constitutivo y alternativo del pre-ARNm así como en diversas modificaciones posteriores al *splicing*, incluyendo la exportación de ARNm nuclear, el proceso de NMD y la traducción del ARNm. Las múltiples actividades que pueden realizar las proteínas SR ponen de manifiesto la importancia de las proteínas SR durante la regulación del metabolismo del ARNm. Estructuralmente, una proteína SR presenta uno o dos dominios RRM en el extremo N-terminal, seguido río abajo por un dominio RS de al menos 50 aminoácidos, el cual se caracteriza por un contenido mayor al 40 % en residuos de serina-arginina.

Este estudio se basó en el análisis de proteínas SR en *Ustilago maydis* por medio de dos abordajes, el primero que consistió en la búsqueda de homólogos para proteínas SR de humano y el segundo en identificar todas aquellas proteínas que tuvieran en su secuencia los dominios RRM que caracterizan a las proteínas SR. Se identificó un total de 76 posibles proteínas SR en *U. maydis*, de las cuales 45 resultaron ser homólogas para proteínas SR de *Homo sapiens*, presentando una identidad que oscila entre el 20 y el 60 %. Estos resultados se obtuvieron por medio de la búsqueda de proteínas SR en la literatura y, empleando el Blast proteína- proteína en la plataforma del NCBI. Por otro lado, el segundo abordaje se llevó a cabo de tal manera que, primero se realizó la búsqueda de todas aquellas proteínas que tuvieran dominios RRM dentro del genoma de *U. maydis* con ayuda de la base datos de Pfam, y posteriormente se analizaron las secuencias aminoacídicas para identificar las regiones ricas en residuos de arginina y serina; solo 18 de las proteínas SR putativas de *U. maydis* presentaban los repetidos RS canónicos y el resto presentaron regiones ricas en R o S de manera separada, la principal función de estos residuos es ser blanco de fosforilaciones y promover la interacción proteína- proteína. Después de la identificación inicial, se seleccionaron 3 homólogos para la obtención de las proteínas recombinantes correspondientes: uRMB39 (90 kDa), uSRSF1 (170 kDa) y uRNPS1 (32 kDa), la obtención de las recombinantes se realizó clonando los productos correspondientes a cada marco de lectura abierta en el vector de reservorio pJET 1.2, seguido de una subclonación en el vector de expresión pPROEX THc. Una vez obtenidas las recombinantes se eligió a uRNPS1, se conoce que la organización de los dominios y los procesos en los cuales se encuentra implicada la proteína RNPS1 se conservan a través de todos los reinos (Califice *et al.*, 2012), para identificar sus posibles secuencias de unión a ARN en el genoma de *U. maydis* empleando el programa GeneFinder, se encontró que 356 genes (5.24 %) poseen el sitio de unión para alguna de las secuencia blanco reportadas para este factor, siendo la secuencia GAAGAAGAA la más prevalente encontrada en 217 genes (3.19 %), sin embargo, debido a esta secuencia posee el motivo GAAG elemento en *cis* muy importante involucrado en la regulación del *splicing alternativo* (Perteau *et al.*, 2007), podrían indicar solo la especificidad que presentan de manera general las proteínas SR pero o por la proteína uRNPS1, por otro lado las secuencias TCTCTCTC y CTTCTTCTC las cuales fueron encontradas en 45 y 78 genes respectivamente se consideran son los blancos específicos para la proteína uRNPS1. Con el fin de integrar todos estos resultados se realizó un análisis *in silico* de las interacciones proteicas predichas presentadas por uRNPS1; realizado con la herramienta STRING se observó que varias de las proteínas con las que interactúa se encuentran asociadas a procesos de *splicing*, forman parte del complejo EJC y algunas

otras corresponden a proteínas SR. Con todos estos datos podemos sugerir que las proteínas SR se encuentran conservadas en *U. maydis*, que probablemente se encuentren llevando a cabo las mismas funciones que RNPS1 de *H. sapiens* y por lo tanto, eventos como: *splicing*, NMD y la exportación de ARNm se encuentran se conservan en el hongo, sin embargo, es necesaria la demostración experimental de cada una de estas conclusiones.

12. INTRODUCCIÓN.

Los genes de organismos eucariotas se encuentran frecuentemente interrumpidos por secuencias no codificantes que deben ser eliminadas durante la expresión, las cuales reciben el nombre de intrones. El *splicing* es el proceso mediante el cual los intrones son removidos, para producir moléculas de ARN mensajero (ARNm) funcionales al ligarse las secuencias codificantes o exones flanqueantes. El *splicing* alternativo ocurre cuando algunos exones, intrones o porciones de los mismos son incluidos de manera diferencial para producir diversas moléculas de ARNm a partir de un mismo precursor inmaduro (pre-ARNm). El *splicing* alternativo es empleado por el 75 % de los genes humanos y constituye la principal fuente de diversidad proteica en eucariotas. Aunque se estima que entre el 15 y el 50 % de las enfermedades genéticas humanas tienen su origen en algún defecto a nivel de *splicing* (Tazi *et al.*, 2008), la extraordinaria contribución del *splicing* alternativo en diferentes procesos biológicos y patológicos en plantas y mamíferos apenas empieza a dilucidarse. A continuación, se presentan los conceptos relacionados con el proceso y la regulación de este mecanismo.

12.1 *Splicing* constitutivo y alternativo.

El concepto simple de un gen como una unidad continua y definida de ADN que da lugar a un único polipéptido se revisó radicalmente por el descubrimiento reportado en 1977 correspondiente a la existencia de "genes interrumpidos o divididos", ya que al hibridizar algunos fragmentos de ADN con el ARNm respectivo para adenovirus 2, se demostró que algunos fragmentos en el ADN no podían aparearse con su ARNm, formando horquillas en algunas regiones (Berget *et al.*, 1977). A partir de este descubrimiento inicial, se empezaron a realizar diversos estudios para identificar tanto el mecanismo responsable de madurar al ARNm, así como la maquinaria que cataliza dicho proceso.

Actualmente entendemos que la mayoría de los genes humanos que codifican para proteínas son sin duda unidades discontinuas que contienen la información que codifica para la proteína en las regiones llamadas exones, los cuales se encuentran interrumpidos y están flanqueadas por secuencias denominadas intrones. De esta manera, el *splicing* del pre-ARNm consiste en eliminar los intrones y así poder empalmar los exones generando un ARNm maduro (Van der Feltz *et al.*, 2012), a este tipo de *splicing* se conoce como *splicing* constitutivo.

La mayoría de los genes que codifican para proteínas humanas (>90 %) contienen intrones y el promedio de la longitud de exones e intrones en los genes humanos es de 145 y 3365 nucleótidos, respectivamente (International Human Genome Sequencing Consortium, 2004).

Por otro lado existe otro tipo de *splicing*, el cual es llamado *splicing* alternativo y es comúnmente utilizado por los metazoarios para generar más de una proteína a partir de un gen (Marshall *et al.*, 2013) y por lo tanto es importante destacar que las estimaciones actuales son que más del 75 % de los pre-ARNm humanos se someten a *splicing* alternativo para producir diferentes isoformas de las

proteínas, aumentando así significativamente la diversidad proteica (Modrek & Lee, 2002) formando parte de la regulación post-transcripcional de la expresión de genes (Harrow *et al.*, 2006).

Se han descrito cinco tipos principales de *splicing* alternativo, los cuales se muestran en la figura 1: sitios de *splicing* 5' alternativos (A), sitios de *splicing* 3' alternativos (B) la inclusión o la exclusión de un exón completo (C); también se puede producir la retención de un intrón (D) y por último el caso de exones mutuamente excluyentes (E), en donde solo uno de los dos exones se encontrará presente en el ARNm maduro final (Sorek *et al.*, 2004).

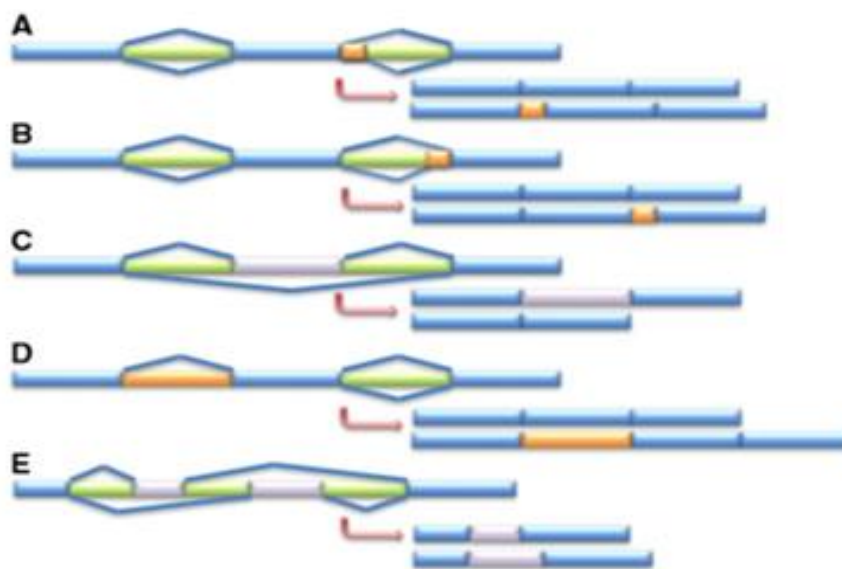


Figura 1. Principales eventos de *splicing* alternativo. En el esquema se observan los exones como recuadros de color azul; los intrones como recuadros de color verde; los recuadros de color naranja son los intrones alternativos retenidos y los recuadros de color rosa son los exones que han sufrido exclusión. El producto obtenido mediante *splicing* se muestra para cada evento (Modificado de Kempken, 2013).

12.2 El Spliceosoma.

El proceso de eliminación de intrones es catalizado por el *spliceosoma*, el cual consiste en un conjunto de cinco partículas pequeñas nucleares ribonucleoproteicas (snRNP): U1, U2, U4, U5 y U6 que a su vez están asociadas a un gran número de proteínas accesorias. Estas proteínas van a reconocer secuencias específicas presentes en el pre-ARNm, de tal forma que la regulación del *splicing* implica un proceso de regulación fina en el cual los elementos en *cis* van a ser reconocidos por factores en *trans* para cada evento de *splicing* (Matlin *et al.*, 2005).

En eucariotas, se han descrito dos tipos de *spliceosoma*: el *spliceosoma* dependiente de U2 y el *spliceosoma* dependiente de U12, que está presente en sólo un subconjunto de organismos (Patel *et al.*, 2003).

12.2.1 Promoción del *spliceosoma*.

Durante el splicing, se llevan a cabo una serie de eventos secuenciales que conducen a catalizar la remoción de las secuencias no codificantes y la ligación de las regiones flanqueantes para formar un ARNm maduro que pueda ser traducido. Dichos eventos se resumen a continuación y se ilustran en la figura 2.

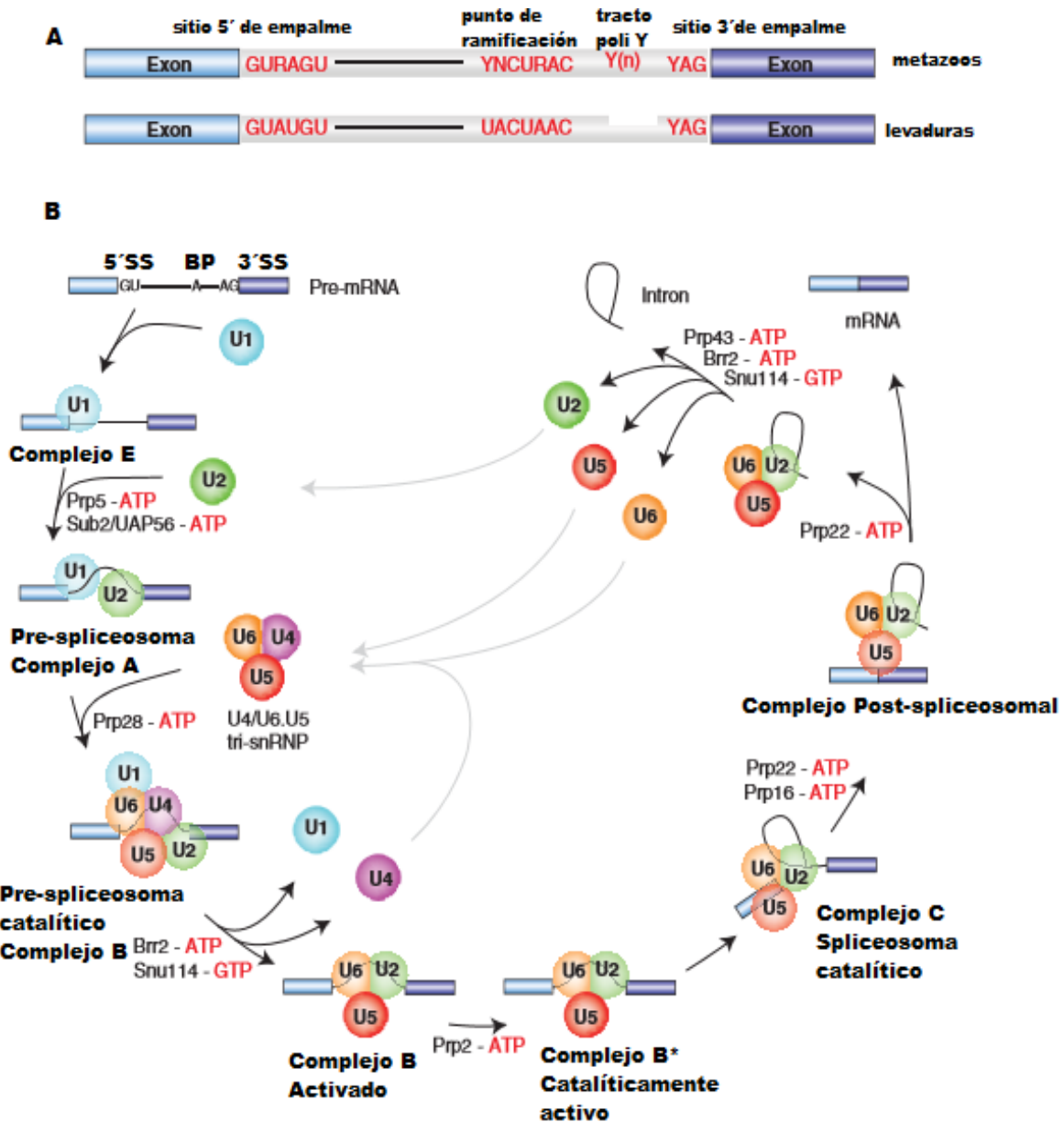


Figura 2. Formación del *spliceosoma*. A) Esquema que presenta los elementos en *cis* de un pre-ARNm de metazoos y levaduras. B) Esquema que muestra los diferentes complejos generados en la formación del *spliceosoma*, se observa en que complejo se encuentra cada U snRNP y la intervención de otras proteínas y moléculas de manera general (Modificado de Lührmann *et al.*, 2011).

- Inicialmente, se forma el complejo E donde U1 snRNP reconoce al dímero GU en el sitio 5' de empalme y la U2 snRNP se une al dímero AG en el sitio de empalme 3'.
- Posteriormente se forma el complejo A cuando la snRNP U2 se reacomoda para reconocer el punto de ramificación.
- Un complejo tri-snRNP que consta de las snRNPs U4/U6-U5 se recluta para formar el complejo B.
- Después de mucha reorganización estructural y de composición, la snRNP U1 se sustituye por la snRNP U6 en la región del sitio 5' de empalme y hay un cambio a un complejo B* activado que puede realizar la primera de las dos reacciones catalíticas.
- Un *spliceosoma* preparado para el segundo paso de la catálisis forma el complejo C. En este paso el *spliceosoma* requiere ATP para catalizar reordenamientos estructurales que son críticos para la progresión a través de todas estas etapas del ensamble.
- Por último, el intrón se remueve en forma de laso junto con las snRNPs 2, 5 y 6, las cuales se disocian para posteriormente ser reutilizadas en otro evento de *splicing*. (Lührmann *et al.*, 2011).

12.3 Factores centrales del *spliceosoma*.

Dentro de la secuencia del pre-ARNm existen varios sitios que son reconocidos por el *spliceosoma* y que se conocen como elementos en *cis* (Fig. 3). Los sitios más conservados son los dinucleótidos GU y AG en los límites de la unión exón-intrón y que se conocen también como sitios de *splicing* 5' y 3' respectivamente. El punto de ramificación (*branchpoint*) es una secuencia degenerada que de manera general posee una adenina conservada. Entre el punto de ramificación y el sitio del dinucleótido AG se encuentra una región rica en pirimidinas, los cuales son usualmente uracilos (Jiménez García *et al.*, 2007).

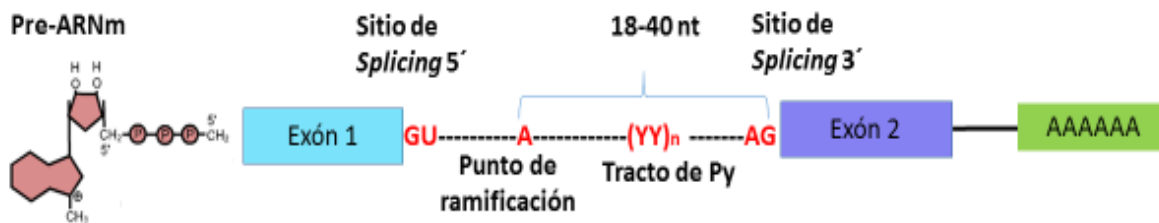


Figura 3. Elementos en *cis* reguladores del *splicing*. En color azul claro y oscuro se muestran los exones y en letras de color rojo están representadas las secuencias consenso que se encuentran dentro del intrón y por tal razón lo delimitan.

Los factores en *trans* corresponden a las proteínas que reconocerán a los elementos en *cis* y que incluyen principalmente a las proteínas snRNPs. Otros factores en *trans* corresponden a las proteínas hnRNP, las cuales poseen la capacidad de silenciar el *splicing* al reconocer secuencias silenciadoras de *splicing* exónicas e intrónicas (ESS, ISS, por sus siglas en inglés). Otra familia de factores reguladores está constituida por las proteínas SR, las cuales de manera general facilitan el *splicing* al reconocer secuencias potenciadoras localizadas también tanto en los exones como en los intrones (ESE, ISE, por sus siglas en inglés) (Matlin *et al.*, 2005).

Los pequeños ARN nucleares cuyas longitudes oscilan entre 60 y 300 nucleótidos presentan la característica de ser ricos en residuos de uracilo (UsnRNAs). Alrededor de estas partículas de ARN, el *spliceosoma* dependiente de U2 se ensambla a partir de las snRNPs U1, U2, U5 y U4/U6 (Fig. 4).

En contraste, las principales subunidades del *spliceosoma* dependiente de U12 son U11, U12, U5, y U4atac/U6atac snRNPs (Patel *et al.*, 2003). Estos ARNs son transcritos por la RNA polimerasa II, a excepción del snRNA U6, que lo transcribe la RNA polimerasa III. En el núcleo, los UsnRNAs se asocian con proteínas para formar ribonucleoproteínas pequeñas nucleares (UsnRNPs); una vez que los snRNAs U1, U2, U4, U5 son transportados al citoplasma, siete proteínas Sm se ensamblan alrededor del snRNA para formar el núcleo de las snRNPs U1, U2, U4, y U5. Las proteínas Sm reconocen una secuencia de ARN de una sola hebra corta en los snRNAs U1, U2, U4 y U5 denominadas sitio Sm, que se ajusta a un consenso AU y secuencias ricas en repetidos de G. Este sitio Sm está situado entre dos estructuras de tallo-bucle en cada uno de estos U snRNA, una característica que puede ser crítica para su asociación estable (Van der Feltz *et al.*, 2012). Las siete proteínas Sm (Sm B/B', D₁, D₂, D₃, -E, -F y -G) son fundamentales para el ensamble, transporte, y la integridad de la U snRNPs. Cada proteína Sm contiene un motivo conservado Sm compuesto de dos segmentos de secuencia primaria cortos denominados Sm1 y Sm2, separadas por una región adaptadora (Will & Lüthmann, 2011).

12.4 Factores auxiliares del *spliceosoma*.

Además de las snRNPs, los principales factores reguladores del *splicing* son las proteínas SR y la familia de las proteínas hnRNP, cada una con secuencias blanco específicas. Actualmente se sabe que la expresión diferencial, así como los cambios en la concentración nuclear de dichos factores pueden ser responsables de modificar las decisiones que ocurren durante el *splicing* dependiendo del tipo celular o del nivel de diferenciación. Por si fuera poco, la expresión de algunos factores de *splicing* puede estar regulada a su vez por eventos de *splicing* alternativo; mientras que la funcionalidad, localización y actividad de varios factores puede regularse también por su estado de fosforilación. De manera general, las partículas ribonucleoproteicas nucleares heterogéneas (hnRNPs) que contienen RNA nuclear heterogéneo (hnRNA) son estabilizadoras; es decir, su función radica en asociarse al pre-ARNm para impedir la formación de estructuras secundarias dadas por el apareamiento de bases. De ésta manera el pre-ARNm permanece accesible para interactuar con otras moléculas.

