



**BENEMÉRITA UNIVERSIDAD AUTÓNOMA
DE PUEBLA**

INSTITUTO DE CIENCIAS CENTRO DE INVESTIGACIÓN EN CIENCIAS
MICROBIOLÓGICAS
POSGRADO EN MICROBIOLOGÍA

**"Identificación de las secuencias de DNA que estructuran los extremos
cromosomales de *Ustilago maydis* cepa PGA 2.1 y
análisis del arreglo que despliegan"**

**TESIS PARA OBTENER EL GRADO DE: MAESTRA EN CIENCIAS
(MICROBIOLOGÍA) CON OPCIÓN EN BIOQUÍMICA Y GENÉTICA
MICROBIANA**

PRESENTA:

Q.C. CINTYA ATONAL AGUILA

ASESORA:

D.C. MA. PATRICIA SÁNCHEZ ALONSO

COASESORA:

D.C LETICIA VEGA ALVARADO

PUEBLA, PUE.

31 DE ENERO 2024



BUAP

Puebla, Pue. a 14 de diciembre 2023.

**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 **Cintya Atonal Águila**, les informamos que hemos revisado el escrito titulado:


“Identificación de las secuencias de DNA que estructuran los extremos cromosomales de *Ustilago maydis* cepa PGA 2.1 y análisis del arreglo que despliegan”

A nuestro juicio, la alumna **Cintya Atonal Águila** puede proceder a la digitalización 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”


M.C. Alejandra Espinosa Texis


D.C. Estela Anastacio Marcelino


D.C. Candelario Vázquez Cruz


D.C. Patricia Lozano Zarain

Índice

Índice de figuras y tablas	4 pág
Agradecimientos.....	5 pág
Resumen.....	6 pág
I. Introducción	7 pág
I.1 Telómeros	7 pág
I.1.1 El DNA telomérico.....	8 pág
I.1.2 Proteínas específicas del telómero y RNA telomérico.....	8 pág
I.1.3 Región subtelomérica.....	9 pág
I.1.4 Genes de contingencia.....	10 pág
I.2 Mantenimiento del extremo cromosomal.....	10 pág
I.2 .1 Vías alternativas.....	11 pág
II. Antecedentes	12 pág
III. Justificación	14 pág
IV. Hipótesis	14 pág
V. Objetivo general	14 pág
V.1 Objetivos específicos.	14 pág
VI. Esquema de trabajo	15 pág
VII. Material y métodos	16 pág
VII.1 Material.....	16 pág
VII.2 Métodos	16 pág
VII.2.1 Secuenciación y ensamblaje	16 pág
VII.2.2 Construcción de catálogo de genes cercanos a región telomérica.....	16 pág
VII.2.3 Identificación de secuencias teloméricas en el genoma.....	17 pág
VII.2.4 Identificación de secuencias asociadas al telómero <i>UTASa</i> Y <i>UTASb</i> en el genoma.....	17 pág
VIII. Resultados	17 pág
VIII. 1 Análisis de calidad de las secuencias	17 pág
VIII. 2 Ensamblaje de secuencias (Canu)	19 pág
VIII. 2 .1 Calidad de ensamblaje de secuencias <i>U. maydis</i>	20 pág
VIII. 2 .2 Resultados evaluación de ensamblajes con la herramienta Mauve.....	20 pág
VIII. 2 .3 Resultados evaluación de ensamblajes con la herramienta Quast.....	23 pág
VIII. 3 Genes cercanos a los extremos teloméricos y secuencias teloméricas en los cromosomas de <i>U. maydis</i>	24 pág
VIII.3.1 Identificación de secuencias asociadas al telómero <i>UTASa</i> y <i>UTASb</i> en el genoma.....	34 pág
IX. Discusión	37 pág
X. Conclusiones	39 pág
XI Perspectivas	40 pág
XII. Referencias	41 pág