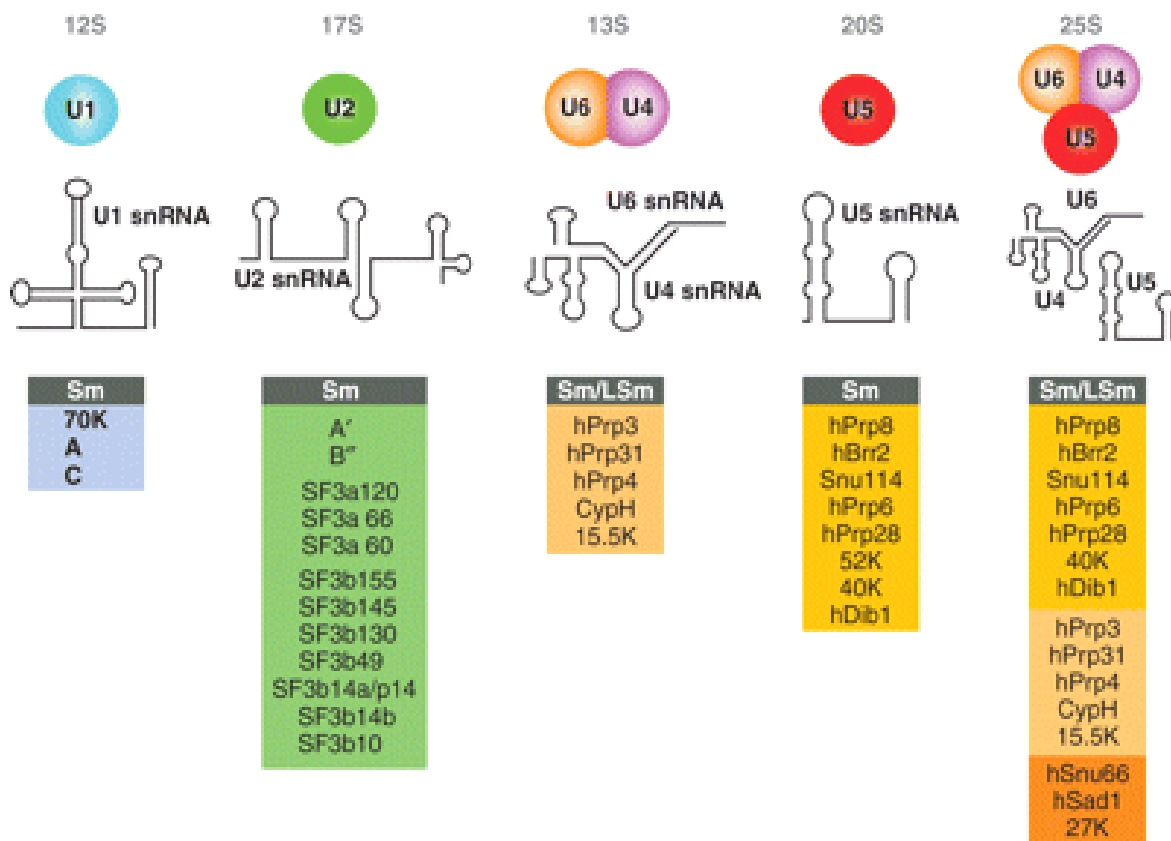


Figura 4. Composición de proteínas y estructura secundaria de snRNA de los principales snRNPs de humanos. Se muestra la estructura secundaria de los ARNs que forman las cinco snRNPs y sus diferentes asociaciones; en la parte de abajo se observan las principales proteínas Sm o proteínas LSm/Lsm que se asocian respectivamente (Will & Lührmann, 2011).

Por otro lado, las proteínas SR (proteínas con dominios ricos en Serina y Arginina) representan una familia de factores de *splicing* que se reporta en vertebrados, invertebrados y plantas. La secuencia de aminoácidos en estas proteínas es rica en residuos de serina (S) y arginina (R) en el extremo carboxilo, formando el dominio SR. Estas proteínas tienen varios papeles reguladores en el *splicing* como son la definición de los sitios de *splicing* y el acercamiento de las regiones de los dinucleótidos. A continuación, profundizaremos en la estructura y función de las proteínas SR.

12.4.1 Proteínas SR.

Definimos una proteína SR como cualquier proteína con los siguientes atributos (Fig. 5): la presencia de uno o dos RRM en el extremo N-terminal, seguido por un dominio RS río abajo de al menos 50 aminoácidos con >40 % de contenido de RS y caracterizado por repetidos de serina-arginina (Long *et al.*, 2009). De acuerdo en un estudio, utilizando la base de datos Uniprot, se encontraron 160 proteínas humanas con dominios RRM; de las cuales 107 proteínas poseen sólo un RRM y 53 proteínas poseen dos RRM.



Figura 5. Dominios estructurales y funcionales característicos de las proteínas SR. En este esquema se muestran los dominios estructurales y funcionales que poseen 9 proteínas SR, en recuadros morados y azules se indican los dominios de reconocimiento de ARN: RRM y RRMH, en recuadros verdes los dominios RS (repetidos de arginina/serina) y en un recuadro azul de menor tamaño el dominio de dedos de Zinc (Shepard *et al.*, 2009).

Actualmente se utiliza una nueva nomenclatura propuesta por Manley y Krainer, se basa en utilizar las letras “SRSF” (SR *splicing* factor) seguido de los números 1-12, esta numeración refleja el orden cronológico en el cual los genes o proteínas fueron descubiertas (Tabla 1) (Manley *et al.*, 2010).

Tabla 1. Nomenclatura actual de las proteínas SR

Proteínas SR	Seudónimo
SRSF1	SF2/ASF, ASF, SF2, SF2p33, SFRS1, SRp30a
SRSF2	PR264, SC-35, SC35, SFRS2, SFRS2A, SRp30b
SRSF3	SFRS3, SRp20
SRSF4	SFRS4, SRP75
SRSF5	HRS, SFRS5, SRP40
SRSF6	B52, HEL-S-91, SFRS6, SRP55
SRSF7	9G8, AAG3, SFRS7
SRSF8	SRSF8, DSM-1, SFRS2B, SRP46
SRSF9	SFRS9, SRp30c
SRSF10	FUSIP1, FUSIP2, NSSR, PPP1R149, SFRS13, SFRS13A, SRp38, SRp40, T ASR, T ASR1, T ASR2
SRSF11	NET 2, SFRS11, dJ677H15.2, p54
SRSF12	SFRS13B, SFRS19, SRp35

12.4.1.1 Mecanismos de las proteínas SR.

Las proteínas SR generalmente se conocen como factores que se unen a secuencias exónicas adyacentes a los sitios de empalme que funcionan como promotoras del reclutamiento de las snRNPs U1 y U2, como se muestra en la figura 6 (Shepard *et al.*, 2009).

Se han identificado otras proteínas denominadas proteínas similares o relacionadas a las SR, las cuales presentan dominios RS, sin embargo, muestran diferencias en la estructura de dominio de reconocimiento a ARN o en algunos casos puede que este dominio no esté presente y en su lugar se encuentre el dominio de dedos de Zinc; también pueden no ser reconocidas por el anticuerpo característico mAb104 (Boucher *et al.*, 2001).

Como se había mencionado antes existen elementos en *cis* que regulan eventos de *splicing*, los cuales se encuentran normalmente en el exón regulado y por lo tanto se conocen como *exonic splicing enhancers* (ESEs). Los elementos ESE son generalmente reconocidos por al menos un miembro de la familia de proteínas SR y reclutan a la maquinaria de *splicing* hacia el intrón adyacente. Las proteínas SR actúan en varias etapas durante el *splicing* y requieren de la fosforilación para el reconocimiento del sitio de empalme y su defosforilación para la catálisis (Busch *et al.*, 2009). Se han identificado varias proteínas quinasas que fosforilan específicamente residuos de serina del dominio RS, incluyendo las proteínas SR quinasa 1 (SRPK1), Clk/Sty quinasa, cdc2p34, y la topoisomerasa (Colwill *et al.*, 1996).

Sin embargo, existen algunos casos en los que las proteínas SR reconocen algún sitio de unión dentro del intrón, donde funcionan como reguladores negativos de *splicing* (Fig. 6). El ejemplo mejor caracterizado se produce durante la infección por adenovirus (Kanopka *et al.*, 1996). En este caso, el empalme es reprimido por la unión del factor SR denominado SF2/ASF a un elemento represor intrónico situado río arriba de la secuencia de ramificación al sitio de empalme 3' en el pre-ARNm de adenovirus.

Estas observaciones muestran que los ESEs no sólo funcionan en el exón facilitando el reconocimiento de sitio de *splicing*, sino que también actúan como barreras para evitar la omisión de exón.

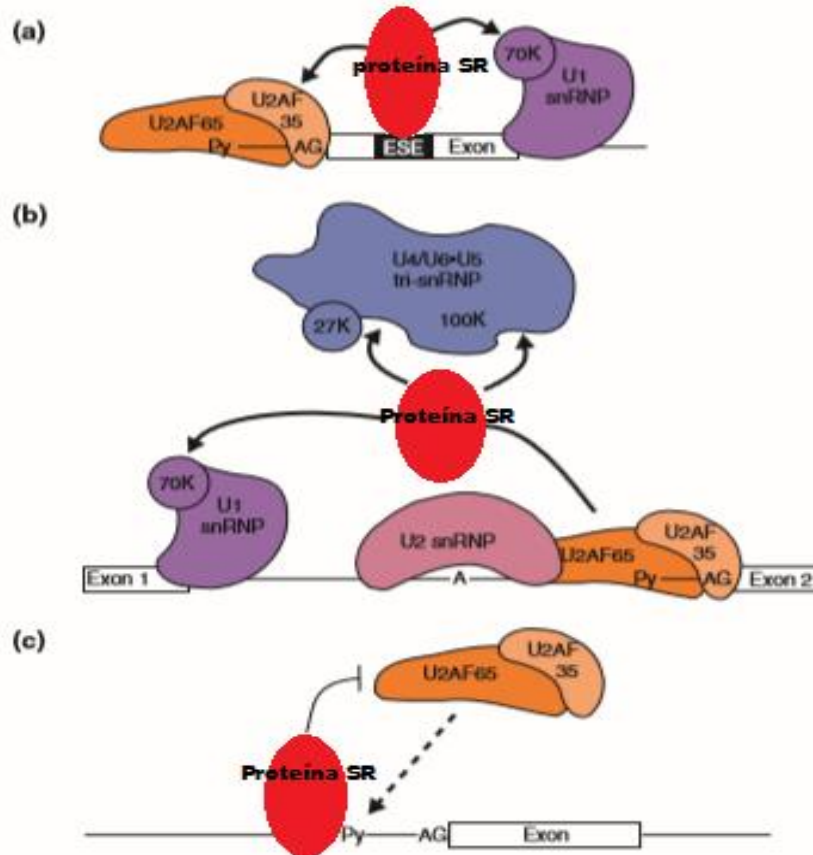


Figura 6. Funciones de proteínas SR en el proceso de *splicing*. a) La proteína SR (roja) se une a un sitio ESE e interactúa con los factores U2AF unidos al sitio 3' río arriba y con la snRNP U1 unida al sitio 5' río abajo. b) Las proteínas SR facilitan la interacción con los sitios de empalme y reclutan el complejo de la tri snRNP U4/U6-U5. c) Represión de *splicing* mediada por Proteínas SR asociadas a secuencias intrónicas (Modificado de Shepard *et al.*, 2009).

Las proteínas SR son proteínas multifuncionales que participan en una amplia diversidad de mecanismos reguladores que controlan el metabolismo del ARNm, tanto en el núcleo como en el citoplasma (Tabla 2). Dicha regulación está condicionada por varios parámetros, tales como la abundancia de proteínas SR, su perfil de fosforilación, su sitio de unión en la molécula del ARNm y su asociación con otros factores dentro del ARNm (Twyffels *et al.*, 2011), es decir, participan en la regulación de *splicing* constitutivo y alternativo de pre-ARNm, en la mediación del transporte eficiente de ARNm, así como en la modulación de la traducción del ARNm (Busch *et al.*, 2009).

Las proteínas SR no solo participan en el *splicing* constitutivo, en el *splicing* alternativo del pre-ARNm y en diversas actividades posteriores al *splicing*, sino que además juegan un papel importante en la exportación de ARNm nuclear, en el mecanismo de NMD, y en la traducción del ARNm. Estas actividades resaltan la importancia de las proteínas SR en varios mecanismos relacionados con la regulación del metabolismo del ARNm (Twyffels *et al.*, 2011).

Tabla 2. Diversidad de funciones de proteínas SR.

Proteína SR	SRSF1	SRSF2	SRSF3	SRSF4	SRSF5	SRSF6	SRSF7	SRSF8	SRSF9	SRSF10	SRSF11	SRSF12
Activador de <i>splicing</i> constitutivo y alternativo	X	X	X	X	X	X	X	X	X			
Estimulador de elongación transcripcional	X	X					X					
Contribuye a la estabilidad del genoma	X	X										
Contribuye a la exportación de ARNm	X		X				X					
Regula la traducción	X											
Activa NMD	X	X										
Promueve la mediación de IRES			X									
Reprime el <i>splicing</i>										X		
Reprime el <i>splicing</i> alternativo											X	X

Algunas proteínas SR como, SF2/ASF (SRSF1), SRp20 (SRSF3), y 9G8 (SRSF7) se transportan continuamente entre el núcleo y el citoplasma; el movimiento de estas proteínas requiere la fosforilación de residuos específicos en el dominio RS y en el dominio de unión al ARN (Cáceres *et al.*, 1998). Una vez más, la fosforilación de residuos de serina específicos dentro del dominio RS parece controlar la eficiencia de la función de exportación del ARNm de Npl3 (Lee *et al.*, 1996). También se ha demostrado que las proteínas SRSF3 y SRSF7 están involucrados en la mediación de la exportación eficiente de ARNm por la proteína asociada-Tap (TAP), que es un factor de exportación nuclear esencial (Huang *et al.*, 2004).

Evidentemente las proteínas SR constituyen una familia de reguladores con funciones importantes en el metabolismo del RNA, lo cual se ejemplifica cuando se considera que los cambios en la función de proteínas SR o en su abundancia con frecuencia se han asociado con la enfermedad humana. Por ejemplo, SRSF1 ha sido descrito como un proto-oncogén y la regulación deficiente de *splicing* alternativo mediado por este factor se ha asociado con varios tipos de cáncer (Faustino *et al.*, 2003). Otro ejemplo es el factor de *splicing* SFRS3, cuya expresión se encuentra elevada en diferentes tipos de cáncer; a pesar de que el mecanismo por medio del cual actúa no está totalmente elucidado, se sabe que esta misma proteína regula la expresión del factor de transcripción FoxM1 y dos de sus blancos transcripcionales (PLK1 y la Cdc25B), que están involucrados en la progresión del ciclo celular y en el control de la proliferación (Jia *et al.*, 2010). Por otro lado, se han detectado concentraciones elevadas de proteínas SR que se expresan durante la infección de las células con

el virus VIH-1 que dan como resultado una fuerte disminución de la progenie viral (Jacquenet *et al.*, 2005).

Si bien se ha demostrado que la participación de las proteínas SR en diversos aspectos de la expresión génica es generalizada, no sería sorprendente demostrar que juegan un papel crítico en otros procesos biológicos importantes.

12.5 *Splicing* en hongos.

Diferentes especies de levaduras tienen bajo número de intrones, *Saccharomyces cerevisiae* tiene sólo 253 intrones y únicamente seis genes contienen dos intrones (Barrass *et al.*, 2003), por lo que consecuentemente el *splicing* alternativo es bajo comparado con algunos hongos filamentosos (Kempken, 2013). Recientemente se describió que del 1.3 al 4.3 % de los genes en ciertas especies de ascomicetos estudiadas ocurren eventos de *splicing* alternativo (Zhao *et al.*, 2013).

Cabe señalar que en *Neurospora crassa* y en todos los ascomicetos filamentosos, los intrones se organizan de forma diferente en sus extremos 3' en comparación con la mayoría de los eucariotas. El punto de ramificación predicho se encuentra más cercano del sitio de empalme 3' y al tracto de polipirimidinas (Henscheid *et al.*, 2008). Se estimó que en el 1.6 % de los genes en *Aspergillus flavus* ocurre *splicing* alternativo (Chang *et al.*, 2010). *Fusarium graminearum* presenta resultados muy similares, en donde se predicen 4 eventos de exclusión de exón, 269 casos de retención de intrón, 25 casos de sitio alternativo 5' y 15 casos de sitio alternativo 3' (Zhao *et al.*, 2013).

Un ejemplo bien estudiado y típico de la retención de intrón fúngico en *S. cerevisiae* es el gen *CYH2*. El ARNm de *CYH2* codifica una proteína ribosomal de 17 kDa. El intrón en el pre-ARNm de *CYH2* se retiene aproximadamente el 50 % de las ocasiones, lo que resulta en un ARNm que codifica para un péptido de 2 kDa sin función conocida. (Schwindinger & Warner., 1987).

En *Saccharomyces cerevisiae*, se demostró que el *spliceosoma* utiliza todos los sitios de empalme 3' disponibles dentro de una cierta distancia del sitio de ramificación y en la mayoría de los casos, esto está mediado por estructuras del pre-ARNm (Luukkonen *et al.*, 1997). Se han identificado los componentes del *spliceosoma* de *D. discoideum* (Hinas *et al.*, 2006) y se compararon con *Drosophila melanogaster*, *Arabidopsis thaliana*, *Saccharomyces cerevisiae* y *Homo sapiens*. Si bien se identificaron los 49 ortólogos de proteínas snRNP de humanos en ameba, mosca, planta y en la levadura, sólo 43 ortólogos de levadura están presentes (Yu *et al.*, 2011). Del mismo modo, se han identificado en la levadura pocos ortólogos para proteínas *spliceosomales* que no son snRNPs y para proteínas ribonucleares heterogéneas (hnRNP) asociadas con el *spliceosoma*. También se identificaron doce proteínas que se asocian con la regulación de *splicing* alternativo (Kempken, 2013).

12.5.1 Proteínas SR en Modelos Fúngicos.

Existen proteínas SR en todas las especies de metazoos, así como en algunos eucariotas inferiores. (Kress *et al.*, 2008). En la levadura, sólo un puñado de genes contienen más de un intrón y la mayoría de los sitios de *splicing* se adhieren a un consenso estricto. La falta de oportunidades para el *splicing* alternativo ha promovido la creencia generalizada de que la levadura también carece de las proteínas SR (Shen *et al.*, 2006). Sin embargo, Npl3, Hrb1 y GBP2 comparten la estructura del dominio básico de las proteínas SR (Häcker *et al.*, 2004); además, Npl3 y GBP2 pueden ser fosforiladas en residuos de serina por la quinasa Sky1, la cual tiene una alta similitud estructural con la Proteína Quinasa 1 (SRPK1) de los mamíferos (Lukasiewicz *et al.*, 2007).

En *S. cerevisiae*, Npl3, modula la eficiencia del *splicing* del pre-ARNm (Kress *et al.*, 2008) y Srp2p es un mediador de la función ESE en la levadura de fisión *Schizosaccharomyces pombe* (Webb *et al.*, 2005).

Más recientemente la proteína SR en levaduras Npl3 se ha implicado en la elongación, terminación de la transcripción y en la traducción (Bucheli & Buratowski, 2005; Dermody *et al.*, 2008; Deka *et al.*, 2008), ya que actúa en la etapa temprana del *splicing*, junto con BBP y con la snRNP U1, estando así ligada al ensamble temprano del *spliceosoma* porque recluta a las snRNPs U1 y U2 (Kress *et al.*, 2008).

12.6 *Ustilago maydis*.

Ustilago maydis es un hongo que induce la formación de tumores en el maíz (*Zea mays*) y teosinte (*Zea mexicana*) comúnmente conocido como huilacoche, el cual puede existir en forma haploide unicelular "levadura" o como una forma dicariótica filamentosa. Esta transición dimórfica se relaciona con la patogenicidad del hongo, de tal manera que el dimorfismo, parasitismo y patogenicidad son aspectos interrelacionados del ciclo de vida de este microorganismo (Banuett, 1995).

U. maydis pertenece a los Basidiomycetes y se clasifica en el orden de los Ustilaginales, que infectan a un gran número de plantas dicotiledóneas y monocotiledóneas causando graves pérdidas económicas a los productores. Los rasgos característicos de los hongos incluyen su dimorfismo y la producción de grandes cantidades de esporas de color oscuro que dan la impresión de hollín (teliosporas) (Banuett, 1995). Otros patógenos pertenecientes a los Ustilaginales son *Ustilago hordei* (carbón cubierto), *Ustilago nuda* (carbón volador), ambos parasitan la cebada; *Ustilago nigra* (carbón parcial) que afecta la avena, entre otros (Herrera, 2008).

Algunas de las características que hacen de *U. maydis* un organismo genéticamente manejable para estudiar diferentes aspectos celulares y moleculares incluyen:

- a) La existencia de una fase haploide donde se forman colonias compactas, que permite el uso de técnicas microbiológicas estándar.

- b) La capacidad de construir células diploides en el laboratorio, lo cual hace que sea posible llevar a cabo pruebas de dominancia y de complementación.
- c) La posibilidad de transformación con ADN exógeno.
- d) La capacidad de sustituir un segmento genómico con una copia mutada de un gen por recombinación homóloga.
- e) La existencia de algunos marcadores seleccionables dominantes para las manipulaciones genéticas (Banuett, 1995).

12.6.1 Ciclo de vida de *Ustilago maydis*.

Aunque *Ustilago maydis* se considera un hongo dimórfico, exhibe tres formas en su ciclo de vida: una forma haploide unicelular, uninucleado, que es saprófita (puede crecer sobre la materia no viviente); una forma filamentosa dicariótica, que es parasitaria (necesita de la planta para su crecimiento) y patógena (induce la enfermedad) y finalmente una forma diploide (la teliospora) que sólo se forma dentro de los tumores en la planta (Fig. 7). Estas transiciones morfológicas se rigen por el apareamiento de dos loci: a y b. La forma haploide unicelular es en forma de cigarro, crece por gemación polar, forma colonias compactas en diferentes medios y no es patógeno; en la naturaleza, es saprófita. La fusión celular se produce sólo si las células llevan diferentes alelos del locus responsable del apareamiento. La fusión celular se realiza normalmente en las superficies de las hojas (Banuett, 1995).

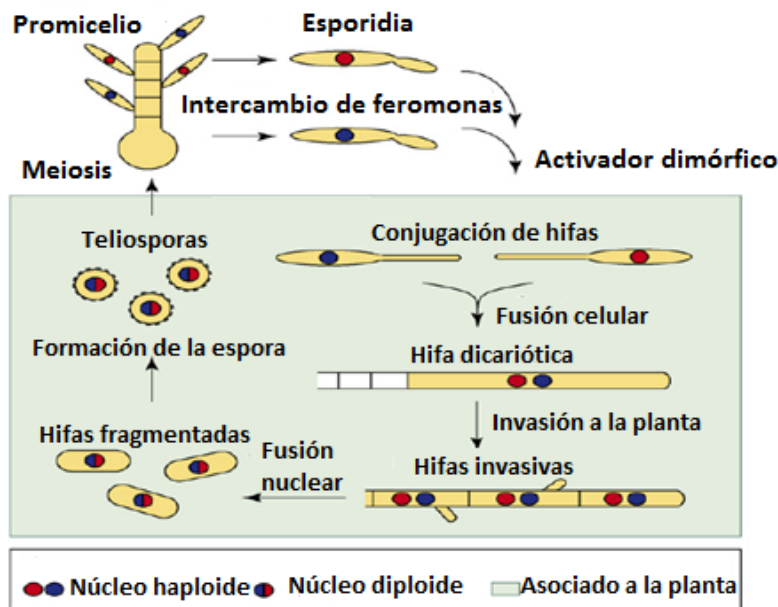


Figura 7. Ciclo de vida de *Ustilago maydis*. Proceso mediante el cual se desarrolla *U. maydis*; la forma saprófita y haploide (esporidio) que posterior a su conjugación genera su forma patógena y dicariótica (teliospora) por medio de un intercambio de feromonas (Steinberg & Perez-Martin, 2008).

Modelos fúngicos, tales como *S. cerevisiae* han mejorado enormemente nuestro conocimiento de la biología celular básica, en parte debido a las numerosas ventajas técnicas que presenta. Sin embargo, algunos procesos celulares que son importantes en los seres humanos están altamente modificados o ni siquiera están presentes en esta levadura. La evidencia experimental obtenida a partir de diversos experimentos realizados en *U. maydis* indicó que este hongo podría llenar este vacío y ser empleado como modelo de estudio. Este organismo tiene una larga historia como un sistema modelo para la reparación del ADN y la recombinación, además de que muestra similitudes adicionales a las células humanas, como la mitosis abierta (Castillo-Lluva *et al.*, 2007) y el transporte a larga distancia (Steinberg, 2007). De hecho, aproximadamente el 10 % del total de las proteínas de *U. maydis* están altamente conservadas y presentan homólogos en humanos (Münsterkötter & Steinberg, 2007).

La inesperada conservación entre *U. maydis* y los humanos podría sugerir que ambos organismos comparten algunos procesos celulares conservados. Sin embargo, es importante tener en cuenta que las predicciones funcionales basadas en homología de secuencia pueden ser sólo un primer paso hacia la comprensión de la función celular. Se necesita un trabajo experimental cuidadoso para demostrar además que *U. maydis* ayuda a la comprensión de la base molecular de las enfermedades humanas. Enfoques bioinformáticos han revelado que 681 homólogos entre *U. maydis* y los seres humanos presentan un porcentaje de identidad mayor al 20 % a nivel de secuencia aminoacídica. Haciendo uso de la base de datos de SIMAP, se encontró que 620 proteínas cumplen este criterio (Fig. 8). Cabe resaltar que este grado de conservación es menor entre *U. maydis* y *S. cerevisiae* (Münsterkötter & Steinberg, 2007).

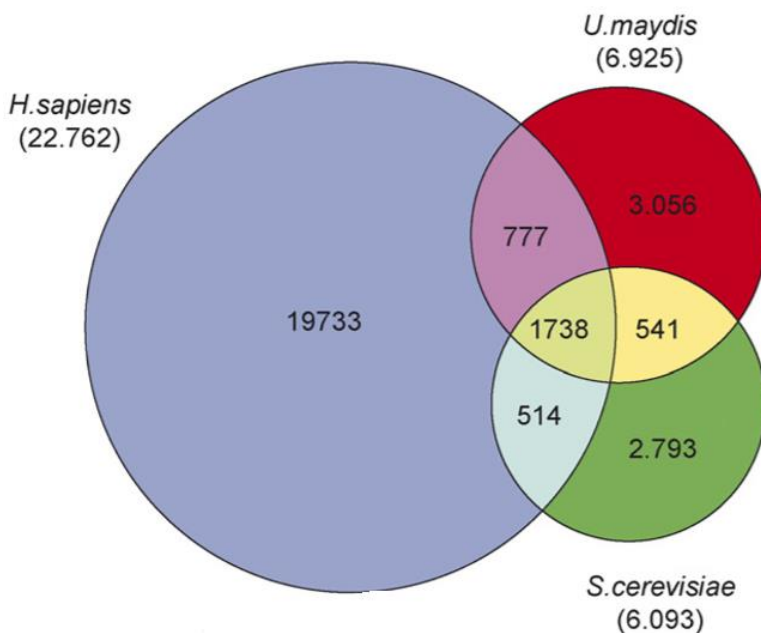


Figura 8. Comparación de secuencias de proteínas predichas en *Ustilago maydis*, *Saccharomyces cerevisiae* y *Homo sapiens*. *H. sapiens* y *S. cerevisiae* tienen un conjunto de 514 genes en común, mientras que 1738 proteínas muestran una identidad mayor al 20 % en los tres organismos. Sin embargo, *U. maydis* comparte un gran número de genes (777) homólogos con respecto al humano (Tomado de Münsterkötter & Steinberg, 2007).

13. JUSTIFICACIÓN.

Debido a que las proteínas SR desempeñan una gran variedad de funciones, dentro de las que se encuentran la exportación de ARNm y procesos de degradación de ARNm (NMD), se han asociado principalmente a procesos de regulación transcripcional, actuando de manera general como potenciadores en los eventos de *splicing*, sumado a esto también se ha demostrado que existen niveles elevados de proteínas SR (SRSF1 y SRSF3) en diversas enfermedades humanas como el cáncer. Como resultado de todas estas observaciones surge la necesidad de estudiar a esta familia de proteínas reguladoras en el modelo experimental *Ustilago maydis*, tomando en cuenta que los resultados arrojados de este estudio ofrecen resultados novedosos en cuanto a la existencia de proteínas SR en este modelo y por otro lado, brinda la posibilidad de obtener una nueva herramienta que nos permita estudiar a estos factores y posiblemente extrapolar los resultados obtenidos sobre la funcionalidad de las proteínas hacia otros organismos.

14. OBJETIVOS.

14.1 Objetivo General.

Identificar homólogos putativos para proteínas SR en *Ustilago maydis* y sus posibles secuencias blanco.

14.2 Objetivos Particulares

- ❖ Identificar *in silico* en el genoma de *Ustilago maydis* los homólogos para las proteínas SR reportadas para humano.
- ❖ Identificar en la anotación del genoma de *Ustilago maydis* proteínas que contengan dominios RRM y RS característicos de las proteínas SR.
- ❖ Comprobar la presencia de los genes codificantes para algunas proteínas SR (SRSF1, RMB39 y RNPS1) en el genoma y a partir de esto generar las proteínas recombinantes.
- ❖ Buscar en el genoma de *Ustilago maydis* las secuencias para los sitios putativos de regulación mediada por proteínas SR.
- ❖ Evaluar la capacidad de interacción de las secuencias reguladoras identificadas y sus respectivas recombinantes.

15. METODOLOGÍA

15.1 Cepas utilizadas, condiciones de cultivo y plásmidos.

15.1.1 *Ustilago maydis* FB1.

Ustilago maydis es nuestro modelo de estudio y la cepa que utilizamos fue FB1 (cepa de referencia, *Broad Institute*), su genoma se obtuvo por medio ensamble de secuencias tomando en cuenta la cepa de *U. maydis* 521, este trabajo se llevó a cabo por el *Broad Institute*.

U. maydis se creció en medio YEPS a 20 °C/250 rpm.

15.1.2 *E. coli* DH5 α .

La cepa de *E. coli* DH5 α se utilizó como reservorio de nuestra construcción y también sirvió para realizar los ensayos de expresión de las proteínas recombinantes generadas. Sus características son LacZ Delta M15, endA1, recA. (Taylor, 1993).

E. coli se creció en medio LB a 37 °C en agitación de 300 rpm. En el caso de la obtención y propagación de las construcciones realizadas y la propagación de plásmidos se creció en medio LB adicionado con Ampicilina a una concentración final de 1 ng/ μ L.

15.1.3 pJET 1.2 (ThermoFisher).

Este vector funciona como reservorio para productos de PCR, posee varias características entre las que se incluyen: rep (pMB1). bla (Ap), eco 47IR, PlacUV5, promotor T7. Sitio múltiple de clonación (SMC).

15.1.4 pPROEX HT (c) (ThermoFisher).

Este vector se utilizó para realizar la expresión de proteínas recombinantes, posee varias características entre las que se encuentran: rep (pBR322), His (6) N-terminal. Bla (Ap), lacI, promotor Tcr, Sitio múltiple de clonación (SMC)

15.2 Medios de cultivo.

15.2.1 YEPS (Yeast Extract, Peptone, Sugar).

Se utilizó para propagar *U. maydis*. Cada litro de medio contiene 20 g de dextrosa, 10 g de extracto de levadura y 20 g de peptona de caseína.

En caso de preparar medio sólido se agregaron 15 g de agar por litro.

15.2.2 LB (Luria- Bertani; Sambrook *et al.*, 1989).

Es un medio de cultivo utilizado de forma habitual para el crecimiento bacteriano. Se utilizó para propagar *E. coli*. Cada litro de medio contiene 5 g de extracto de levadura, 10 g de NaCl y 10 g de peptona de caseína.

En caso de preparar medio sólido se agregaron 15 g de agar por litro.

15.2.3 SOC.

Este medio se utilizó para rescatar las células de *E. coli* DH5 α que fueron transformadas por electroporación. Este medio contiene 2 % de triptona, 0.5 % de extracto de levadura, 10 mM de NaCl, 2.5 mM de KCl, 20 mM de glucosa, 10 mM de MgSO $_4$ y de MgCl $_2$.

15.2.4 GYT.

Este medio se utilizó para almacenar las células electrocompetentes de *E. coli*

Cada 100 mL de medio contiene 10 mL de glicerol, 0.125 g de extracto de levadura y 0.25 g de triptona.

15.3 Análisis bioinformático.

15.3.1 Análisis *in silico* de homólogos para proteínas SR de humanos en el genoma de *Ustilago maydis*.

Para este abordaje se inició con una búsqueda de proteínas SR de humanos reportadas en la literatura, posteriormente se realizó un alineamiento comparando la secuencia aminoacídica de la proteína SR de humano contra el genoma de *Ustilago maydis* utilizando un Blast para proteínas en la plataforma del NCBI (*National Center for Biotechnology Information*), el cual utiliza el algoritmo BLOSUM62. Posteriormente los locus identificados fueron enlistados, se anotó el número de aminoácidos correspondientes para las proteínas SR de humano y para los locus identificados de *U. maydis*; se identificaron las diferentes denominaciones que recibían todas las proteínas SR de humano identificadas en la literatura y se realizó la búsqueda bibliográfica de los homólogos ya descritos para todas estas proteínas en diferentes organismos. Estos datos se correlacionaron con los resultados obtenidos del segundo abordaje (Metodología 4.4.2) con el fin de confirmar y reforzar la identidad de las proteínas SR putativas identificadas.

15.3.2 Análisis *in silico* de proteínas SR putativas en *Ustilago maydis*.

Para identificar de manera exhaustiva la existencia de proteínas SR en el genoma de *Ustilago maydis* se utilizó la base de datos de Pfam, esta base de datos agrupa las proteínas en familias de

acuerdo con los dominios funcionales; por tal razón nos permitió identificar todas aquellas proteínas que contenían dominios RRM, que permiten la unión de la proteína con el ADN. Esta búsqueda se realizó utilizando la plataforma del *Broad Institute*. Posteriormente se identificaron los dominios RS, con ayuda de la plataforma del *KEGG (Kyoto Encyclopedia of Genes and Genomes)* se corroboró el número de dominios RRM de la proteína y el tamaño de cada uno de ellos. Por otro lado, se utilizó la herramienta *KEGG ORTHOLOGY*, para identificar el homólogo de cada una de las proteínas con dominio RRM y conocer si han sido descritas con alguna función. Estos datos fueron utilizados para realizar una correlación con los resultados obtenidos del apartado anterior (Metodología 4.4.1) de tal manera que se completa la búsqueda y se genera una lista que incluya tanto los homólogos putativos de proteínas SR humanas, como posibles proteínas SR específicas del basidiomiceto.

15.3.3 Análisis *in silico* de interacción con otras proteínas.

Este análisis se realizó utilizando la base de datos *STRING (Search Tool for the Retrieval of Interacting Genes/Proteins)*; la cual nos permitió conocer las posibles interacciones proteicas para algunas de las proteínas SR putativas identificadas en *Ustilago maydis*.

STRING es un sistema de predicción el cual se basa en la extrapolación de datos reportados de interacciones proteicas hacia la proteína de interés, a partir de los cristales de proteínas reportadas en las bases de datos, tomando en cuenta la presencia de dominios funcionales y su homología. Esta herramienta predice interacciones proteicas basándose tanto en interacciones reportadas y validadas mediante datos experimentales como en datos de homología estructural y funcional, empleando métodos de predicción computacional.

15.3.4 Búsqueda de sitios blanco.

Proteínas recombinantes uRNPS1, uSRSF1, uRBM39; se buscó en la literatura la secuencia reportada como sitio de unión reportados para los homólogos humanos y posteriormente se alimentó el programa GenFinder (Hernández-Pérez *et al.*, en proceso) con las secuencias. El empleo de este programa nos permite identificar las secuencias blanco presentes en cada gen, así como su ubicación. El listado arrojado se analizó en función de la frecuencia, prevalencia y posición de las secuencias blanco identificadas.

15.4 Técnicas de Biología Molecular.

15.4.1 Obtención de ADN genómico a partir de la cepa FB1 de *Ustilago maydis* (*Kit Wizard® Genomic DNA Purification, Promega*).

La obtención de DNA genómico se realizó de acuerdo con las indicaciones del proveedor, como se indica a continuación. Inicialmente se cultivó el hongo en 5 ml de medio líquido YEPS durante toda la noche a 30 °C en agitación constante. El cultivo se centrifugó 2 min/13000 rpm, se resuspendió en

500 μ L de EDTA 50 mM. Para lisar las células se agregaron 7 μ L de liticasa (10 μ g/mL) mezclando suavemente con la punta, se incubó a 37 °C/45 min, se centrifugó y desechó el sobrenadante, posteriormente se adicionaron 300 μ L de *Nucleic lysis solution* en la cual se resuspendió la pastilla. Para la precipitación de proteínas se adicionaron 100 μ L de *protein precipitation solution*, se mezcló con vortex y se incubó en hielo ambos casos por 10 min. Tras centrifugar a 1300 rpm/3 min se realizó la precipitación de ADN, para esto el sobrenadante se transfirió a un tubo de 1.5 mL y se adicionaron 300 μ L de isopropanol frío y se incubó en hielo 30 min; se centrifugó y la pastilla resultante se lavó con 300 μ L de etanol al 70 %, nuevamente se centrifugó y se aspiró por completo el etanol con ayuda de una punta; la pastilla se dejó secar al aire por 20 min. Para rehidratar el ADN se adicionaron 50 μ L de *RNase solution*, se mezcló con una punta y se incubó a 37 °C/15 min, dejando enfriar a temperatura ambiente. Por último, se dejó incubando a 4°C toda la noche. Las muestras fueron almacenadas a -80 °C.

15.4.2 Amplificación de fragmentos de ADN mediante PCR (*New England BioLabs*).

Los *primers* se diseñaron de tal manera que nos permitieran amplificar el marco de lectura abierta completo en la siguiente tabla se muestran los locus utilizados, el tamaño del amplicón y los *primers* utilizados.

Tabla 3. Oligonucleótidos empleados para obtener las proteínas recombinantes de *U. maydis*.

Locus en <i>U. maydis</i>	Nombre asignado	Tamaño amplicón	Fwd (5'-3')	Rev (5'-3)	Referencia
UMAG_0016 4	uRBM39	1923 pb	ATGTATGCCGA CTTGGAAGAG	CTACTGCGCGC TGATTGGC	Este estudio
UMAG_0255 2	uSRSF1	4011 pb	ATGTCGGCAAT CAACCACGCA	CTACGATGCGTT GCTTCCAG	Este estudio
UMAG_3178	uRNPS1	594 pb	ATGGATCGAAC AGCTGATTCA	TCATCGTGTGTG TTGGCGG	Este estudio

La reacción de amplificación por PCR se realizó utilizando 1 μ L de ADN como templado (10-100 ng), 5 μ L de ThermoPol Reaction Buffer 10X, 1 μ L de dNTPs [10mM], 1 μ L de cada oligonucleótido [10 mM], 0.25 μ L de Vent DNA Polymerase (5 U), en un volumen final de 50 μ L.

Al término de la reacción, los productos de PCR se visualizan mediante una electroforesis en gel de agarosa al 1 %, los cuales se corrieron a 100 mV/30 min para comprobar la obtención de los amplificados esperados.

Tabla 4. Condiciones para las reacciones de PCR

Proteína Homóloga en <i>U. maydis</i>	Templado FB1	Tm °C	Tiempo de extensión	Tamaño del amplificado	Referencia
uRBM39	ADN	62	2'	1923 pb	Este trabajo
uSRSF1	ADN	62	4'10''	4011 pb	Este trabajo
uRNPS1	ADN	60	40''	594 pb	Este trabajo

El programa de amplificación empleado de manera general se describe a continuación: pre-desnaturalización de 95 °C/5 min, posteriormente 30 ciclos de desnaturalización a 95 °C/30 s, 30 s a la temperatura de alineamiento correspondiente (Tabla 4) y una extensión a 72 °C durante el tiempo requerido de acuerdo al tamaño del amplicón. Por último, se realizó una extensión final a 72 °C durante 5 min y el termociclador se programó para almacenar las reacciones a 10 °C.

15.4.3 Extracción de ADN plásmidico mediante Lisis alcalina (Green & Sambrook, 2012).

Para la extracción de ADN plásmidico se cultivó cada una de las clonas obtenidas en 5 mL de medio LB-Amp (1 µg/1 mL), incubando toda la noche a 37 °C/150 rpm. Se centrifugó el cultivo a 6000 rpm/7 min, se desechó el sobrenadante y se agregaron 200 µL de solución I (50 mM de glucosa, 25 mM de Tris-HCl pH 8.0, 10 mM de EDTA, pH 8), se resuspendió la pastilla y se trasladó a un tubo de 1.5 mL donde se mezcló con vortex hasta lisar la pastilla. Para realizar la lisis alcalina se adicionaron 200 µL de solución II (200 mM de NaOH, 1 % SDS) y se mezcló por inversión 10 veces. La precipitación de proteínas se realizó adicionando 200 µL de solución III (5 M de Acetato de potasio/2 M de ácido acético, pH 5,5), se mezcló con vortex al máximo por 10 s y se centrifugó al máximo por 5 min; posteriormente para aislar el ADN plasmidico se recuperó el sobrenadante y se adicionaron 800 µL de etanol absoluto, se mezcló y se incubó a -4° C/30 min. Se centrifugó al máximo 5 min, se desechó el sobrenadante y se lavó la pastilla con etanol al 70 %; nuevamente se centrifugó y se desechó el sobrenadante, dejando que la pastilla se secase al aire. Por último, se resuspendió la pastilla en 50 µL de agua estéril con RNAsa (1 µg/ml).

15.4.4 Electroforesis de ADN en geles de agarosa (Green & Sambrook, 2012).

Se utilizaron geles de agarosa al 1 % para visualizar las muestras de ADN genómico, plásmidos y productos de digestiones enzimáticas realizadas. Los geles se prepararon utilizando buffer TAE (40 mM de Tris-acetato, pH 8, 1 mM de EDTA) y agregando 1 gota de Bromuro de etidio (1 µg/ml); el corrimiento se realizó empleando el mismo buffer. Las muestras se mezclaron con buffer de carga (0.25 % de azul de bromofenol (p/v), 0.25 % de xilencianol (p/v), 15 % de glicerol (p/v), en agua, pH 8). Como referencia se utilizó el marcador de peso molecular de 1 kb (fermentas).

Todas las electroforesis se migraron a 100 mV y se visualizaron con un transluminador de luz ultravioleta (300 nm de longitud de onda).

15.4.5 Clonación de los productos de PCR en el vector pJET 1.2 (CloneJET PCR Cloning Kit- Thermo Fisher Scientific).

La clonación de los productos de PCR se realizó utilizando el vector pJET 1.2, el cual permite clonar de manera directa productos de PCR sin necesidad de purificarlos. Este sistema posee una eficiencia del 99 %, las características importantes que presenta este vector son: un promotor de transcripción T7, un gen letal el cual es interrumpido cuando se inserta ADN en los sitios de clonación, un casete de resistencia a ampicilina y sitios múltiples de clonación y de restricción.

La clonación en este vector funcionó como reservorio de nuestro producto de PCR y nos brindó sitios de corte específicos con los cuales posteriormente se realizó una subclonación en el vector de expresión pPROEX HT.

Tabla 5. Componentes y cantidades para las reacciones de ligación.

Componentes	Volumen
Buffer 10 X	1 µL
Productos de PCR no purificados	1 µL - 0.15 pmol
pJET 1.2/blunt cloning vector (50 ng/µL)	1 µL
Agua libre de nucleasa	6 µL
Ligasa T4	1 µL
Volumen total	10 µL

La reacción de ligación se realizó de acuerdo a los componentes y las cantidades mostradas en la tabla 5 y se dejó incubando durante 2 días a temperatura ambiente.

15.4.6 Subclonación de los productos digeridos con NotI-XbaI en el vector pPROEX-HTc (Addgen).

A partir de las clonas obtenidas en dirección 5'-3' de la transformación en el vector pJET 1.2 blunt, se realizó la subclonación en el vector pPROEX- HTc.

El vector pPROEX-HT posee un promotor Tcr, un gen *LacI*, un ori *C*, un gen β -lactamasa, casete de resistencia a ampicilina y múltiples sitios de corte.

Las digestiones se realizaron de acuerdo a lo siguiente: el ADN plasmídico y el vector pPROEX-HTc fueron digeridos con las enzimas NotI-XbaI (Fermentas), de la digestión del ADN plasmídico se obtuvo el inserto, el cual se ligó en el vector pPROEX- HTc cortado con las mismas enzimas para que la clonación fuera dirigida.

Para cada una de las digestiones se tomaron 43 µL de ADN plasmídico, 2 µL de enzima XbaI y 5 µL de buffer adecuado para la enzima, se dejó incubando toda la noche, posteriormente se agregaron 2 µL de enzima NotI, 10 µL de buffer adecuado para la enzima y 38 µL de agua libre de nucleasas, la reacción se incubó 37 °C/4 h. Las digestiones se cargaron en un gel de agarosa al 8 % y se cortaron las bandas correspondientes a los insertos, con la ayuda de una navaja estéril.

Los productos de ADN se purificaron con “High Pure PCR Product Purification Kit- Roche”, de acuerdo con las indicaciones del fabricante. Posteriormente, las ligaciones se realizaron de acuerdo con las cantidades indicadas en la tabla 5. La reacción se llevó a un volumen final de 10 µL, se incubó a temperatura ambiente durante dos días.

Tabla 6. Componentes y cantidades para las reacciones de ligación

Componentes	Volumen
Buffer 10X	1 µL
Vector cortado y purificado(pPROEX-THc-NotI-XbaI) Productos de digestión purificados	1:3
Agua libre de nucleasa	4 µL
Ligasa T4	1 µL
Volumen total	10 µL

15.4.7 Purificación de ADN (High Pure PCR Product Purification Kit- Roche).

Para purificar muestras de ADN, se realizó lo siguiente: el volumen de cada muestra se ajustó a 100 µL, se adicionaron 400 µL de *Binding buffer*, se mezcló la muestra con vortex y se centrifugó por 10 s. Posteriormente, se insertó una columna dentro de un tubo colector y usando una micropipeta se transfirió la muestra en el centro de la columna, se centrifugó 30 s/8000 g, se desconectó la columna del tubo colector y se desechó el sobrenadante. Después se ensambló la columna y se adicionaron 400 µL de *Wash buffer*, se centrifugó 30 s/8000 g, se desechó el sobrenadante, se volvió a lavar, pero ahora con 300 µL de *Wash buffer*, se desechó el sobrenadante nuevamente y se centrifugó a velocidad máxima por 1 min. Por último, la columna se conectó a un tubo nuevo de 1.5 mL y se adicionaron 12 µL de *Elution buffer* y para colectar el eluido se centrifugó a 8000 g/1 min.