Índice de figuras y tablas

Figura. 1.- La región que se encuentra en el borde físico de los cromosomas.....	7 pág
Figura.2.-El ciclo de reacción de la telomerasa.....	11 pág
Figura 3.-Imagen Mazorca de maíz con tumores de <i>U. maydis</i>	12 pág
Figura 4.- Imagen Tumores de <i>U. maydis</i>	12 pág
Figura 5.-Imagen Cepa (PGA 2.1) de <i>U. maydis</i> en la electroforesis.....	13 pág
Figura 6.-Esquema de trabajo.....	15 pág
Figura 7. - Gráfica Longitud de distribución de secuencias de la cepa <i>U. maydis</i> (521).....	18 pág
Figura 8. - Gráfica Longitud de distribución de secuencias de la cepa <i>U. maydis</i> (PGA2.1).....	18 pág
Figura 9. - Gráfica Longitud de distribución de secuencias de la cepa <i>U. maydis</i> (PGAtrt1-4).....	19 pág
Figura 10. - Imagen Genoma de referencia <i>U. maydis</i> 521 GenBank Ensamble CANU <i>U. maydis</i> 521...	20 pág
Figura 11. - Imagen Genoma de referencia <i>U. maydis</i> 521 GenBank comparado con Ensamble CANU <i>U. maydis</i> PGA2.1.....	21 pág
Figura 12. - Imagen Genoma de Ensamble CANU <i>U. maydis</i> PGA21 comparado con Ensamble CANU <i>U. maydis</i> PGAtrt1.....	22 pág
Figura 13.- Representación de genes que se encuentran en los 50 kb cercanas al extremo telomérico izquierdo.....	26 pág
Figura 14.- Representación de genes que se encuentran en los 50 kb cercanas al extremo telomérico izquierdo.....	27 pág
Figura 15.- Representación de genes que se encuentran en los 50 kb cercanas al extremo telomérico derecho.....	28 pág
Figura 16.- Representación de genes que se encuentran en los 50 kb cercanas al extremo telomérico derecho.....	29 pág
Figura 17. - Gráfica Longitud de telómeros de cepa <i>U. maydis</i> 521	33 pág
Figura 18. - Gráfica Longitud de telómeros de cepa <i>U. maydis</i> PGA2.1.....	33 pág
Figura 19. - Gráfica Longitud de telómeros de cepa <i>U. maydis</i> PGAtrt1-4.	34 pág
Figura 20.- Organización de secuencias en el extremo cromosomal de <i>U. maydis</i> 521.....	35 pág
Figura 21.- Organización de secuencias en el extremo cromosomal de <i>U. maydis</i> PGA2.1.....	36 pág

Tablas

Tabla 1. - Tabla Parámetros bioinformáticos de ensamble cepas <i>U. maydis</i> (521, PGA2.1 y PGAtrt1)....	19 pág
Tabla 2. - Tabla Resultados de análisis QUAST ensamble cepa <i>U. maydis</i> (521, PGA2.1 y PGAtrt1).....	23 pág
Tabla 3. - Tabla Contigs de las cepas <i>U. maydis</i> 521 y PGA2.1 relacionados con cromosomas	25 pág
Tabla 4. -Tabla Genes que no se encontraron en cromosomas de la cepa <i>U. maydis</i> PGA2.1 en el catálogo de genes.....	32 pág

Agradecimientos

Quiero expresar mi gratitud a todas las personas e instituciones que contribuyeron de manera significativa a la realización de esta tesis.

Expreso mi agradecimiento al Consejo Nacional de Ciencia y Tecnología (CONACYT) por su apoyo a través de la beca de estudios de posgrado que hizo posible la realización de este proyecto de tesis. También quiero reconocer el respaldo financiero proporcionado por el Instituto de Ciencias de la Benemérita Universidad Autónoma de Puebla y la Vicerrectoría de Investigación y Estudios de Posgrado, que me permitieron llevarlo a su exitosa conclusión. Su apoyo ha sido fundamental en mi camino académico y de investigación.

Agradezco al Laboratorio Nacional de Supercómputo del Sureste de México (LNS) por contribuir a este proyecto.

En primer lugar, deseo agradecer a mi director de tesis, D.C. Ma. Patricia Sánchez Alonso y a mi coasesora D.C Leticia Vega Alvarado, por su orientación y apoyo a lo largo de este proyecto de investigación. Sus consejos y conocimientos fueron fundamentales para el desarrollo de este trabajo.

Agradezco a mi comité de tesis, a D.C. Candelario Vázquez Cruz, M.C. Alejandra Espinosa Taxis, D.C. Estela Anastacio y a D.C. Patricia Lozano Zaráin por su valiosa retroalimentación y sugerencias constructivas que ayudaron a mejorar la calidad de este proyecto.

Mi gratitud se extiende a mis compañeros de laboratorio Adalberto, Adriana, Marina, Hugo, Elena y Miguel y a mis amigas Ceci y Jearim, quienes me brindaron un ambiente de apoyo durante estos años. Sus consejos y apoyo en los momentos difíciles y de éxito fueron fundamentales para el avance de mi proyecto de investigación.

A mi familia, a mis papas (Enedina y Carlos) que estuvieron en todo momento incondicionalmente a lo largo de este proceso de la investigación, por su amor incondicional, aliento constante y comprensión durante esta etapa de mi vida, a mis hermanos (Sandra, Juliana y Juan Carlos) que siempre tenían un sabio consejo ante una situación difícil, su apoyo emocional fue mi mayor fortaleza.