Para el caso de las purificaciones realizadas a partir de muestras de ADN en banda se realizó lo siguiente: se pesó la banda y por cada 100 mg de gel se agregaron 100 µL de *Binding enhancer*, se calentó la muestra a 56 °C hasta que la agarosa se disolvió (no más de 20 min). Después se trasladó a un tubo nuevo y se realizó el proceso mencionado anteriormente para purificar ADN en solución.

15.4.8 Análisis de restricción enzimática.

El ensayo con enzimas de restricción es muy importante ya que debemos asegurarnos que el inserto se encuentre clonado de manera adecuada, es decir, se encuentre en dirección 5'-3' en el plásmido: este proceso se llevó a cabo tomando 5 µL de ADN plasmídico, 1.5 µL de buffer específico para cada enzima utilizada, 8.3 µL de agua libre de nucleasas y 0.2 µL de enzima, la reacción se incubó a 37 °C/2 h.

15.5 Técnicas de transferencia genética.

15.5.1 Preparación de células *E. coli* DH5α electrocompetentes (Green y Sambrook, 2012).

La preparación de células electrocompetentes se realizó mediante un tratamiento con glicerol frío al 15 %. En este procedimiento se parte del cultivo de una colonia de *E. coli* DH5α en 25 ml de medio LB, se incuba a 37 °C/250 rpm durante toda la noche; al siguiente día se inocularon 250 mL de medio LB con 12.5 mL del precultivo, se incubó a 37 °C/300 rpm hasta alcanzar una D. O₆₀₀ de aproximadamente 0.5-0.7. Una vez alcanzada la DO se trasladaron rápidamente a un baño en hielo durante 20 min, se dividió el cultivo en tubos estériles de 50 ml previamente puestos en un baño con hielo y se colectaron las células por centrifugación a 4 °C/1000 g/20 min. A continuación, se decantó el sobrenadante y se lavó la pastilla suavemente con 50 mL de agua fría con ayuda de una pipeta, nuevamente se colectaron las células por centrifugación a 4 °C/1000 g/20 min, se decantó el sobrenadante y la pastilla fue resuspendida en 25 mL de glicerol frío al 10 %. Posteriormente se repitió el procedimiento anterior, pero ahora se resuspendieron en 12.5 mL de glicerol frío al 10 %, nuevamente las células fueron centrifugadas en las mismas condiciones y resuspendidas en 1 mL de medio GYT frío. Por último, se resuspendieron y se realizaron alícuotas de 40 µL las cuales fueron almacenadas a -80°C hasta el momento de su uso.

15.5.2 Electroporación de células *E. coli* DH5α electrocompetentes (Green y Sambrook, 2012).

Las ligaciones correspondientes a pJET-uRBM39, pJET-uSRF1 y pPROc-uSRSF1 fueron introducidas a las células de *E. coli* DH5α electrocompetentes por medio de electroporación.

La electroporación de las células se realizó de la siguiente manera. En un tubo de 1.5 ml se colocaron 40 µL de células electrocompetentes descongeladas con 4 µL de la ligación (10 -100 ng) se mezclaron y se encubaron en hielo 1 min; se trasladó el contenido a una cubeta de 0.1 cm, se colocó en la cámara y usando el MicroPulser- Bio Rad se pulso "Ec1"; se removió la cubeta de la cámara e inmediatamente se adicionó 1ml de medio SOC; se resuspendieron las células y se trasladaron a un tubo de 15 ml, se incubaron a 37 °C/1 h/225 rpm.

Posteriormente, se sembraron en placas LB-ampicilina. Una placa se inoculó tomando el 10 % de las células (100 μ L de cultivo) y otra placa se inoculó con el resto de las células con la ayuda de un asa triangular de metal.

15.5.3 Preparación de células *E. coli* DH5 α competentes (Green y Sambrook, 2012).

Las células quimiocompetentes se prepararon de acuerdo con las siguientes indicaciones: se inició un precultivo de 5 ml de caldo LB inoculado con 1 UFC de *E. coli* DH5 α (37 °C/12 h), con el cual se inoculó un cultivo de 200 ml de caldo LB, se incubó a 37 °C hasta alcanzar una DO₆₀₀ de 0.5, se transfirió a tubos estériles de 50 ml previamente enfriados, se centrifugó 3000 rpm/10 min/4 °C, se eliminó el sobrenadante y se resuspendió la pastilla en 12.5 ml de CaCl₂ 0.1 M. Esta mezcla se almacenó en hielo por 10 min, se centrifugó 3000 rpm/10 min/4 °C, se eliminó el sobrenadante, se resuspendió la pastilla en 1 ml de CaCl₂ 0.1 M/Glicerol al 15 %. Las células se alicuotaron y almacenaron a -70 °C para su uso posterior.

15.5.4 Transformación de células *E. coli* DH5 α competentes (Green y Sambrook, 2012).

La transformación de las células se realizó de la siguiente manera: se tomaron 100 μ L de células competentes en tubos de 1.5 ml en hielo y 10 μ L de la ligación, se incubó la mezcla 40 min en hielo, se indujo choque térmico a 42 °C/90 s, se transfirió en tubo a hielo durante 90 segundos, se agregó 750 μ L de caldo LB, incubó 37 °C/40 min/300 rpm, se sembraron en placas LB-ampicilina. Una placa se inoculó tomando el 10 % de las células (100 μ L de cultivo) y otra placa se inoculó con el resto de las células con la ayuda de un asa triangular de metal.

Esta metodología se utilizó para realizar las subclonaciones de pPROc-uRBM39, pJET-uRNPS1 y pPROc-uRNPS1.

15.6 Técnicas de Proteínas recombinantes.

15.6.1 Inducción de proteínas recombinantes con IPTG (pPRoEX-HT Procariotic Expression System).

Para poder obtener las proteínas recombinantes inicialmente es necesario inducir su expresión. En este caso, el vector que utilizamos posee un promotor híbrido Trp-Lac, por lo cual puede ser activado en presencia de lactosa, sin embargo, debido a que necesitamos una expresión constante es necesario utilizar IPTG (Isopropil- β -D-1-tiogalactopiranosido), el cual es un análogo de la lactosa y posee la ventaja de no ser degradado. Esta inducción se realizó inoculando un precultivo de 2 ml de caldo LB-Amp con una UFC, incubando el cultivo 37 °C/200 rpm/12 h, después se transfirió 0.1 ml de precultivo en 10 ml de caldo LB-Amp el cual se incubó 37 °C/ 200 rpm hasta alcanzar una D.O₅₉₀ de 0.5-1.0. En este punto se tomó 1 ml correspondiente al tiempo 0, sin inducción. Al resto

del cultivo se adicionó IPTG a una concentración final de 0.6 mM, se continuó con la incubación y se tomaron alícuotas a la hora 1, 2 y 3 después de la inducción, se centrifugaron todas las muestras a 13000 rpm/1 min, se lavaron las células en 100 µL de PBS. En el caso de uRMB39 y uSRSF1 se incubaron a temperatura ambiente/12 h, después de la inducción, las alícuotas se resuspendieron en volúmenes iguales de buffer para muestra 2x [125 mM Tris-HCl (pH 6.8), 4 % (w/v) SDS, 20 % (v/v) Glicerol, 0.01 % (w/v) azul de bromofenol, 5 % de β-mercaptoetanol], se hirvieron las muestras y analizaron por electroforesis en gel de poliacrilamida desnaturizante.

15.6.2 Electroforesis de proteínas en gel de poliacrilamida SDS (Miniprotean4 BioRad).

Para poder visualizar las proteínas inducidas se realizó una electroforesis en gel desnaturizante de poliacrilamida (SDS-PAGE). Se utilizaron dos concentraciones diferentes con las siguientes cantidades para el gel concentrador y de separación:

- Gel al 12 % para uRNPS1: 4 ml de poliacrilamida [1:29], 3.4 mL de agua, 2.5 ml de buffer [Tris- HCl 1.5 mM] para el gel de separación y 2.5 mL de buffer [0.5 M Tris-HCl, pH 6.8] para el gel concentrador.
- Gel al 8 % para uRMB39/ uSRSF1: 2.7 ml de poliacrilamida [1:29], 4.7 mL de agua, 2.5 ml de buffer [Tris- HCl 1.5 mM] para el gel de separación y 2.5 mL de buffer [0.5 M Tris-HCl, pH 6.8] para el gel concentrador).

En ambos casos para polimerizar se utilizaron 5 µL de TEMED y 50 µL de APS al 10 % para el gel concentrador, en cuanto al gel de separación se agregaron 10 µL de TEMED y 100 µL de APS al 10 %.

Se cargaron 20 µL de las muestras en los pocillos, como patrón de referencia se utilizó el marcador para proteínas con un rango de 15-170 kDa (Page Ruler Fermentas). La electroforesis se corrió en buffer Tris – Glicina (500 mM de Tris, 500 mM de Glicina) a 80 mV los primeros 30 min y posteriormente a 100 mM durante 2 - 3 h, dependiendo del tamaño de la proteína a identificar.

15.6.3 Tinción de Geles de acrilamida con Coomassie.

Una vez terminada la electroforesis, se colocó el gel en un recipiente para inmediatamente ser teñido, se adicionó azul de coomassie (0.1 % de azul brillante de coomassie, 45 % de metanol y 10 % de ácido acético) lo suficiente para cubrir la superficie del gel, se dejó incubando por 15 min a temperatura ambiente en un agitador orbital. Se decantó el exceso de colorante y se lavó con agua. Para poder observar las bandas se adicionó una solución para desteñir (10 % de metanol y 5 % de ácido acético) y se mantuvo durante 18 h en el agitador orbital a velocidad media.

15.7 Western Blot.

Las proteínas recombinantes obtenidas de la inducción se cuantificaron por medio del método de Bradford y se corrieron en un gel de poliacrilamida al 8 % o al 12 %, dependiendo del tamaño de la proteína de acuerdo a las especificaciones previamente descritas. Posteriormente se realizó la transferencia a una membrana de PVDF (Millipore) previamente activada en metanol absoluto/ 5 min. La transferencia se realizó durante 16 h/30 mA, la membrana se bloqueó con PBS/Casac [5 %]

(Nestlé) con el anticuerpo monoclonal anti-His (Santa Cruz Biotechnology) a una dilución 1:500 toda la noche. Al día siguiente se hicieron 5 lavados con PBS-Tween 0.05 % previos a la incubación con el anticuerpo secundario acoplado a peroxidasa (goat anti mouse IgG-HRP), el cual se utilizó a una dilución 1:5000. Por último, se reveló con luminol (Immobilon Western Chemiluminescent HRP Substrate- Millipore) y la membrana se expuso a una placa radiográfica para obtener la impresión de la señal (Kodak). La incubación de la membrana con el sustrato fue de 5 min y la exposición con la placa fue de 30 s, la placa fue sumergida en solución reveladora y fijadora (CARESTREAM dental) por 15 s y lavada con agua corriente.

15.8 Ensayo de retardamiento

Los ensayos de retardamiento se llevaron a cabo utilizando aproximadamente 30 µg de la proteína uRNPS1, los cuales fueron incubados en 2 µL de buffer de unión BB 5X (50 mM HEPES-KOH pH 7.6, 40 mM MgCl₂, 0.1 mM DTT, 0.5 mM EDTA, 50 % Glicerol), 2 µg de ADN de esperma de salmón (Invitrogen) y 25 ng de oligo biotilado (Metodología 5.9.1). Las mezclas se llevaron a un volumen final de 20 µL y fueron incubadas durante 15 minutos a 4 °C, para permitir la formación de complejos.

Los complejos resultantes se analizaron en un gel nativo de poliacrilamida al 6 % (acrilamida-bisacrilamida 29:1) en TBE 0.5X (445 mM Tris- Base, 445 mM ácido bórico, 10 mM EDTA) el cual se corrió a 80 volts por 1 hora y media (en un carril se cargó buffer de carga para poder monitorear la migración) al término de la separación electroforética el gel se transfirió a una membrana de nylon previo equilibrio en buffer de transferencia (25 mM Tris-base, 192 mM pH Glicina 8.3), durante 15 minutos. La transferencia se realizó en una cámara de transferencia semi-seca Trans-blot semidry (Bio-Rad) durante 45 minutos a 15 V. Una vez transferidos los complejos RNA-proteína a la membrana de nylon, éstos fueron detectados con el sistema Chemiluminiscent Nucleic Acid Detection Module Kit (Thermo Fisher Scientific) siguiendo las especificaciones del proveedor. La membrana fue expuesta a una placa radiográfica hasta que apareciera la señal de los compuestos formados entre el ADN y uRNPS1.

Nota. - Para preparar el gel se utilizaron 1.5 mL de acrilamida, 1 mL de TBE 5X, 7.3 mL de agua, 10 µL de APS y 1 µL de TEMED. Este gel vacío fue pre-corrido a 150 v durante 30 minutos en TBE 0.5X.

15.8.1 Marcaje de oligos con biotina

Se utilizaron dos oligos diferentes para realizar este ensayo, se nombraron como oligo 1 y oligo 2 (CTT CTT CTT CCT TCT TCC TCC TTC TTC TTC y GAA₍₁₀₎ respectivamente).

Para llevar a cabo esta reacción se empleó buffer de TdT a una concentración final de 1X, 10 µL de oligo al 1 mM, Biotina -UTP a una concentración final de 0.5 µM (Roche) y 0.2 U de enzima TdT (fermentas); después de haber incubado la reacción durante 30 minutos a 37 °C se agregaron 5 µL de EDTA 0.5 para detener la reacción; posteriormente se lavó con 50 µL de cloroformo: alcohol isoamílico, se mezcló con ayuda de un vortex y se centrifugó durante 2 min a 1200 rpm, a

ESTUDIO DE PROTEÍNAS SR PUTATIVAS EN *Ustilago maydis*

continuación se recuperó la fase acuosa con un volumen igual de etanol absoluto y 1 μL de esperma de Salmón, se dejó precipitar durante 30 min a 4 °C. Por último, se centrifugó 10 min/1200 rpm, se eliminó el sobrenadante, se secó la pastilla y se resuspendió en 20 μL de agua libre de nucleasas

Nota. - La transferasa terminal deoxinucleotidil (TdT) se diluyó en buffer de reacción justo antes de usar a una concentración de 2 U/ μL .

16. RESULTADOS.

16.1 Identificación *in silico* de los homólogos para las proteínas SR reportadas para humano, en el genoma de *Ustilago maydis*.

Inicialmente se hizo un compendio de las proteínas SR reportadas en la literatura para *Homo sapiens*. A partir de esta búsqueda se obtuvo un total de 48 proteínas SR de humano encontradas en la literatura y se analizaron sus características generales, como la presencia de dominios funcionales, dominios estructurales, tamaño, etc., las cuales se resumen en la tabla 7.

Tabla 7. Proteínas SR humanas y sus características generales.

	Proteínas SR identificadas en humano	UniProt	No Acceso NCBI	Longitud (aa)	Dominios funcionales	Dominios estructurales
1	SRSF1/SF2/ASF	Q07955	NP_008855.1	248	RRM1, RRM2 Motivo de interacción con SAFB1	Región rica en G Región rica en RS
2	SRSF2/SC35	Q01130	NP_003007.2	221	RRM	Región rica en G, Región rica en RS
3	SRSF3/SRp20	P84103	NP_003008.1	164	RRM Motivo de interacción con NXF1	Región rica en RS
4	SRSF4/SRp75	Q08170	NP_005617.2	494	RRM1, RRM2	Región rica en G Región rica en RS
5	SRSF5/SRp40	Q13243	NP_001034554.1	272	RRM1, RRM2	Región rica en G Región rica en RS
6	SRSF6/SRp55	Q13247	NP_006266.2	344	RRM1, RRM2	Región rica en G Región rica en RS
7	SRSF7/9G8	Q16629	NP_001026854.1	238	RRM Dedo de Zinc CCHC Motivo de interacción con NXF1	Región rica en RS
8	SRSF8/SRp46	Q9BRL6	NP_115285.1	282	RRM	Región rica en R Región rica en RS
9	SRSF9/SRp30c	Q13242	NP_003760.1	221	RRM1, RRM2 Motivo de Interacción con SAFB1	Región rica en G Región rica en RS
10	SRSF10/SRp38	O75494	NP_473357.1	262	RRM	Región rica en RS
11	SRSF11/p54	Q05519	NP_004759.1	484	RRM	Región rica en G Región rica en RS Región rica en R
12	SRSF12	Q8WXF0	NP_542781.3	261	RRM	Región rica en RS
13	TRA-2 beta	P62995	NP_004584.1	288	RRM	2 Regiones ricas en RS
14	TRA-2 alpha	Q13595	NP_037425.1	282	RRM	2 Regiones ricas en RS
15	U1-70K	P08621	NP_003080.2	437	RRM	Región rica en R/D/E Región rica en G Región rica en R/D/E Región rica en RS
16	Urp	Q15696	NP_005080.1	482	RRM Dedo de Zinc C3H1 tipo 1 Dedo de Zinc C3H1 tipo 2	Región rica en RS Región rica en E

ESTUDIO DE PROTEÍNAS SR PUTATIVAS EN *Ustilago maydis*

17	RNPS1	Q15287	NP_006702.1	305	RRM Motivos de interacción con p110 Motivo de interacción con SRP54 Motivo de interacción con UPF2 y UPF3B Motivo de interacción con PNN y exón omitido Motivo de interacción con SAP18 y ACIN1 Motivo de interacción con TRA2B y de localización nuclear.	Región rica en RS Región rica en K Región rica en R/P
18	SRrp53	Q961Z7		334		Hélice superenrolladas Región rica en RS
19	SRrp86	Q8WXA9	NP_631907.1	508	RRM	Región rica en R/E/K/S Región rica en RS
20	SRrp130	Q8TF01	NP_001309341.1	805		3 hélices superenrolladas Región rica en Q Región rica en P Región rica en S Región rica en RS
21	SRrp129	Q99590	NP_004710.2	1463	Dedo de zinc tipo RING degenerado	Región rica en R Región rica en P Región rica en RS
22	XE7	Q02040	NP_005079.2	695	RRM Motivo de unión a la subunidad PKA-RI y PKA-RII Motivo de unión a la subunidad PKA-Riα	Región rica en R Región rica en RS
23	RBM5	P52756	NP_005769.1	815	RRM1, RRM2, Dedo de Zinc tipo RanBP2 Dedo de Zinc tipo C2H2 atípico Parche-G Motivo de interacción con U2AF2 Motivo de interacción con ACIN1, SNRPB, SNRPN, SNRNP70 y SNRNP200	Región rica en RS
24	RBM23	Q86U06	NP_001070819.1	439	RRM1, RRM2	Región rica en R Región rica en A
25	RMB39 / HCC1/CAPER	Q14498	NP_909122.1	530	RRM1, RRM2, RRM3 Motivo de interacción con Jun Motivo de activación Motivo de interacción con ESR1 y ESR2 Motivo de interacción con NCOA6	Región rica en RS
26	Proteína SR140 asociada a U2	O15042	NP_001073884.1	1029	RRM Motivo SURP CID	3 Hélices superenrolladas Región rica en RS Región rica Región rica en P Región rica en D Región rica en E
27	DHX38/PRP16	Q92620	NP_054722.2	1227	Helicasa de unión a ATP Helicasa en el C-terminal Caja DEAH	Región rica en RS
28	DDX46/hPRP5	Q7L014	NP_055644.2	1031	Helicasa unión a ATP Helicasa C terminal	Hélice superenrollada Caja DEAD Motivo Q Región rica en R
29	SRm160	Q8IYB3	NP_005830.2	904	PWI Motivo de localización en matriz y	Región rica en RS Región rica en P

ESTUDIO DE PROTEÍNAS SR PUTATIVAS EN *Ustilago maydis*

					speckles	
30	SRm300	Q9UQ35	NP_057417.3	2752	Motivo de unión a RNA	Hélice superenrollada Región rica en RS Región rica en A
31	HsWAP	Q12872		951	Motivo SURP 1 Motivo SURP 2	Región rica en SR
32	PININ	Q9H307	NP_002678.2	717	Motivo de interacción con RNPS1 Motivo de interacción con KRT8, KRT18 v KRT19 Motivo de interacción con complejo PSAP Motivo de interacción con PPIG	4 hélices superenrolladas Región rica en E Región rica en Q Región rica en RS
33	LUC7B1	Q9NQ29	NP_958815.1	371		2 hélices superenrolladas Región rica en RS
34	Acinus	Q9UKV3	NP_055792.1	1341	SAP Motivo de interacción con RNPS1 y SAP18 y formación del complejo ASAP	Región rica en E Región rica en S Región rica en R/D/E/K Región rica en RS
35	SR-A1	Q9H7N4	NP_067051.2	1312	Motivo de interacción con el dominio CTD de POLR2A	Región rica en P Región rica en SR Región rica en E
36	ZNF265	O95218	NP_976225.1	330	Dedo de Zinc RanBP2-tipo 1 Dedo de Zinc RanBP2-tipo 2 Motivo de señalización nuclear	Región rica en RS
37	Prp22	Q14562	NP_004932.1	1220	Motivo S1 Helicasa de unión a ATP Helicasa en el C-terminal Caja DEAD	Región rica en K Región rica en RS
38	U5-100k	Q9BUQ8	NP_004809.2	820	Helicasa de unión a ATP Helicasa en el C-terminal Motivo Q Caja DEAD	Región rica en R Región rica en E Región rica en RS
39	Prp4k	Q13523	NP_003904.3	1007	Proteína cinasa	Región rica en RK Región rica en H Región rica en RS
40	CrkRS	Q9NYV4	NP_057591.2	1490	Proteína cinasa	Región rica en A Región rica en P Región rica en RS
41	CDC2L5	Q14004	NP_003709.3	1512	Proteína cinasa	Región rica en RS
42	CIR1	Q86X95		450	Motivo de interacción con RBPJ Motivo de interacción con RP9 Motivo de señalización nuclear	Región rica en RK Región rica en RS
43	Cyclin L1	Q9UK58	NP_064703.1	526	Ciclina like-1 Ciclina like-2	Región rica en RS
44	Cyclin L2	Q96S94	NP_112199.2	520	Ciclina like-1 Ciclina like-2	Región rica en RS
45	SR-cyp	Q13427	NP_004783.2	754	PPIasa tipo ciclopilina	Región rica en RS Región rica en RK
46	Proteína 27kDa asociada a U4/U6 · U5 tri-snRNP	Q8WVK2	NP_006848.1	155		Región rica en R Región rica en RS
47	U2AF65	P26368	NP_009210.1	475	Motivo de interacción con PRPF19	Región rica en RS
48	U2AF35	Q01081	NP_006749.1	240	Motivo RRM	Región rica en RSG

Según la clasificación más reciente, las proteínas del grupo 1 (SFRS1, SFRS4, SFRS5, SFRS6, y SFRS9) tienen dos dominios RMM, y el segundo dominio RRM tiene un motivo heptapéptido SWQDLKD invariable y conservado evolutivamente. En contraste, las proteínas de los grupos 2 (SFRS12, y SR140) y 3 (SFRS2, SFRS3 y SFRS7) tienen un único dominio RRM y carece del motivo SWQDLKD (Manley *et al.*, 2010); sin embargo, algunas proteínas contienen otro tipo de dominios que permiten su interacción con RNA, como es el caso del dominio PWI y los dedos de Zinc. Como puede observarse de los datos mostrados en la tabla 7, otras proteínas SR identificadas además de presentar dominios RRM poseen dominios de interacción con otras proteínas, por ejemplo, la proteína SRSF7/9G8 que posee dominio de unión a la proteína NXF1 también conocida como TAP, la cual está implicada en la exportación de ARNm del núcleo hacia el citoplasma en humano (Herold *et al.*, 2000), sin embargo, la característica que comparten todas estas proteínas es el dominio estructural RS. En este sentido cabe destacar que no en todos los casos el dominio se encuentra en el extremo amino terminal, en ocasiones estos dominios pueden estar distribuidos a lo largo de la proteína y pueden contener hasta dos dominios RS como es en el caso de la proteína RNPS1, dominio que permite la interacción entre proteínas (Long & Cáceres, 2009). Es interesante resaltar que en los dominios estructurales se evidenciaron no solo repetidos ricos en R y S, sino también repetidos ricos en otros aminoácidos como A, E, D, RK, K, etc. Aunque de manera clásica se considera que las proteínas SR presentan regiones ricas en ambos residuos, ahora se ha observado que de manera individual los residuos de arginina o serina son susceptibles a sufrir eventos de fosforilación u otro tipo de modificación postraducciona para activar cierta función de las proteínas SR, por ejemplo, algunos factores de *splicing* como SRSF1 (SF2/ASF) son fosforilados en los residuos de serina, provocando que se concentren en las espículas nucleares, estos hallazgos se realizaron al observar la localización de proteínas SR a las cuales se les había eliminado el dominio RS y proteínas SR completas (Rossi *et al.*, 1996). Sin embargo, se necesitan realizar diversos estudios para demostrar la funcionalidad de residuos particulares para la actividad de una proteína SR específica.

Una vez identificadas las proteínas SR reconocidas para humano, se realizó una búsqueda del homólogo correspondiente en las secuencias anotadas en la base de datos para *Ustilago maydis*. Los resultados obtenidos para esta comparación se muestran en la tabla 8, en donde se resumen los homólogos identificados en *U. maydis* para proteínas SR reportadas en *H. sapiens*, mostrando el porcentaje de identidad obtenido para cada péptido. En algunos casos, se encontró que un mismo locus de *U. maydis* podía corresponder al homólogo para más de una proteína SR humana; dichos casos se resaltan con color en la siguiente tabla.

Tabla 8. Secuencias homólogas para proteínas SR de humanos en *Ustilago maydis*.

Proteína SR <i>H. sapiens</i>	Tamaño aa <i>H. sapiens</i>	Locus en <i>U. maydis</i>	Identidad %	Función asignada
SRSF1	248	UMAG_02552	28/78 (36 %)	Proteína hipotética
SRSF2	221	UMAG_02412	31/76 (41 %)	Proteína hipotética
SRSF3	164	UMAG_10162	33/89 (37 %)	Proteína hipotética
SRSF4	494	UMAG_15022	76/190 (40 %)	Factor de <i>splicing</i> , rico en arginina/serina
SRSF5	272	UMAG_15022	71/190 (37 %)	Factor de <i>splicing</i> , rico en arginina/serina
SRSF6	344	UMAG_10375	66/209 (32 %)	Proteína hipotética
SRSF7	238	UMAG_01265	24/68 (35 %)	Proteína hipotética
SRSF8	282	UMAG_06193	49/147 (33 %)	Nucleolina
SRSF9	221	UMAG_02469	31/109 (28 %)	Proteína hipotética
SRSF10	262	UMAG_06193	23/78 (29 %)	Nucleolina
SRSF11	484	UMAG_04390	41/138 (30 %)	Proteína hipotética
SRSF12	261	UMAG_00516	38/ 119 (32 %)	Proteína nucleolar 4
TRA 2 alfa	282	UMAG_10239	51/101 (50 %)	Proteína transformadora 2
TRA 2 beta	288	UMAG_10239	62/138 (45 %)	Proteína transformadora 2
SRp86	508	UMAG_05363	22/78 (28 %)	Factor de <i>splicing</i> U2AF subunidad 65 kDa
SRp129	1463	UMAG_11405	12/30 (40 %)	Complejo promotor del anafase subunidad 11
SRp130	805	UMAG_00275	12/25 (48 %)	ADN polimerasa
SRm160	904	UMAG_12176	69/128 (54 %)	Proteína de la matriz con repetidos serina/ arginina
SRm300	2752	UMAG_06397	58/164 (35 %)	Proteína hipotética
SR-A1	1312	UMAG_05527	27/102 (26 %)	Proteína hipotética
Sr-cyp	754	UMAG_06259	92/175 (53 %)	U-snRNP putativa asociada a ciclofilina
SR140-U2	1029	UMAG_10477	184/719 (26 %)	Proteína hipotética
27 kDa peptide tri-snRNP-U4/U6 · U5	155	UMAG_04183	37/81 (46 %)	Proteína 3 asociada a tri-snRNP U4/U6 · U5
RBM5	815	UMAG_02655	58/187 (31 %)	Proteína hipotética
RBM23	439	UMAG_00164	101/283 (36 %)	Proteína 39 de unión a ARN
RBM39/ HCC1	530	UMAG_00164	150/429 (35 %)	Proteína 39 de unión a ARN
DHX38 o PRP16	1227	UMAG_04188	620/1236 (50 %)	Factor de <i>splicing</i> ATP dependiente de ARN Helicasa

ESTUDIO DE PROTEÍNAS SR PUTATIVAS EN *Ustilago maydis*

DDX46	1031	UMAG_01174	319/583 (55 %)	ARN Helicasa dependiente de ATP
Prp4k	1007	UMAG_11912	191/347 (55 %)	Proteína cinasa serina/treonina
Prp22	1220	UMAG_03936	696/1236 (56 %)	ARN Helicasa dependiente de ATP
U1-70K	437	UMAG_04775	105/233 (45 %)	Ribonucleoproteína pequeña nuclear de 70 kDa
U5-100k	488	UMAG_10666	324/721 (45 %)	ARN Helicasa dependiente de ATP
Urp	482	UMAG_03893	65/166 (39 %)	Factor de <i>splicing</i> U2AF subunidad 35 kDa
CrkRS	1490	UMAG_04925	138/302 (46 %)	Subunidad alfa Cinasa CTD
CDC2L5	1512	UMAG_04925	144/327 (44 %)	Subunidad alfa Cinasa CTD
Cyclin L1	526	UMAG_10398	92/256 (36 %)	Proteína hipotética
Cyclin L2	226	UMAG_10398	92/256 (36 %)	Proteína hipotética
U2AF35	240	UMAG_03893	118/212 (56 %)	Factor de <i>splicing</i> U2AF subunidad 35 kDa
U2AF65	475	UMAG_05363	171/367 (47 %)	Factor de <i>splicing</i> U2AF subunidad 65 kDa
XE7	695	UMAG_10161	47/171 (27 %)	Proteína hipotética
RNPS1	305	UMAG_03178	60/186 (32 %)	Proteína de unión a ARN con dominio rico en serinas
Pinin	717	UMAG_02686	16/46 (35 %)	Proteína hipotética
LUC7B1	371	UMAG_10299	86/224 (38 %)	Proteína hipotética
Acinus	1341	UMAG_00992	16/40 (40 %)	MAP cinasa y fosfatasa
ZNF265	330	UMAG_02423	11/21 (52 %)	Proteína hipotética

De las 48 proteínas SR reportadas para humano, encontramos el homólogo respectivo para 45 de ellas (94 %) en *U. maydis* (Tabla 8). En algunos casos, dos proteínas SR diferentes presentaban el mismo homólogo en *U. maydis*, por ejemplo, TRA2 alfa y TRA2 beta. Estas observaciones sugieren que la diversidad de proteínas en *U. maydis* es menor a la que se presenta en humano; sin embargo, también indica que una misma proteína en un evento determinado pudiera llegar a suplir la actividad de dos o más proteínas y que podrían tener una función redundante en algunos casos.

Por otra parte, se observó que el porcentaje de identidad para los homólogos identificados se encuentra en el rango de 20-60 %. Los porcentajes de identidad más altos se observaron para los homólogos de las proteínas Prp22 (56 %), U2AF35 (56 %), Prp4 (55 %), DDX46 (55 %), SRm160 (54 %), Sr-cyp (53 %), ZNF265 (52 %), DHX38 (50 %) y Tra2 alfa (50 %). Sin embargo, los porcentajes para los homólogos de proteínas SR clásicas oscilan entre el 28 y 40 %. También se observó que en los casos donde dos proteínas SR diferentes presentaban el mismo homólogo en *U. maydis*, los porcentajes de identidad para ambos casos son muy parecidos. Finalmente, para algunos de los locus identificados se muestra ya una función putativa en la base de datos de *U.*

maydis. Aunque la función asignada no incluye la de ser factores SR, algunas de las funciones descritas pueden ser relevantes en el metabolismo del ARN.

16.2I Identificación de proteínas que contienen dominios RRM y RS característicos de las proteínas SR en la anotación de genoma de *Ustilago maydis*

Ya que los primeros resultados solo se fundamentaron en el hecho de obtener los homólogos de *Ustilago maydis* a partir de proteínas SR reportadas en humano, existía la posibilidad de que *U. maydis* tuviera otras proteínas SR además de las que se pudieran compartir con humano. Para resolver esta incógnita y completar una búsqueda exhaustiva que incluyera a todas las posibles proteínas SR en el hongo, se realizó una segunda búsqueda bioinformática. Empleando este abordaje se identificaron en la base de datos para *U. maydis* 60 proteínas que poseen dominios RRM, para las cuales se identificó de forma manual si poseían además dominios ricos en residuos de serina, arginina o ambos. En la tabla 9 se resume la información para las 60 proteínas identificadas, incluyendo el número de dominios RRM que poseen, también se agregó la información que corresponde a las regiones ricas en S, R y RS; donde **SR-amino**, **SR-carboxilo** y **SR-toda** significa que los repetidos SR se encuentran en el extremo amino, carboxilo o a lo largo de toda la proteína; de manera similar se indican las regiones ricas en R y S. Estos datos sugieren que *U. maydis* podría llegar a tener una gran variedad de estas proteínas cumpliendo diferentes funciones.

Tabla 9. Proteínas con dominios RRM y RS identificadas en la base de datos de *Ustilago maydis*.

	Locus	Dominio RRM	Dominio S, R o SR		Locus	Dominio RRM	Dominio S, R o SR
1	UMAG_00164.1	3	SR-amino	31	UMAG_03178.1	1	S-carboxilo
2	UMAG_00254.1	2	S-carboxilo	32	UMAG_03208.1	1	S-amino
3	UMAG_10966.1	2	SR-amino	33	UMAG_10836.1	2	S-toda
4	UMAG_00516.1	6		34	UMAG_11084.1	1	S-toda
5	UMAG_00601.1	2	S-amino	35	UMAG_03494.1	4	
6	UMAG_00667.1	1	S-carboxilo	36	UMAG_10477.1	1	SR-toda
7	UMAG_00716.1	2		37	UMAG_03724.1	2	R-carboxilo
8	UMAG_12111.1	2	S-carboxilo	38	UMAG_03771.1	1	S-amino
9	UMAG_10793.1	1	SR-carboxilo	39	UMAG_11511.1	1	SR-amino
10	UMAG_01182.1	3	S-carboxilo	40	UMAG_03939.1	1	R-carboxilo
11	UMAG_01265.1	1	S-amino	41	UMAG_03952.1	1	R-carboxilo
12	UMAG_01271.1	2	S-amino	42	UMAG_04149.1	1	S-amino
13	UMAG_10239.1	1	SR-toda	43	UMAG_04269.1	1	R-carboxilo
14	UMAG_01414.1	1	S-amino	44	UMAG_04299.1	1	SR-toda
15	UMAG_01536.1	1	S-carboxilo	45	UMAG_11148.1	1	SR-toda

ESTUDIO DE PROTEÍNAS SR PUTATIVAS EN *Ustilago maydis*

16	UMAG_02292.1	3	<u>R-carboxilo</u>	46	UMAG_04390.1	1	SR-amino
17	UMAG_02346.1	4		47	UMAG_04398.1	4	SR-toda
18	UMAG_02401.1	1	<u>R-carboxilo</u>	48	UMAG_10903.1	1	SR-toda
19	UMAG_02412.1	1		49	UMAG_04679.1	2	
20	UMAG_02420.1	2	SR-carboxilo	50	UMAG_10674.1	2	SR-carboxilo
21	UMAG_02469.1	1	<u>R-carboxilo</u>	51	UMAG_04775.1	1	SR-toda
22	UMAG_10162.1	2	<u>R-amino</u>	52	UMAG_04835.1	2	S-toda
23	UMAG_02472.1	2	S-amino	53	UMAG_12284.1	1	SR-amino
24	UMAG_02552.1	2	S-toda	54	UMAG_05363.1	3	SR-amino
25	UMAG_02609.1	1	SR-amino	55	UMAG_11178.1	2	S-amino
26	UMAG_02696.1	1	S-toda	56	UMAG_10723.1	2	S-carboxilo
27	UMAG_02881.1	1		57	UMAG_11591.1	1	S-toda
28	UMAG_12203.1	1	S-carboxilo	58	UMAG_06090.1	1	S-toda
29	UMAG_02994.1	2	S-toda	59	UMAG_06183.1	3	S-toda
30	UMAG_10375.1	4	SR-amino	60	UMAG_06193.1	2	S-toda

Se encontró que de las 60 proteínas con dominios RRM, 18 presentaban dominios SR notables, mientras que 27 proteínas contenían repetidos de Serina; en menor cantidad se identificaron 8 proteínas que contenían repetidos de arginina y finalmente 7 proteínas con dominios RRM no contenían ninguna región rica en estos residuos de manera aparente; de tal forma que en este abordaje se obtuvieron 53 posibles proteínas SR en el genoma de nuestro modelo de estudio. De manera adicional se identificó la localización de cada uno de los dominios encontrados para cada uno de los homólogos en *U. maydis*; es decir, se ubicó la región en la cual se encontraban cada uno, ya fuera en el extremo carboxilo, en el extremo amino o a lo largo de toda la proteína.

En un intento por identificar a las proteínas SR que tuvieran una mayor posibilidad de ser funcionales en *U. maydis*, se decidió realizar una correlación global de los resultados obtenidos en ambos abordajes para identificar a las proteínas SR putativas que presenten un mejor sustento. Al correlacionar los dos abordajes empleados se llegó a un total de 15 coincidencias que fueron identificadas por las dos metodologías empleadas y que podrían corresponder a proteínas SR putativas, las cuales se muestran en la tabla 10.

Tabla 10. Proteínas SR putativas en *Ustilago maydis*.

	Locus en <i>U. maydis</i>	Proteína SR en humano	Score (Bits)
1	UMAG_02552	SRSF1	48.5
2	UMAG_02412	SRSF2	56.6102
3	UMAG_01265	SRSF3	44.2838
4	UMAG_10375	SRSF4 SRSF5	51.6026 51.2174
5	UMAG_04775	SRSF6 U170K	49.2914 129.413
6	UMAG_04149	SRSF7	42.743
7	UMAG_06193	SRSF8 SRSF10	59.6918 54.299
8	UMAG_02881	SRSF9	45.8246
9	UMAG_04390	SRSF11	41.9726
10	UMAG_10239	TRA2 β TRA2 α SRSF12	112 111 51.6026
11	UMAG_00164	RBM23 RBM39	150.984 130.183
12	UMAG_02401	SRp86, SRp508	43.5134
13	UMAG_03939	U2AF65	45.8246
14	UMAG_03178	RNPS1	53.9138
15	UMAG_10477	proteína SR140 asociada a U2	162.925

Es importante destacar que a partir de esta correlación de resultados se encontró que las 15 proteínas obtenidas corresponden a proteínas SR clásicas o canónicas. Esta observación nos permite proponer que el estudio de cada una de estas proteínas arrojaría resultados importantes y novedosos. Sin embargo, para continuar con este trabajo se eligieron solo tres de los quince homólogos identificados, para los cuales se procedió a la obtención de las proteínas recombinantes respectivas. Para dicho fin se eligieron los homólogos para las proteínas SRSF1, ya que es una de las proteínas SR que más ha sido estudiada, RBM39 que participa en mecanismos de NMD y la proteína RNPS1 que forma parte de complejo EJC. Cabe mencionar que las proteínas SRSF2, SRSF9, TRA2 β , TRA2 α , SRSF12 no se tomaron en cuenta porque poseen intrones en su secuencia. Aunque este hallazgo resulta interesante porque existen diversos factores de splicing que contienen intrones y cuya expresión es regulada a su vez por splicing, en este caso en particular la presencia de los intrones podría dificultar el desarrollo experimental y la obtención de las proteínas recombinantes planteada como objetivo de este proyecto.

Finalmente, a partir de esta correlación se identificaron un total de 78 proteínas SR putativas en *U. maydis*, las cuales presentan una estructura y organización variables. Aunque resulta interesante el estudio y un análisis a profundidad de estas proteínas, es necesario un trabajo posterior para caracterizarlas funcionalmente.

16.3 Comprobación de los genes codificantes para las proteínas uSRSF1, uRMB39 y uRNPS1 en *Ustilago maydis* y generación de sus respectivas proteínas recombinantes.

16.3.1 Extracción de ADN y amplificación de fragmentos de ADN por medio de PCR.

A partir de los homólogos identificados se eligieron únicamente a tres locus para los cuales se obtuvieron las proteínas recombinantes uRMB39, uSRSF1 y uRNPS1, correspondientes a los locus UMAG_00164, UMAG_02552 y UMAG_3178 y se diseñaron los *primers* que amplificaron el marco abierto de lectura completo (Metodología 4.5.1).

El ADN extraído de la cepa FB1 de *Ustilago maydis* (Metodología 4.5.2), fue visualizado por medio de una electroforesis en gel de agarosa (Metodología 4.5.5) y se utilizó como templado para realizar las reacciones de PCR (Metodología 4.5.3), obteniendo así los siguientes amplificados mostrados en la figura 9.

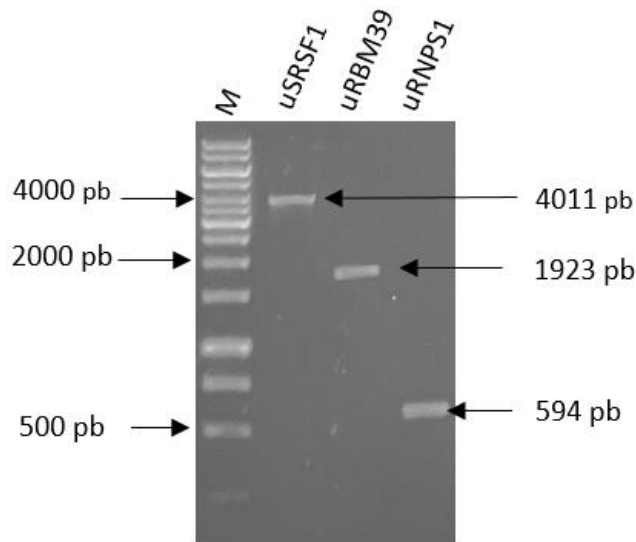


Figura. 9 Productos amplificados por PCR. Para la proteína uSRSF1, uRMB39 y uRNPS1 se obtuvieron los amplificados de 4011 pb, 1923 pb y 594 pb, respectivamente.

16.3.2 Clonación en el vector pJET 1.2.

Los productos amplificados por PCR se ligaron en el vector pJET 1.2 (Metodología 4.5.6). Considerando que el vector presenta un tamaño de 2974 pb y los amplificados para uRNPS1 (594 pb), uRMB39 (1923 pb) y uSRSF1 (4011 pb), el tamaño total para cada construcción fue de 3568 pb, 4897 pb y 6985pb respectivamente. Una vez realizadas las ligaciones se electroporaron en células de *E. coli* DH5 α electrocompetentes (Metodología 4.6.1 y 4.6.2) o bien se transformaron en células quimiocompetentes de *E. coli* DH5 α (Metodología 4.6.3 y 4.6.4), según sea el caso.

La clonación en este vector nos sirvió de reservorio para posteriormente realizar la subclonación, posteriormente se realizó la extracción de plásmido (Metodología 4.5.4) y las clonas obtenidas fueron sometidas a un análisis de restricción (Metodología 4.5.9), para saber si el inserto se encontraba en la orientación 5' a 3'.

Los mapas de restricción, así como las reacciones de digestión enzimática que permitieron confirmar la identidad de cada clona se muestran a continuación para uSRSF1 (Fig. 10), uRBM39 (Fig. 11) y uRNPS1 (Fig. 12).

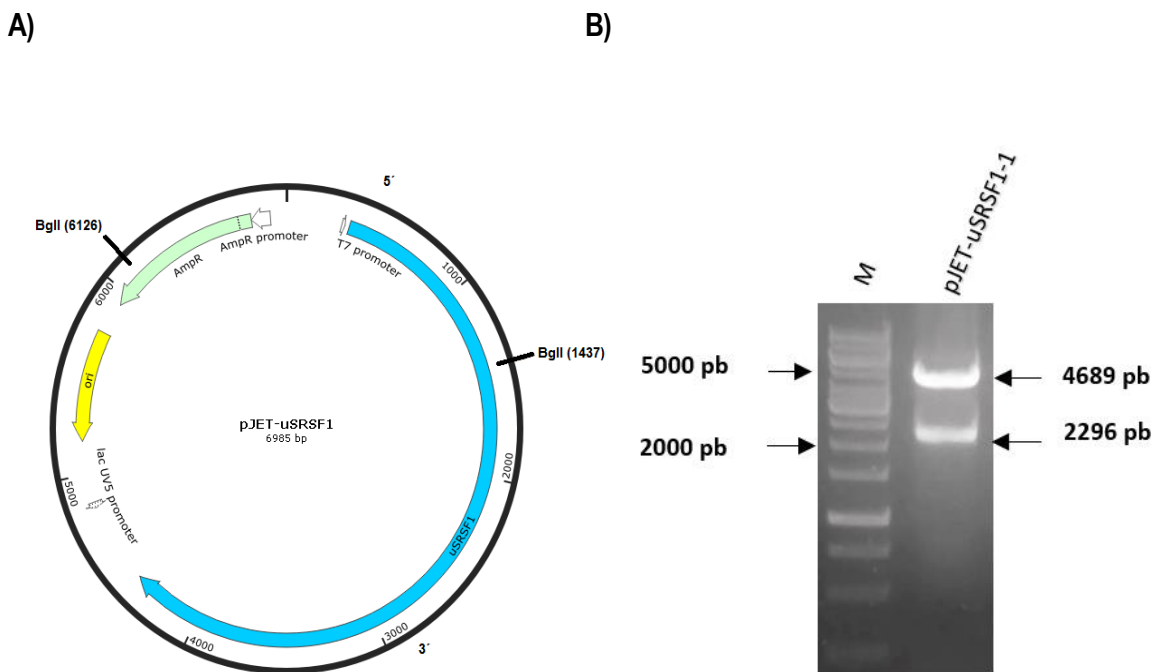


Figura 10. Digestión de la construcción pJET-uSRSF1 con la enzima BglII. A) Mapa de la construcción pJET-uSRSF1, se indican los sitios de corte de la enzima BglII y el inserto uSRSF1 de 4011 pb. B) Electroforesis en gel de agarosa al 1 %, se muestran las bandas de 4689 pb y 2296 pb correspondientes al análisis restricción, que indican la dirección 3' - 5' del inserto dentro del plásmido.

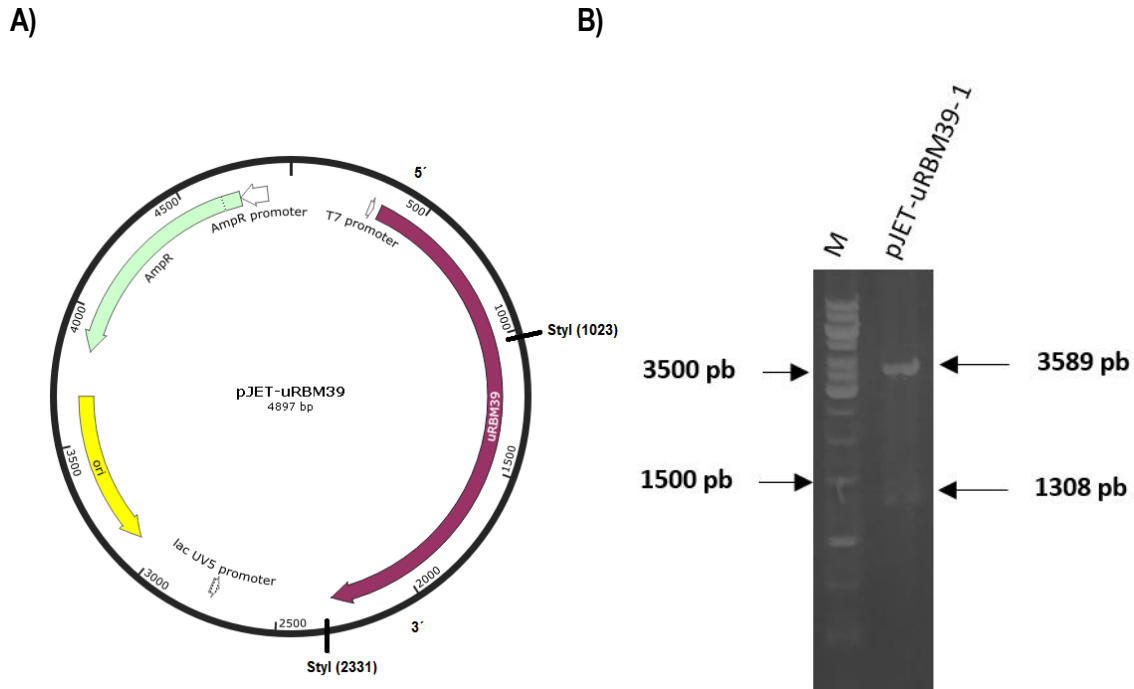


Figura. 11 Digestión de la construcción pJET-uRBM39 con la enzima StyI. A) Mapa de la construcción pJET-uRBM39, se indica el sitio de corte de la enzima StyI, y el inserto uRBM39 de 1923 pb. B) Electroforesis en gel de agarosa al 1 %, se muestran las bandas de 3589 pb y 1308 pb correspondientes al análisis de restricción, que indican la dirección 3´-5´ del inserto dentro del plásmido.

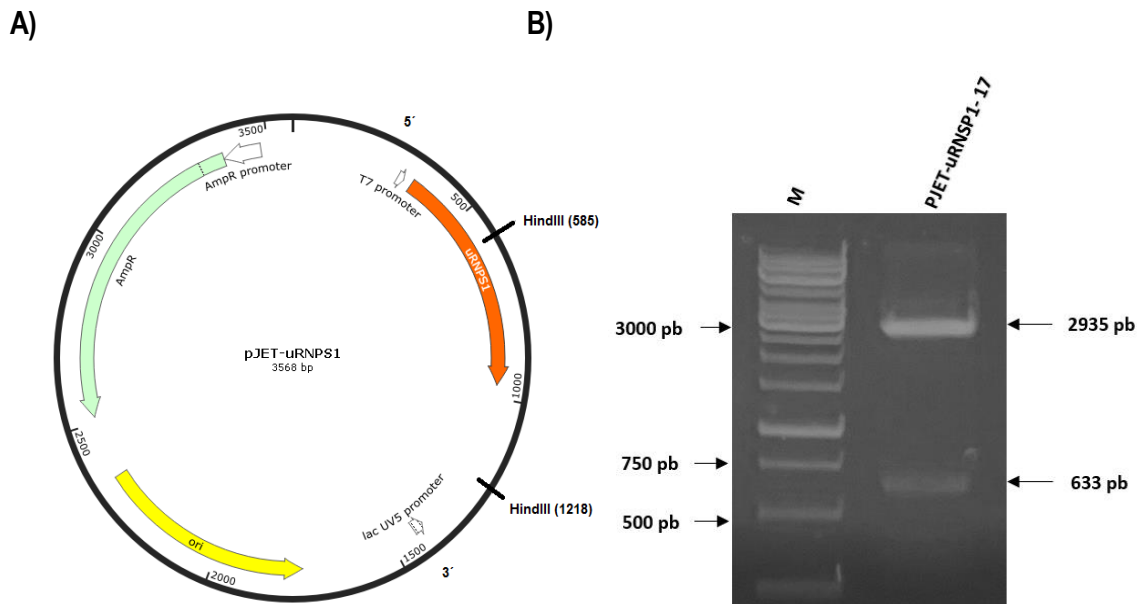


Figura 12. Digestión de la construcción pJET-uRNPS1 con la enzima HindIII. A) Mapa de la construcción pJET-uRNPS1, se indica el sitio de corte de la enzima HindIII y el inserto uRNPS1 de 594 pb. B) Electroforesis en gel de agarosa al 1 %, se muestran las bandas de 2935 pb y 633 pb correspondientes al análisis restricción, que indican la dirección 3´-5´ del inserto dentro del plásmido.

16.3.3 Subclonación en el vector pPROEX HTc.

Una vez obtenidas las clonas en orientación 5'-3' dentro del vector pJET se realizó la subclonación en el vector pPROExHT. Para poder realizarlo, primero se digirió el ADN plásmidico obtenido a partir de 5 mL de cultivo (Metodología 4.5.4) para una clona de pJET previamente identificada por análisis de restricción. Por otro lado, se digirió el vector pPROEx HTc de manera independiente. En ambos casos, se utilizaron las enzimas NotI-XbaI para que la clonación fuera dirigida. Posteriormente, tanto el vector como el inserto fueron cortados de gel y purificados (Metodología 4.5.8) para luego ser ligados en una proporción 1:3 (Metodología 4.5.7).

Una vez realizadas las ligaciones, se electroporaron en células de *E. coli* DH5 α electrocompetentes (Metodología 4.6.1 y 4.6.2) o bien se transformaron en células quimiocompetentes de *E. coli* DH5 α (Metodología 4.6.3 y 4.6.4), dependiendo de la construcción.

Posteriormente se realizó la extracción de plásmido (Metodología 4.5.4) y las clonas obtenidas fueron sometidas a un análisis de restricción (Metodología 4.5.9) para confirmar la correcta inserción del fragmento en orientación 5'a 3'. El mapa de restricción y los análisis correspondientes se muestran para uSRSF1 (Fig. 13), para uRBM39 (Fig. 14) y para uRNPS1 (Fig. 15), respectivamente.

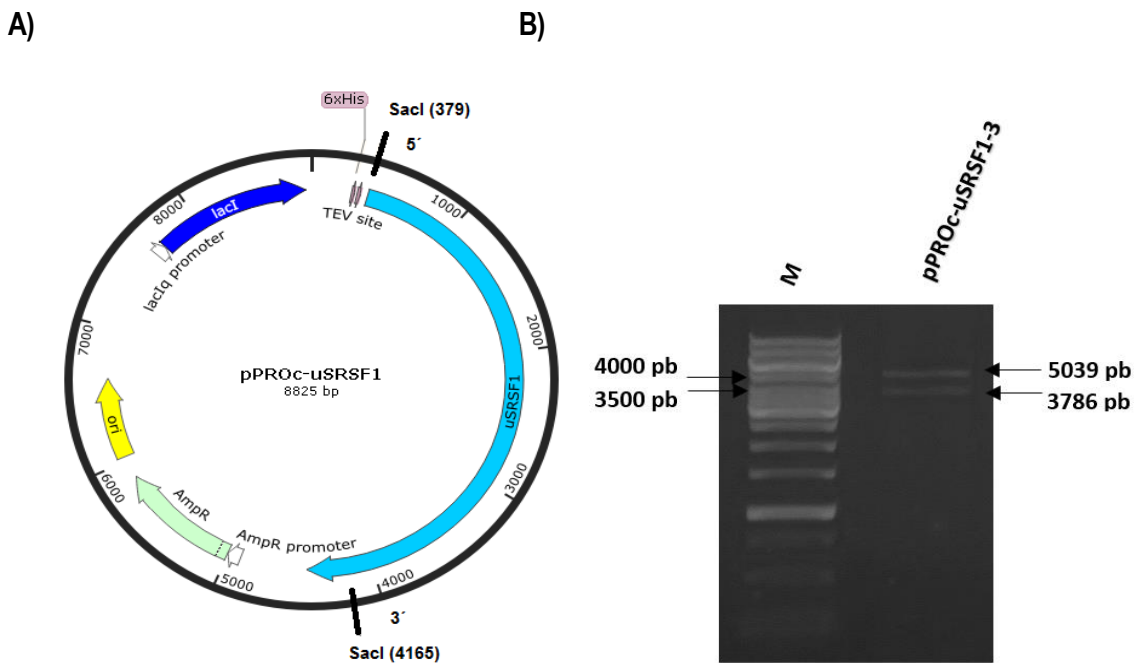


Figura. 13 Análisis de restricción de pPROc-uSRSF1 con la enzima SacI. A) Mapa de la construcción pPROc-uSRSF1, se indica el sitio de corte de la enzima SacII. B) Electroforesis en gel de agarosa al 1 %, se muestran las bandas de 5039 pb y 3786 pb correspondientes al análisis restricción, que indican la dirección 3' - 5' del inserto dentro del plásmido.

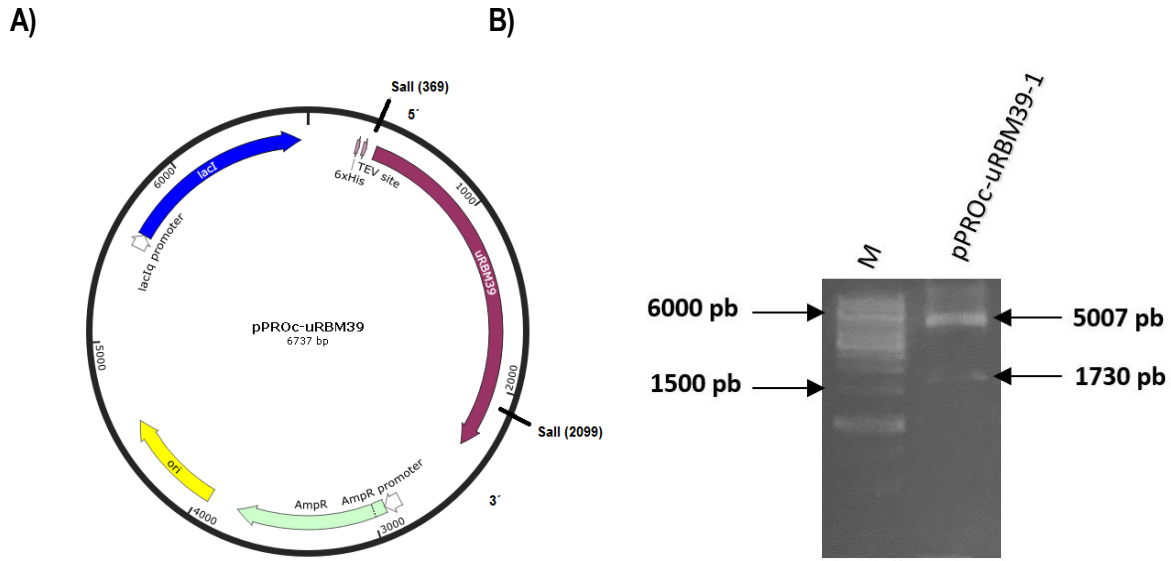


Figura. 14 Análisis de restricción de pPROc-uRBM39 con la enzima Sall. A) Mapa de la construcción pPROc-uRBM39, se indica el sitio de corte de la enzima Sall. B) Electroforesis en gel de agarosa al 1 %, se muestran las bandas de 5007 pb y 1730 pb correspondientes al análisis restricción, que indican la dirección 3' - 5' del inserto dentro del plásmido.

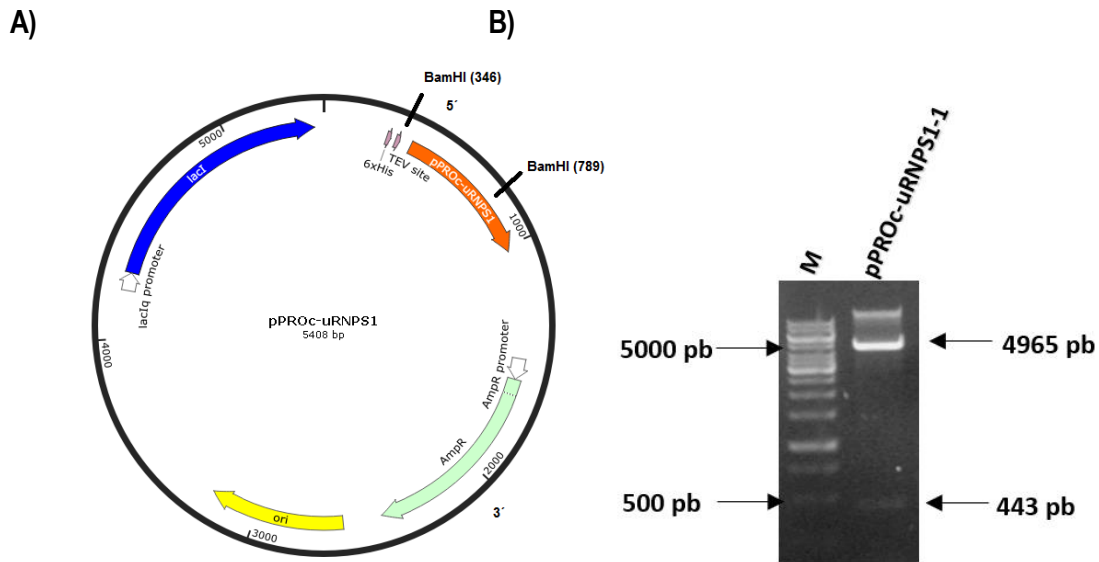


Figura 15 Análisis de restricción de pPROc-uRNPS1 con la enzima BamHI. A) Mapa de la construcción pPROc-uRNPS1, se indica el sitio de corte de la enzima BamHI. B) Electroforesis en gel de agarosa al 1 %, se muestran las bandas de 4965 pb y 443 pb correspondientes al análisis restricción, que indican la dirección 3' - 5' del inserto dentro del plásmido.

16.3.4 Inducción de proteínas recombinantes (IPTG).

La clonación en este vector nos sirvió como sistema de expresión ya que posee el gen *LacI*, el cual favorece la transcripción de la proteína de interés en presencia de lactosa; en el laboratorio se utilizó isopropil tgalactósido (IPTG) que es un análogo de la lactosa no hidrolizable, el cual permitirá mantener una concentración constante durante todo el experimento de inducción. Por otro lado, el vector empleado permite la adición de una cola de 6 histidinas en el extremo 5', esta característica es de suma importancia ya que favorece los procesos de purificación de la proteína recombinante y su identificación mediante el empleo de anticuerpos. Para obtener las proteínas recombinantes se probaron diferentes condiciones de temperatura, tiempo, presencia de ciertos osmolitos (sacarosa, manitol) y concentración de IPTG.

Una vez que se llevó a cabo la inducción de las proteínas recombinantes uSRSF1 (Fig. 16), uRBM39 (Fig. 17) y uRNPS1 (Fig. 18 y Fig. 19), se verificó su producción mediante una electroforesis en gel desnaturalizante de poliacrilamida.

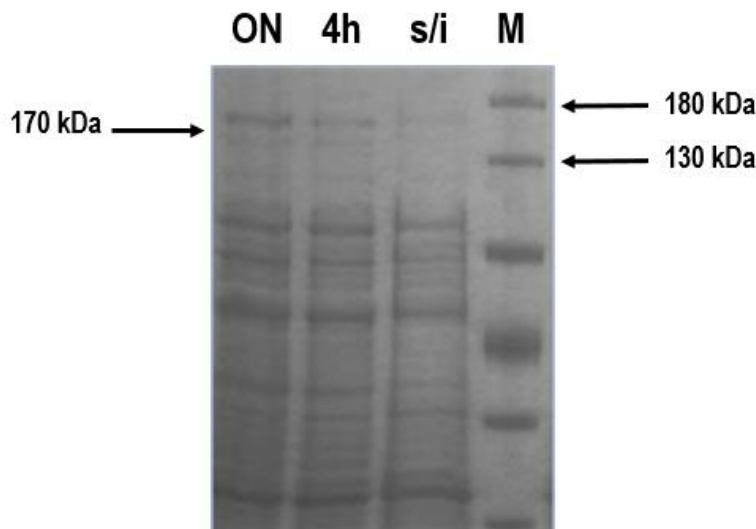


Figura 16. Identificación en SDS PAGE de la proteína recombinante uSRSF1. Fotografía de SDS Page al 8 %: en el carril 1 se cargó la inducción después de toda un anoche (ON), la flecha indica la banda que corresponde a la proteína de 170 kDa, en el segundo carril (4h) l inducción correspondiente a 4 horas, en el carril 3 (s/i) se cargó la muestra sin inducir y el ultimo carril (M) se encuentra el marcador de peso molecular; para realizar la inducción se utilizó una concentración final de 0.3 mM de IPTG.

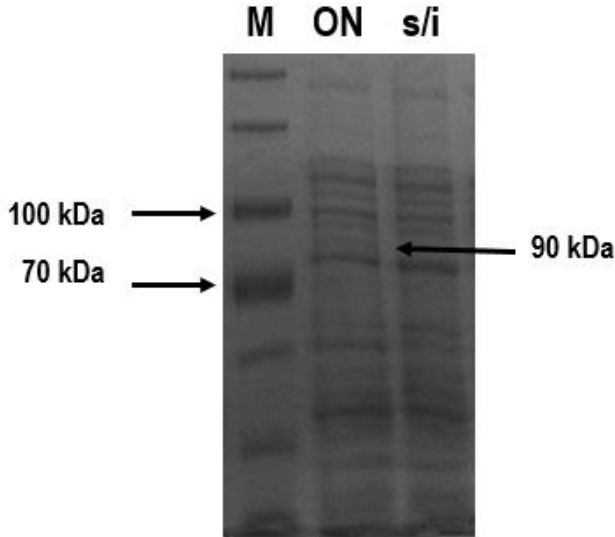


Figura 17. Identificación en SDS PAGE de la proteína recombinante uRBM39. Fotografía de SDS Page al 10 %, donde en el primer carril se muestra el marcador de peso molecular (M), en el segundo carril se cargó la inducción correspondiente a toda la noche (ON) en la cual se observa una banda que corresponde aproximadamente a 90 kDa, en el último carril se cargó la muestra sin inducir (s/i); para realizar esta inducción se utilizó una concentración final de 0.3 mM de IPTG.

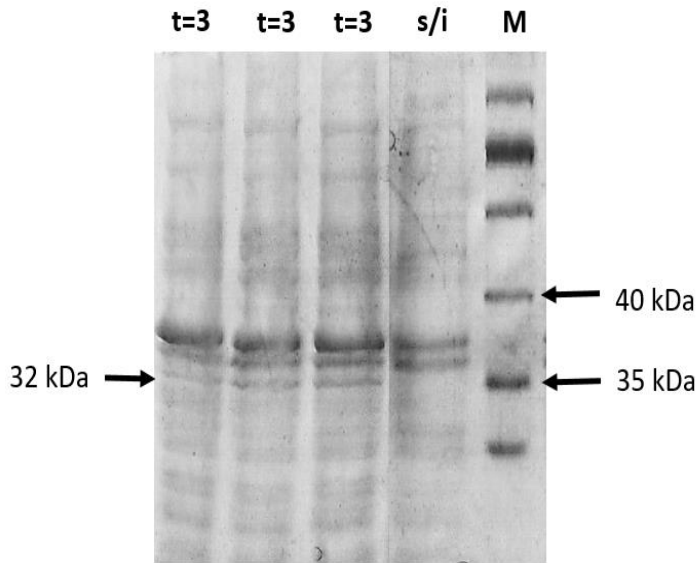


Figura 18. Identificación en SDS PAGE de la proteína recombinante uRNPS1. Fotografía de SDS Page al 12 %, los primeros tres carriles muestran la inducción de la proteína después de tres horas (t= 3), con una flecha se indica la banda de aproximadamente 32 kDa, en el cuarto carril se cargó la muestra sin inducir (s/i) y por último se observa el marcador de peso molecular (M); para realizar esta inducción se utilizó una concentración final de 0.6 mM de IPTG.

Para verificar la presencia de las proteínas recombinantes se realizó un Western Blot utilizando el anticuerpo que reconoce la cola de Histidinas (Santa Cruz). Debido a la baja cantidad de proteína obtenida de la inducción de uSRSF1 y uRBM39 solo se realizó el Western para uRNPS1 (Fig. 19). Se obtuvo la señal que corresponde para una proteína de aproximadamente 32 kDa.

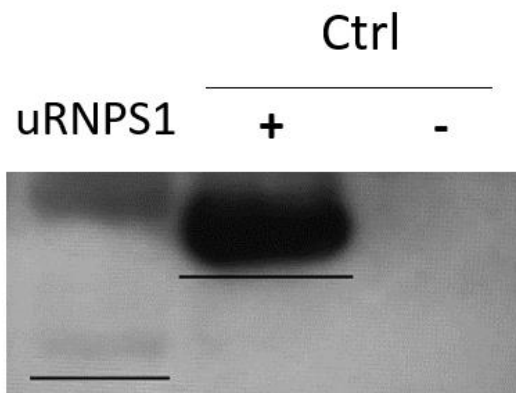


Figura 19. Western blot de la proteína recombinante uRNPS1. En la fotografía el carril 1 se presenta la señal correspondiente a la proteína uRNPS1, como control positivo (+) se utilizó la proteína recombinante uUPF1 (37 kDa) y como control negativo (-) se utilizó el extracto total de *E. coli* DH5α.

16.4 Identificación de secuencias de regulación putativas de las proteínas SR en el genoma de *Ustilago maydis*.

Primero se realizó una búsqueda de los sitios de unión reportados para las proteínas SR de humano e inicialmente sólo se encontraron sitios de unión reportados para la proteína SRSF1. Sin embargo, recientemente se reportaron los sitios de unión para SR45, el ortólogo de RNPS1 en *Arabidopsis thaliana* (Xing y cols., 2015) y debido a que esta era otra de las proteínas recombinantes obtenidas en este trabajo, se decidió incorporar dichos sitios de unión a nuestro estudio. Considerando que las proteínas recombinantes uSRSF1 y uRBM39 obtenidas se expresan en pequeñas cantidades y son proteínas de mayor tamaño, planteamos el empleo de la proteína recombinante uRNPS1 para los ensayos de interacción propuestos en este trabajo, ya que se expresa en mayor cantidad y podría ser de mejor manipulación para dichos experimentos. En la siguiente tabla se muestran los sitios de unión reportados para las proteínas SRSF1 y RNPS1, indicando los métodos mediante los cuales fueron identificadas estas secuencias.

Tabla 11. Sitios de unión reportados para SRSF1 y RNPS1.

Proteína SR	Sitio de unión	Método
SRSF1 de humano	RGAAGAAC	SELEX (Tacke <i>et al.</i> , 1995)
	AGGACRRAGC	SELEX (Tacke <i>et al.</i> , 1995)
	SRSASGA	SELEX funcional (Liu <i>et al.</i> , 1998)
	UGRWG	CLIP (Sanford <i>et al.</i> , 2008)
Ortólogo de RNPS1 en <i>A. thaliana</i>	GAGAGAGAGA	RIP-Seq (Xing <i>et al.</i> , 2015)
	TCTCTCTCTC	
	CTTCTTCTTC	
	GAAGAAGAA	

A partir de la información obtenida en la bibliografía y resumida en la tabla anterior, se realizó la búsqueda de las 4 secuencias blanco identificadas como sitios de unión para el ortólogo de RNPS1 en el genoma de *U. maydis* empleando el programa GenFinder.

Esta búsqueda consistió en identificar cada una de las secuencias en los 6785 genes que posee *Ustilago maydis*. Los resultados obtenidos a partir de dicha búsqueda se presentan en la siguiente grafica (Fig. 20).

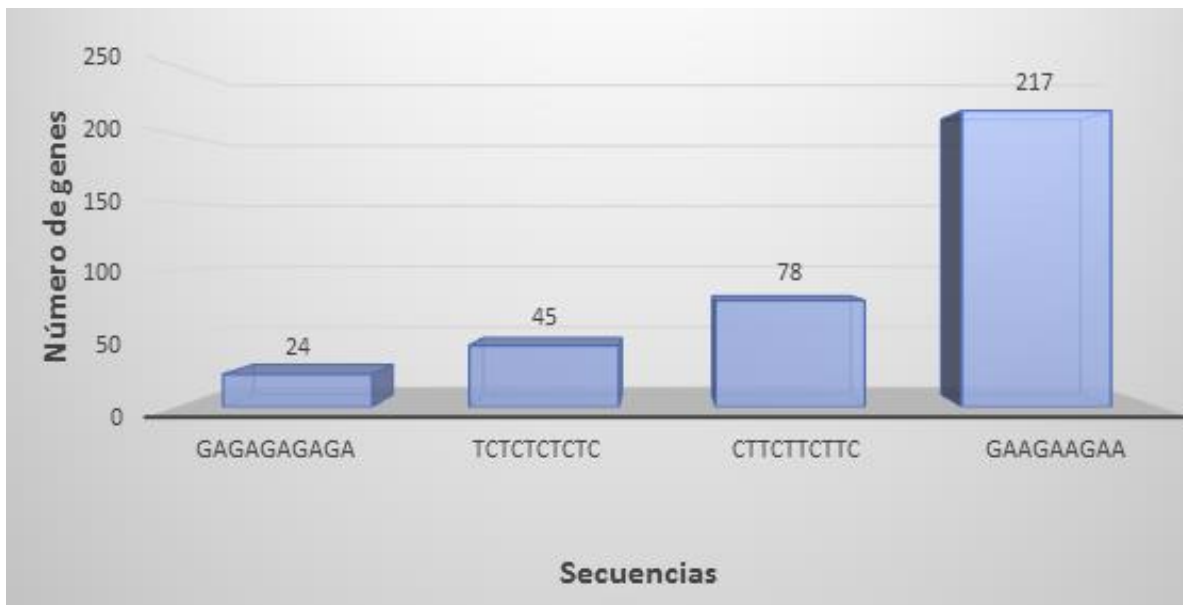


Figura 20. Frecuencia de genes que muestran secuencias de pegado para RNPS1 en el genoma de *Ustilago maydis*. En esta gráfica se muestra el número de genes de *U. maydis* que presentaron secuencias de unión para RNPS1, con ayuda del Programa GenFinder (Hernández-Pérez *et al.*, en proceso).

A partir de los resultados mostrados en la figura 20 se puede observar que 356 genes (5.24 % del genoma) de *U. maydis* presentan secuencias blanco para RNPS1. Como se puede observar la secuencia GAGAGAGAGA es la menos prevalente al estar presente en sólo 24 genes de *U. maydis*. El repetido TCTCTCTCTC se encontró en 45 genes, mientras que la secuencia CTTCTTCTTC se encontró en 78 genes. La secuencia GAAGAAGAA, que corresponde a una secuencia blanco característica para proteínas SR resultó ser la más abundante presentándose en 217 genes de *U. maydis*.

A partir de estos datos se diseñaron los oligos para realizar los ensayos de retardamiento para la proteína uRNPS1.

16.5 Análisis de interacción de uRNPS1 *in silico*.

Por otra parte, se realizó el interactoma de la proteína uRNPS1, empleando la herramienta STRING. Este programa realiza predicciones de interacciones entre proteínas; principalmente se basa en extrapolar datos de interacciones reportados de proteínas que resultan homóloga a la proteína de interés. Se identificaron las 20 interacciones principales predichas para esta proteína, las cuales se muestran en la figura 21.

Posteriormente, nos dimos la tarea de buscar los homólogos en *Homo sapiens* para cada uno de los locus de *U. maydis* que formaban parte del interactoma. En la tabla 12 se muestran los locus correspondientes a las proteínas de *U. maydis* con las cuales interactúa uRNPS1 (locus UM3178.1), así como las proteínas putativas que corresponden a algunos de los locus. De manera adicional, se muestra el porcentaje de identidad del factor identificado en *U. maydis* con el homólogo respectivo en *H. sapiens*, así como el tamaño de cada proteína de *U. maydis*. En color rojo se indican las proteínas que forman parte del complejo de unión exón-exón (EJC), aquellas proteínas que forman el núcleo de NMD se encuentran de color amarillo y en color azul se muestran otras proteínas SR.

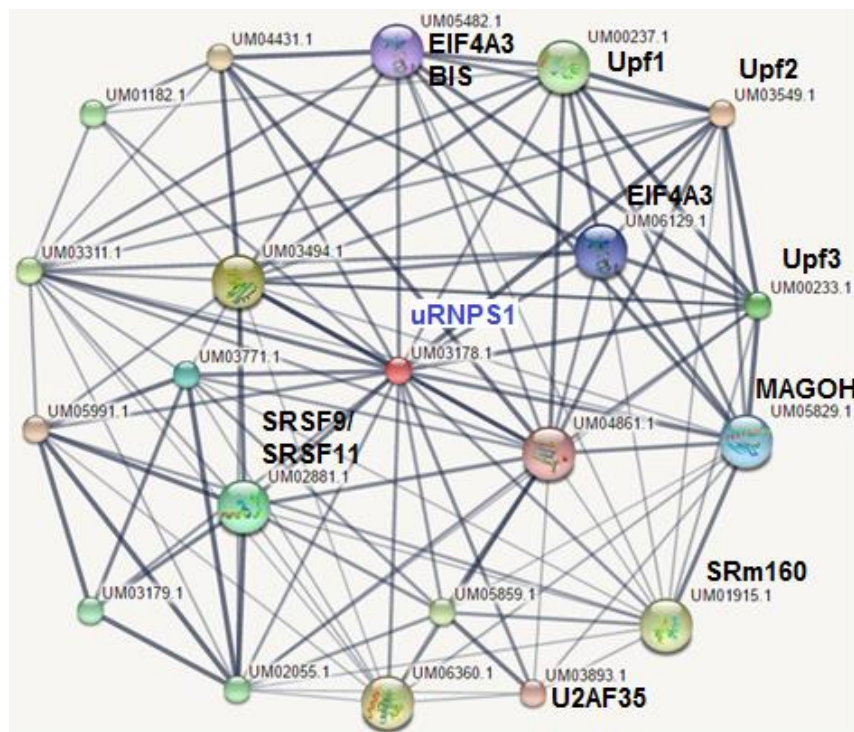


Figura 21. Interacción de RNPS1 con diferentes proteínas. Se muestran las interacciones predichas para RNPS1 (locus UM03178.1) con 20 proteínas, empleando la herramienta STRING. Las esferas representan cada una de las proteínas, al lado se indica el locus al que pertenecen incluyendo el homólogo correspondiente identificado para algunos blancos. El locus para RNPS1 en *Ustilago maydis* se encuentra justo es el centro. Las líneas se emplean para representar la interacción, el grosor o intensidad de la línea indica el grado de confiabilidad de la interacción dada la evidencia disponible.

Tabla 12. Características de las proteínas que interactúan con uRNPS1.

Locus en <i>U. maydis</i>	Tamaño (aa)	Proteína putativa	Identidad con <i>H. sapiens</i> (%)
UM00233.1	559	Upf3	33
UM00237.1	1088	Upf1	79
UM03549.1	1235	Upf2	33
UM01915.1	281	SRm160	51
UM02881.1	403	SRSF9/SRSF11	68
UM03893.1	279	U2AF35	61
UM05482.1	411	EIF4A3 BIS	74
UM05829.1	145	MAGOH	67
UM06129.1	397	EIF4A3	79
UM01182.1	464	--	29
UM02055.1	645	--	34
UM03179.1	1185	--	93
UM03311.1	1081	--	43
UM03494.1	651	--	48
UM03771.1	738	--	32
UM04431.1	1740	--	33
UM04861.1	211	--	61
UM05859.1	313	--	56
UM05991.1	667	--	50
UM06360.1	126	--	74

Las identidades oscilan entre el 29 % y el 79 %, las proteínas que presentaron un porcentaje mayor son UPF1 y EIF4A3 (79 %), EIF4A3BIS (74 %) SRSF9/SRSF11 (68 %) y MAGOH (67 %), todas estas proteínas se encuentran formando parte del complejo de unión exón-exón (EJC), del mecanismo de NMD y del proceso de splicing, incluyendo otras proteínas SR.

17. DISCUSIÓN DE RESULTADOS.

17.1 Identificación *in silico* de los homólogos para las proteínas SR reportadas para humano, en el genoma de *Ustilago maydis*.

En este trabajo identificamos 45 homólogos de proteínas SR de *Homo sapiens* contenidas en el genoma *Ustilago maydis*, estos resultados fueron obtenidos por medio de la búsqueda de proteínas SR en la literatura y la identificación de su homólogo utilizando un alineamiento de sus secuencias aminoacídicas, empleando el Blast de la plataforma del NCBI (Tabla 8) Se conoce que algunos componentes básicos de la maquinaria de *splicing* como los 5 snRNAs se encuentran conservadas en levaduras de fisión y mamíferos (Jurica & Moore, 2003). Sin embargo, respecto a los factores auxiliares como proteínas SR se han identificado en menor cantidad, por ejemplo, el caso *Saccharomyces cerevisiae* donde solo se han identificado tres proteínas SR, dentro de las cuales la proteína NPL3, se encuentra llevando a cabo procesos de *splicing* (Kress *et al.*, 2008). De acuerdo con este abordaje se sugiere que probablemente el mecanismo de *splicing* en el hongo *U. maydis* también es regulado por esta familia de proteínas y que se encuentran en mayor abundancia, debido principalmente a la presencia de intrones dentro de sus genes en comparación con otras levaduras que debido a la ausencia de intrones, presentan también la pérdida en el *splicing alternativo* (Shen & Green, 2006). Aunque todavía es necesario validar funcionalmente la participación de estas proteínas en la regulación de *splicing* tanto *in vitro* como *in vivo* en el hongo, los hallazgos en este trabajo sugieren que las proteínas SR putativas identificadas pueden tener una función en la regulación del *splicing* en *U. maydis*, como lo sugiere el interactoma de RNPS1, donde se observa que esta proteína interacciona con proteínas SR U2AF35, SRSF9 - SRSF11 y Sm160, U2AF35 forma parte del *spliceosoma*, ya que es la subunidad pequeña que forma parte de U2AF junto con la subunidad grande U2AF65, este heterodímero se une directamente con el tracto de polipirimidinas, sin embargo, la subunidad U2AF35 reconoce el sitio de empalme 3' (Wu *et al.*, 1999).

Por otro lado, se ha demostrado al menos en *H. sapiens* que las proteínas SR tienen una participación también en el mecanismo de NMD (nonsense mediated mRNA decay) (Le Hir *et al.*, 2001), el cual consiste en la degradación de ARNm generados por *splicing* que presentan codones de paro prematuros, UTRs largos, etc. (Green *et al.*, 2003; Mendell *et al.*, 2004). En este sentido, cabe destacar que en ocasiones los codones de paro prematuros son introducidos cuando éstos se encuentran localizados en un exón o intrón alternativo y que durante el proceso de *splicing* se deposita alguna proteína SR sobre el ARNm, posteriormente dicha secuencia es arrestada por la maquinaria de NMD y puede ser degradada (Twyffels *et al.*, 2011). Por otra parte, dentro de nuestro grupo de trabajo se ha identificado que las proteínas que forman el núcleo del mecanismo de NMD: UPF1, UPF2 y UPF3 (Andersen *et al.*, 2000) también poseen homólogos en *U. maydis*. En conjunto, las observaciones referentes a los procesos de *splicing* y NMD indican que ambos procesos se encuentran activos en el hongo y que éstos son más parecidos a los procesos descritos en *H. sapiens* que a los que se han demostrado para *S. cerevisiae*. La demostración de que el procesamiento del ARNm ocurre de manera muy semejante entre *H. sapiens* y *U. maydis* deberá ser abordado en otros estudios.

17.2 Identificación de proteínas que contienen dominios RRM y RS característicos de las proteínas SR en la anotación de genoma de *Ustilago maydis*.

Por otra parte, a partir del segundo abordaje también se identificaron un número importante de posibles factores SR, con un total de 53 proteínas putativas en *Ustilago maydis* (Tabla 9). En primera instancia se realizó la búsqueda de todas aquellas proteínas que tuvieran dominios RRM dentro del genoma de *U. maydis* con ayuda de la base datos de Pfam, posteriormente se analizaron las secuencias aminoacídicas para identificar las regiones ricas en residuos de arginina y serina. Se identificaron 18 proteínas que poseían repetidos RS, esto es un hallazgo importante ya que la presencia de proteínas SR canónicas indica la conservación de estos factores en nuestro organismo de estudio, esto coincide con otros reportes donde se menciona la conservación evolutiva y estructural de esta familia de factores (Zahler *et al.*, 1993). También se identificaron 27 proteínas que contienen regiones ricas en Ser, de acuerdo con esto, si recordamos las proteínas SR sufren eventos de fosforilación en el dominio RS principalmente en los residuos de Ser, de tal manera que estas características brindan a la proteína mayor versatilidad, un ejemplo claro de esto es, la capacidad que tienen algunas proteínas para trasladarse del núcleo al citoplasma acompañando al ARNm (Caceres *et al.*, 1998; Cazalla *et al.*, 2002), sin embargo, para que esta deslocalización se lleve a cabo la proteína SR debe estar en un estado hipofosforilado (Huang *et al.*, 2004). Por el contrario, para que estas proteínas regresen al núcleo es necesario que sean fosforiladas precisamente en residuos de Ser por medio de SPK1 (quinasas de proteínas SR) o Clk/Sty (Shepard & Hertel, 2009; Siebel *et al.*, 1999), un claro ejemplo es la fosforilación de la proteína Npl3 de *S. cerevisiae* por la cinasa de proteínas SR, Sky1p, del mismo organismo; donde no solo se evidenció la presencia de estos factores, sino también la habilidad de proteína para transportarse del núcleo hacia el citoplasma (Siebel *et al.*, 1999). De manera semejante, los residuos de R al ser metilados provocan el arresto de proteínas SR en el núcleo como ocurre en el caso de la proteína SRSF1, la cual es metilada en los residuos R93, R97 y R109, (Sinha *et al.*, 2011). De acuerdo con nuestras observaciones, es posible que las proteínas SR putativas identificadas en el hongo podrían tener los residuos necesarios para sufrir las modificaciones indicadas. Nuevamente, es necesaria la demostración experimental posterior que indique tanto que las reacciones son posibles, como que los cambios de localización y de funcionalidad son posibles.

17.3 Comprobación de la presencia de los genes codificantes para las proteínas SR: SRSF1, RMB39 y RNPS1 en *Ustilago maydis* y generación de la proteína recombinante uRNPS1.

A partir de las 15 proteínas SR putativas obtenidas de la correlación realizada con los resultados de los dos abordajes anteriormente mostrados se decidió trabajar solo con tres locus UMAG_00164, UMAG_02552 y UMAG_3178, que corresponden a los homólogos para las proteínas SRSF1, RBM39 y RNPS1; SRSF1 es una de las proteínas SR más estudiadas se encuentra involucrada

como activador de *splicing* (Wu & Maniatis, 1993), en procesos de transporte nuclear (Serrano *et al.*, 2016) y actuando como regulador de procesos apoptóticos y proliferativos en distintos tejidos (Anczuków *et al.*, 2012); con respecto a RBM39 se ha identificado como supresor o activador de *splicing* alternativo (Mai *et al.*, 2016) y también se ha identificado su baja expresión en carcinoma de pulmón de células pequeñas (Bangur *et al.*, 2002); por último RNPS1 como se ha mencionado con anterioridad forma parte del complejo de unión exón-exón y por lo tanto se encuentra involucrada con mecanismos de NMD, por otro lado, también se ha visto que brinda estabilidad a los transcritos nacientes (Li *et al.*, 2007). Con todo lo mencionado en este apartado podemos decir que, resulta importante el hallazgo y generación de estas proteínas en *U. maydis*. Como se observa en la figura 9, se obtuvieron los amplificadores de cada uno de los locus correspondientes a las proteínas SR, esto resulta importante ya que efectivamente *U. maydis* posee los genes que codifican para estas tres proteínas SR putativas. Posteriormente, se llevó a cabo la clonación y transformación en células de *E. coli* DH5 α , con respecto a esto se presentaron ciertas complicaciones, ya que las inducciones de las proteínas recombinantes uSRSF1 y RBM39 no fueron tan evidentes, probablemente este resultado se debe al hecho de haber utilizado un sistema de expresión heterólogo; cabe resaltar que en ninguno de nuestros experimentos se utilizaron inhibidores de proteasas. Sin embargo, con respecto a la proteína uRNPS1 putativa se aprecia una mayor inducción, su presencia se corroboró realizando un western blot utilizando el anticuerpo anti-His (Fig. 19). Obtener la proteína recombinante esperada amplía la posibilidad de tener una proteína funcional.

Con la finalidad de seguir con este estudio las proteínas uSRSF1 y uRBM39 ya no se utilizaron, sin embargo, se plantea seguir trabajando con ellas en futuros estudios dado el gran potencial que representan.

17.4 Búsqueda de las secuencias de regulación para proteínas SR en el genoma de *Ustilago maydis*.

Para poder identificar los sitios putativos de regulación para alguna de nuestras proteínas recombinantes SR de *Ustilago maydis*, en primer lugar, se buscaron sitios de unión a ARN de las proteínas SR reportadas, se encontraron sitios de unión para la proteína SRSF1, SRSF2 (Tacke & Manley, 1997), SRSF5 (Tacke *et al.*, 1997), SRSF3, SRSF7 (Cavaloc *et al.*, 1999) y SR45 (RNPS1) (Xing *et al.*, 2015); posteriormente a partir de estos datos se decidió utilizar las secuencias de unión para la proteína SR45, que corresponde al ortólogo de RNPS1 en *Arabidopsis thaliana*, la razón por la cual se decidió utilizar esta proteína, se debió a que su recombinante fue la que mejor se expresó en nuestros experimentos. A continuación, empleando el programa denominado GeneFinder (Hernández – Pérez *et al.*, en proceso), se identificaron los sitios putativos de unión para las proteínas SR, para este caso en particular se buscaron las secuencias GAAGAAGA, CTTCTTCTTC, TCTCTCTCTC y GAGAGAGAGA (Xing *et al.*, 2015). Es importante recalcar que este programa se encuentra alimentado con el genoma completo de *U. maydis* y los resultados que nos arroja corresponden a la abundancia de la secuencia en el genoma, el número de genes que poseen la

secuencia, así como también el gen en el que se encuentra la secuencia, el número de veces que se encuentra tal secuencia en el gen y la posición en la cual se encuentra la secuencia dentro del gen.

A partir de los resultados mostrados en la figura 16 se puede observar que 356 genes (5.24 % del genoma) de *U. maydis* presentan secuencias blanco para RNPS1, no existen reportes en los cuales se indique la abundancia de cada una de las secuencias de unión para RNPS1, sin embargo, estos sitios blanco (Tabla 11) se identificaron en un total de 4361 transcritos derivados 4262 de los genes de *A. thaliana*, estos transcritos son denominados como SARs (SR45-associated RNAs) (Xing *et al.*, 2015). Por otro lado, se encontró que la secuencia con mayor prevalencia en el genoma de *U. maydis* fue GAAGAAGAA y fue identificada en 217 genes, se ha reportado recientemente que en plantas el motivo GAAG es un elemento en *cis* muy importante involucrado en la regulación del *splicing alternativo* (Perteza *et al.*, 2007; Thomas *et al.*, 2012). La secuencia de GAGAGAGAGA a pesar de poseer los mismos nucleótidos que la secuencia GAAGAAGAA, se encontró en menor abundancia con respecto a todas presentándose solo en 24 genes, estos resultados podrían indicarnos la especificidad que presentan de manera general las proteínas SR por este tipo de secuencias, debido a que tanto la proteína TRA2 como SRSF10 se unen a motivos GAAGAA y AGAGAAA respectivamente, resulta tener una gran similitud entre ellas (Tacke *et al.*, 1998; Tacke *et al.*, 1997), En cuanto a las secuencias TCTCTCTC y CTCTCTTC las cuales fueron encontradas en 45 y 78 genes respectivamente, podemos decir, que probablemente sean blancos específicos para la proteína uRNPS1.

17.5 Interacción de uRNPS1 con otras proteínas.

La proteína recombinante uRNPS1 se obtuvo a partir del amplificado de 594 pb que corresponde al locus UMAG_03178, este amplificado se clonó en primer lugar en el vector pJET1.2, el cual funciona como reservorio para nuestro producto; posteriormente, se llevó a cabo una subclonación en el vector pPROEXHTc, para realizar este paso se realizó una doble digestión de la construcción pJET-uRNPS1 y el vector pPROEXHTc con las enzimas de restricción NotI- XbaI, a continuación se llevó a cabo la ligación de los productos obtenidos que dio lugar a la generación de la construcción pPROEX-uRNPS1, ambas clonaciones se transformaron en células DH5 α . A partir de esta construcción se expresó la proteína uRNPS1 de 32 kDa; para la inducción se utilizó una concentración final de 0.6 mM de IPTG, durante 3 horas, a 37°C en agitación constante. La razón principal por la cual se utilizó la recombinante uRNPS1 para realizar todos los análisis anteriores parte del hecho de que, no se logró la expresión esperada de las proteínas uRBM39 y uSRSF1, las razones de este resultado son variadas como se mencionó anteriormente, se utilizó un sistema de expresión heterólogo, en el caso de la proteína uSRSF1 su tamaño era de aproximadamente 170 kDa, esto implica un mayor desgaste de la célula bacteriana para sintetizar la proteína, no se utilizó ningún tipo de inhibidor de proteasas, variables en el uso preferente de codones de paro del organismo, etc. Para futuros ensayos se propone utilizar otro tipo de metodologías que confieran un resultado exitoso, sin embargo, para fines prácticos y para poder seguir con el trabajo estipulado se decidió dejar a un lado estos problemas técnicos y trabajar con la proteína uRNPS1.

Se conoce que la organización de los dominios y los procesos en los cuales se encuentra implicada la proteína RNPS1 se conservan en protozoarios, plantas, hongos, bacterias y metazoos (Califice *et al.*, 2012), este reporte nos brinda cierto grado de seguridad en cuando a lo realizado a lo largo de todo nuestro estudio. En el análisis del interactoma de uRNPS1 realizado con la herramienta STRING se observó que varias de las proteínas con las que interactúa se encuentran asociadas a procesos de *splicing*, forman parte del complejo EJC y algunas otras corresponden a proteínas SR. De manera general, se ha demostrado que la proteína RNPS1 en *Homo sapiens* estimula y regula el *splicing* de los sitios 3' alternativos (Mayeda *et al.*, 1999), también se ha observado en otros estudios que en *A. thaliana* y *Drosophila melanogaster* su homólogo para RNPS1 se encuentra jugando un papel importante en el *splicing* alternativo (Aly *et al.*, 2003; Ashton-Beaucage *et al.*, 2010), de esta manera se explica el hecho de que interactúe con otras proteínas SR (U2AF35 y Sm160); también se ha establecido que participa en la exportación del ARNm del núcleo al citoplasma formando parte transitoriamente del complejo de unión exón-exón (EJC) el cual es depositado después del empalme de los exones y está formado principalmente por las proteínas Y14, Magoh, eIF4AIII y MLN51 (Tange *et al.*, 2005), en el interactoma realizado *in silico* en nuestro estudio, pudimos observar que la uRNPS1 putativa identificada tiene la capacidad de interactuar con los homólogos equivalentes en *Ustilago maydis* para estas mismas proteínas a excepción de MNL5; este mismo complejo se encuentra íntimamente asociado con mecanismos de NMD, por ejemplo, en el caso de *D. melanogaster* se han identificado los homólogos para las proteínas UPF1–3, SMG1, SMG5 and SMG6 (Gatfield *et al.*, 2003), de manera similar en estudios realizados por nuestro grupo de trabajo se han identificado los homólogos para las proteínas UPF1 (Martínez- Montiel *et al.*, 2016), UPF2 y UPF3 (Martínez- Montiel., en proceso) , las cuales forman el núcleo de mecanismo de NMD, reiterando la presencia de estas proteínas de acuerdo con las interacciones predichas en nuestro ensayo *in silico* donde también se identificó que uRNPS1 interactúa con las proteínas UPF1, UPF2 y UPF3 (Tabla 12). Nuevamente, estas observaciones sugieren que el homólogo de RNPS1 en *U. maydis* puede tener una función semejante a la que se ha descrito para esta proteína en humano, indicado por las reacciones de interacción propuestas. Sin embargo, se requerirá de evidencia experimental adicional para validar dichas interacciones.

18. CONCLUSIONES.

- ✓ El 94 % de proteínas SR reportadas en *Homo sapiens* se encuentran presentes en *U. maydis*.
- ✓ De las 53 proteínas SR putativas con dominios RRM de *U. maydis*, 18 contienen dominios SR canónicos, 27 poseen repetidos de Serina y 8 incluyen repetidos de Arginina.
- ✓ *Ustilago maydis* posee 78 locus que corresponden a proteínas SR putativas.
- ✓ En comparación *U. maydis* posee más proteínas SR que *Saccharomyces cerevisiae*.
- ✓ La secuencia de unión para uRNPS1 es contenida en 356 genes de *U. maydis*, lo que corresponde al 5.24 % de su genoma.
- ✓ La secuencia de pegado GAAGAAGAA es la más prevalente para uRNPS1, presentándose en 217 genes de *U. maydis*.

19. PERSPECTIVAS.

- ❖ Completar los estudios de interacción ARN-proteína para uRNPS1.
- ❖ Realizar los estudios de interacción proteína – ARN, utilizando las proteínas uRMB39 y uSRSF1 recombinantes obtenidas.
- ❖ Analizar la estructura de los genes que contienen las secuencias de pegado de *Ustilago maydis*.
- ❖ Evaluar la funcionalidad de las proteínas SR recombinantes obtenidas mediante ensayos de *splicing in vitro*.
- ❖ Validar la funcionalidad en *splicing* de las proteínas recombinantes obtenidas en un sistema heterólogo (células cancerosas en cultivo).

20. BIBLIOGRAFÍA.

1. Ali GS, Golovkin M, Reddy AS. (2003) Nuclear localization and in vivo dynamics of a plant-specific serine/arginine-rich protein. *Plant J* 36: 883–893.
2. Ast G. (2004) How did alternative splicing evolve? *Nature Reviews Genetics* 5(10): 773-782.
3. Anczuków O, Rosenberg AZ, Akerman M, Das S, Zhan L, Kami R, Muthuswamy SK, Krainer AR. (2012). The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nature structural & molecular biology* 19(2): 220-228.
4. Ashton-Beaucage D, Udell CM, Lavoie H, Baril C, Lefrançois M, Chagnon P, Gendron P, Caron-Lizotte O, Bonneil E, Thibault P, Therrien M. (2010) The exon junction complex controls the splicing of MAPK and other long intron-containing transcripts in *Drosophila*. *Cell* 143: 251–262.
5. Bangur CS, Switzer A, Fan L, Marton MJ, Meyer MR, Wang T. (2002) Identification of genes over-expressed in small cell lung carcinoma using suppression subtractive hybridization and cDNA microarray expression analysis. *Oncogene* 21(23): 3814-3825.
6. Banuett F. (1995) Genetics of *Ustilago maydis*, a fungal pathogen that induces tumors in maize. *Annual review of genetics*, 29(1): 179-208.
7. Barrass JD, Beggs JD. (2003) Splicing goes global. *Trends in Genetics* 19(6): 295-298.
8. Berget SM, Moore C, Sharp PA. (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences* 74(8): 3171-3175.
9. Bucheli ME, Buratowski S. (2005). Npl3 is an antagonist of mRNA 3' end formation by RNA polymerase II. *The EMBO journal* 24(12): 2150-2160.
10. Boucher L, Ouzounis CA, Enright AJ, & Blencowe BJ. (2001) A genome-wide survey of RS domain proteins. *Rna* 7(12): 1693-1701
11. Busch A., Hertel KJ. (2012) Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdisciplinary Reviews. RNA* 3(1): 1-12.
12. Cáceres JF, Sreaton GR, Krainer, AR. (1998) A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes & development* 12(1): 55-66.
13. Caputi M, Mayeda A, Krainer AR, Zahler AM. (1999) hnRNP A/B proteins are required for inhibition of HIV-1 pre-mRNA splicing. *The EMBO journal* 18(14): 4060-4067.
14. Cavaloc Y, Bourgeois CF, Kister L, Stévenin J. (1999) The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *Rna* 5(3): 468-483.
15. Califice S, Baurain D, Hanikenne M, Motte P. (2011) A single ancient origin for prototypical SR splicing factors. *Plant physiology* 158: 546-560.
16. Cazalla D, Zhu J, Manche L, Huber E, Krainer AR, Cáceres, JF. (2002) Nuclear export and retention signals in the RS domain of SR proteins. *Molecular and cellular biology* 22(19): 6871-6882.

17. Chang KY, Georgianna DR, Heber S, Payne GA, Muddiman DC. (2010). Detection of alternative splice variants at the proteome level in *Aspergillus flavus*. *Journal of proteome research* 9(3): 1209-1217.
18. Cho S, Hoang A, Sinha R, Zhong, XY, Fu XD, Krainer AR, Ghosh G. (2011) Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proceedings of the National Academy of Sciences* 108(20): 8233-8238.
19. Colwill K, Feng LL, Yeakley JM, Gish GD, Cáceres JF, Pawson T, Fu XD. (1996) SRPK1 and Clk/Sty protein kinases show distinct substrate specificities for serine/arginine-rich splicing factors. *Journal of Biological Chemistry* 271(40): 24569-24575.
20. Castillo-Lluva S, Alvarez-Tabarés I, Weber I, Steinberg G, Pérez-Martín J. (2007) Sustained cell polarity and virulence in the phytopathogenic fungus *Ustilago maydis* depends on an essential cyclin-dependent kinase from the Cdk5/Pho85 family. *Journal of cell science* 120(9): 1584-1595.
21. Cohen SN, Chang AC, Hsu L. (1972) Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. *Proceedings of the National Academy of Sciences* 69(8): 2110-2114.
22. David, CJ, Manley JL. (2010) Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes & development* 24(21): 2343-2364.
23. Dermody JL, Dreyfuss JM, Villén J, Ogundipe B, Gygi SP, Park PJ, Ponticelli CL, Buratowski S, Bucheli ME. (2008) Unphosphorylated SR-like protein Npl3 stimulates RNA polymerase II elongation. *PLoS one* 3(9): e3273.
24. Deka P, Bucheli ME, Moore C, Buratowski S, Varani G. (2008) Structure of the yeast SR protein Npl3 and Interaction with mRNA 3'-end processing signals. *Journal of molecular biology* 375(1): 136-150.
25. Faustino NA, Cooper TA. (2003) Pre-mRNA splicing and human disease. *Genes & development* 17(4): 419-437.
26. Gatfield D, Unterholzner L, Ciccarelli FD, Bork P, Izaurralde E. (2003) Nonsense-mediated mRNA decay in *Drosophila*: at the intersection of the yeast and mammalian pathways. *The EMBO journal* 22(15): 3960-3970.
27. Green MR, Sambrook J. (2012) *Molecular cloning a laboratory Manual*. Cold Spring Harbor Laboratory Press.
28. Green RE, Lewis BP, Hillman RT, Blanchette M, Lareau LF, Garnett AT, Brenner SE. (2003) Widespread predicted nonsense-mediated mRNA decay of alternatively-spliced transcripts of human normal and disease genes. *Bioinformatics* 19: 118-121.
29. Häcker S, Krebber H. (2004) Differential export requirements for shuttling serine/arginine-type mRNA-binding proteins. *Journal of Biological Chemistry* 279(7): 5049-5052.
30. Harrow J, Denoeud F, Frankish A, Reymond A, Chen CK, Chrast J, Rossier C. (2006). GENCODE: producing a reference annotation for ENCODE. *Genome biology* 22:1760-1774.
31. Henscheid KL, Voelker RB, Berglund JA. (2008) Alternative modes of binding by U2AF65 at the polypyrimidine tract. *Biochemistry* 47(1): 449-459.

32. Herrera, JR. (2008) *Ustilago maydis*: ascenso de un hongo mexicano de la gastronomía local al mundo científico. *Nova Scientia* 1(1): 118-135.
33. Herold A, Suyama M, Rodrigues JP, Braun IC, Kutay U, Carmo-Fonseca M, Izaurralde E. (2000) TAP (NXF1) belongs to a multigene family of putative RNA export factors with a conserved modular architecture. *Molecular and cellular biology* 20(23): 8996-9008.
34. Hinas A, Larsson P, Avesson L, Kirsebom LA, Virtanen A, Söderbom F. (2006) Identification of the major spliceosomal RNAs in *Dictyostelium discoideum* reveals developmentally regulated U2 variants and polyadenylated snRNAs. *Eukaryotic cell* 5(6): 924-934.
35. Huang Y, Yario TA, Steitz, JA. (2004) A molecular link between SR protein dephosphorylation and mRNA export. *Proceedings of the National Academy of Sciences of the United States of America* 101(26): 9666-9670.
36. International Human Genome Sequencing Consortium. (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431(7011): 931-945
37. Jacquenet S, Decimo D, Muriaux D, Darlix JL. (2005) Dual effect of the SR proteins ASF/SF2, SC35 and 9G8 on HIV-1 RNA splicing and virion production. *Retrovirology* 2: 33.
38. Jia R, Li C, McCoy JP, Deng CX, Zheng ZM. (2010) SRp20 is a proto-oncogene critical for cell proliferation and tumor induction and maintenance. *Int J Biol Sci* 6(7): 806-826.
39. Jiménez-García LF, Spector DL. (1993) In vivo evidence that transcription and splicing are coordinated by a recruiting mechanism. *Cell* 73(1):47-59.
40. Jurica MS, Moore MJ. (2003). Pre-mRNA splicing: awash in a sea of proteins. *Molecular cell* 12(1): 5-14.
41. Kanopka A, Mühlemann O, Akusjärvi G. (1996) Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature* 381(6582): 535-538.
42. Kempken F. (2013) Alternative splicing in ascomycetes. *Applied microbiology and biotechnology* 97(10): 4235-4241.
43. Kress TL, Krogan NJ, Guthrie C. (2008) A single SR-like protein, Npl3, promotes pre-mRNA splicing in budding yeast. *Molecular cell* 32(5): 727-734.
44. Lee MS, Henry M, Silver PA. (1996) A protein that shuttles between the nucleus and the cytoplasm is an important mediator of RNA export. *Genes & development* 10(10):1233-1246.
45. Le Hir H, Gatfield D, Izaurralde E, Moore MJ. (2001) The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *The EMBO journal* 20(17): 4987-4997.
46. Li X, Niu T, Manley JL. (2007) The RNA binding protein RNPS1 alleviates ASF/SF2 depletion-induced genomic instability. *Rna* 13(12): 2108-2115.
47. Liu HX, Zhang M, Krainer, AR. (1998) Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes & development* 12(13): 1998-2012.
48. Long JC, Caceres, JF. (2009) The SR protein family of splicing factors: master regulators of gene expression. *Biochemical Journal* 417(1):15-27
49. Lukasiewicz R, Velazquez-Dones A, Huynh N, Hagopian J, Fu XD, Adams J, Ghosh G. (2007) Structurally unique yeast and mammalian serine-arginine protein kinases catalyze

- evolutionarily conserved phosphorylation reactions. *Journal of Biological Chemistry* 282(32): 23036-23043.
50. Luukkonen BM, Séraphin B. (1997) The role of branchpoint-3' splice site spacing and interaction between intron terminal nucleotides in 3' splice site selection in *Saccharomyces cerevisiae*. *The EMBO journal* 16(4): 779-792.
 51. Lykke-Andersen J, Shu MD, Steitz, JA. (2000) Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. *Cell* 103(7): 1121-1131.
 52. Mai S, Qu X, Li P, Ma Q, Cao C, Liu X. (2016). Global regulation of alternative RNA splicing by the SR-rich protein RBM39. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* 1859(8): 1014-1024.
 53. Manley JL, Krainer AR. (2010) A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes & development* 24(11): 1073-1074.
 54. Marshall AN, Montealegre MC, Jiménez-López C, Lorenz MC, Van Hoof A. (2013) Alternative splicing and subfunctionalization generates functional diversity in fungal proteomes. *PLoS Genet* 9(3): e1003376
 55. Martínez-Montiel N, Morales-Lara L, Hernández-Pérez JM, Martínez-Contreras RD. (2016) In Silico Analysis of the Structural and Biochemical Features of the NMD Factor UPF1 in *Ustilago maydis*. *PloS one* 11(2): e0148191.
 56. Matlin AJ, Clark F, Smith CW. (2005) Understanding alternative splicing: towards a cellular code. *Nature reviews Molecular cell biology* 6(5): 386-398.
 57. Mendell JT, Sharifi NA, Meyers JL, Martinez-Murillo F, Dietz HC. (2004) Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nature genetics* 36(10): 1073-1078.
 58. Modrek B, Lee C. (2002). A genomic view of alternative splicing. *Nature genetics* 30(1): 13-19.
 59. Münsterkötter M, Steinberg G. (2007) The fungus *Ustilago maydis* and humans share disease-related proteins that are not found in *Saccharomyces cerevisiae*. *BMC genomics* 8(1): 473.
 60. Nagy E, Maquat LE. (1998) A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends in biochemical sciences* 23(6): 198-199.
 61. Patel AA, Steitz, JA. (2003) Splicing double: insights from the second spliceosome. *Nature Reviews Molecular Cell Biology* 4(12): 960-970.
 62. Perteza M, Mount SM, Salzberg SL. (2007) A computational survey of candidate exonic splicing enhancer motifs in the model plant *Arabidopsis thaliana*. *BMC bioinformatics* 8(1):159.
 63. Rossi F, Labourier E, Forné T, Divita G, Derancourt J, Riou JF, Tazi J. (1996) Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I. *Nature* 381: 80-82.

64. Sanford JR, Coutinho P, Hackett JA, Wang X, Ranahan W, Caceres JF. (2008) Identification of nuclear and cytoplasmic mRNA targets for the shuttling protein SF2/ASF. *PLoS One* 3(10): e3369.
65. Schwindinger WF, Warner JR. (1987) Transcriptional elements of the yeast ribosomal protein gene CYH2. *Journal of Biological Chemistry* 262(12): 5690-5695.
66. Serrano P, Aubol BE, Keshwani MM, Forli S, Ma CT, Dutta SK, Adams JA. (2016) Directional Phosphorylation and Nuclear Transport of the Splicing Factor SRSF1 Is Regulated by an RNA Recognition Motif. *Journal of molecular biology* 428(11): 2430-2445.
67. Shen H, Green MR. (2006) RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. *Genes & development* 20(13): 1755-1765.
68. Shepard PJ, Hertel KJ. (2009) The SR protein family. *Genome biology* 10: 242-242.9.
69. Siebel CW, Feng L, Guthrie C, Fu XD. (1999) Conservation in budding yeast of a kinase specific for SR splicing factors. *Proceedings of the National Academy of Sciences* 96(10): 5440-5445.
70. Steinberg G. (2007) Tracks for traffic: microtubules in the plant pathogen *Ustilago maydis*. *New Phytologist* 174(4): 721-733.
71. Steinberg G, Perez-Martin J. (2008) *Ustilago maydis*, a new fungal model system for cell biology. *Trends in cell biology* 18(2): 61-67.
72. Tacke R, Manley JL. (1995) The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *The EMBO journal* 14(14): 3540-3551.
73. Tacke R, Chen Y, Manley JL. (1997) Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. *Proceedings of the National Academy of Sciences* 94(4): 1148-1153.
74. Tacke R, Tohyama, M, Ogawa S, Manley JL. (1998) Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell* 93(1): 139-148.
75. Tange TO, Shibuya T, Jurica MS, Moore MJ. (2005) Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *Rna* 11(12): 1869-1883.
76. Taylor RG, Walker DC, McInnes RR. (1993) E. coli host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic acids research* 21(7): 1677-1678.
77. Tazi J, Bakkour N, Stamm S. (2009). Alternative splicing and disease *Biochimica Biophysica Acta (BBA)-Molecular Basis of Disease* 1792(1): 14-26.
78. Twyffels L, Gueydan C, Kruijs V. (2011) Shuttling SR proteins: more than splicing factors. *FEBS Journal* 278(18): 3246-3255.
79. Thomas J, Palusa SG, Prasad KV, Ali GS, Surabhi GK, Ben-Hur A, Reddy AS. (2012) Identification of an intronic splicing regulatory element involved in auto-regulation of alternative splicing of SCL33 pre-mRNA. *The Plant Journal* 72(6): 935-946.
80. Van der Feltz C, Anthony K, Brilot A, Pomeranz Krummel DA. (2012) Architecture of the spliceosome. *Biochemistry* 51(16): 3321-3333.

81. Webb CJ, Romfo CM, van Heeckeren WJ, Wise JA. (2005) Exonic splicing enhancers in fission yeast: functional conservation demonstrates an early evolutionary origin. *Genes & development* 19(2): 242-254.
82. Will CL, Luhmann R. (2011) Spliceosome structure and function. *Cold Spring Harb perspect biol* 3(7): a003707.
83. Wu JY & Maniatis T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* 75(6): 1061-1070.
84. Wu S, Romfo CM, Nilsen TW, Green MR. (1999) Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* 402(6763): 832-835.
85. Xing D, Wang Y, Hamilton M, Ben-Hur A, Reddy AS. (2015) Transcriptome-Wide Identification of RNA Targets of Arabidopsis SERINE/ARGININE-RICH45 Uncovers the Unexpected Roles of This RNA Binding Protein in RNA Processing. *The Plant Cell* 27(12): 3294-3308.
86. Yu B, Fey P, Kestin-Pilcher KE, Fedorov A, Prakash A, Chisholm RL, Wu JY. (2011) Spliceosomal genes in the *D. discoideum* genome: a comparison with those in *H. sapiens*, *D. melanogaster*, *A. thaliana* and *S. cerevisiae*. *Protein & cell*, 2(5): 395-409.
87. Zahler AM, Neugebauer KM, Lane WS, Roth MB. (1993) Distinct functions of SR proteins in alternative pre-mRNA splicing. *Science* 260(5105): 219-222.
88. Zhao C, Waalwijk C, de Wit PJ, Tang D, & van der Lee T. (2013) RNA-Seq analysis reveals new gene models and alternative splicing in the fungal pathogen *Fusarium graminearum*. *BMC genomics*, 14(1): 1.