Este trabajo no habría sido posible sin el apoyo de todas estas personas e instituciones. Gracias por creer en mí y en este proyecto.

Resumen

Ustilago maydis es un hongo fitopatógeno estudiado ampliamente con diversos enfoques como patogenicidad, dimorfismo, señalización molecular, control del ciclo celular, biorremediación, metabolismo de RNA, recombinación y reparación de DNA, también se perfila como modelo para el estudio del metabolismo del telómero. La cepa PGA2.1 fue caracterizada previamente en el grupo de Microbiología Molecular y Celular (CICM-ICUAP) y se diferencia de otras cepas por la baja cantidad de elementos asociados al telómero *UTASa*, los cuales son secuencias moderadamente repetidas tipo *a* muy conservadas.

El presente estudio se enfoca en la identificación y análisis de las secuencias de DNA que componen los extremos cromosomales en la cepa patogénica PGA 2.1 de *Ustilago maydis*, porque son cruciales para la estabilidad genómica. Mediante el uso de técnicas de secuenciación de nueva generación (NGS) se han identificado las secuencias teloméricas y las secuencias de las regiones subteloméricas, así como otros elementos que forman parte de esta estructura cromosomal en diferentes organismos. Para consolidar el trabajo se obtuvieron las secuencias del genoma de *U. maydis* cepas 521, PGA 2.1 y la mutante PGA2.1trt1-1, las cuales se analizaron para determinar la longitud telomérica y localizar los elementos de la región subtelomérica de sus cromosomas. Los resultados respecto a los ensamblajes realizados en las cepas de estudio han permitido reconstruir parcialmente la estructura de los extremos cromosomales y revelar su organización y su importancia funcional en *trans* de estas mismas. Los hallazgos en relación a la variabilidad en la longitud de los telómeros en las cepas, así como el mosaicismo genético que se presenta en las regiones cercanas a los extremos teloméricos presentados y la variabilidad de la distribución de secuencias *UTASa*, *UTASb* y *rumT* en las cepas, aquí tienen implicaciones significativas para la comprensión de la biología y la estructura de los extremos cromosomales de *U. maydis*. Además, ofrecen una base para investigaciones futuras sobre la dinámica genómica en hongos patógenos y la evolución de sus extremos cromosomales.

Palabras clave: *Ustilago maydis*, extremos cromosomales, secuenciación de nueva generación, elementos teloméricos, organización de extremos cromosomales

I.-Introducción

Existen estructuras especializadas llamadas telómeros que se encuentran en los extremos de los cromosomas y desempeñan un papel fundamental en la integridad y estabilidad del material genético. Los telómeros actúan como "tapones" protectores, evitando la pérdida de secuencias esenciales de ácido desoxirribonucleico (DNA) en los extremos de los cromosomas. Esto previene la fusión entre cromosomas y la degradación del DNA dañado (Blackburn, 1994; Greider, 1991).

Además de su función en el mantenimiento de la integridad cromosómica, los telómeros están implicados en procesos clave como la replicación del DNA y el desencadenamiento del envejecimiento celular. Esto está relacionado con el "problema de la replicación del extremo final de los cromosomas", ya que a medida que las células se dividen, los telómeros se acortan, lo que está relacionado con el envejecimiento y la senescencia celular (Harley 1990; Blackburn, 1994; Zakian, 1996; Olovnikov, 1972).

Las células eucariotas dependen de los telómeros para mantener la estabilidad y la funcionalidad de los cromosomas, ya que proporcionan una vía para la replicación precisa y completa de los extremos de los cromosomas. Los telómeros son estructuras nucleoproteicas esenciales que participan también en la correcta segregación de la información genética durante la división celular y para preservar la estructura cariológica característica de cada especie. Estos mismos evitan la activación de la maquinaria de respuesta al daño en el DNA, conocida como rotura de doble hebra (DDR) (d'Adda di Fagagna *et al.*, 2003; Grenon *et al.*, 2001)

I.1 Telómeros

Estructura de los extremos cromosomales

Los telómeros consisten en DNA telomérico, proteínas específicas y ácido ribonucleico (RNA) telomérico (Azzalin & Lingner, 2015). Los extremos cromosomales constan de dos dominios constituidos por secuencias repetidas de diferente complejidad como se presenta en la figura 1, así como las diferentes proteínas que se unen a estos dominios, denominadas proteínas del telosoma, o también conocida como shelterina en eucariotes vertebrados (McEachern, et al 2000; Pryde & Louis, 1997; Flint et al., 1997; De Lange, 2018).

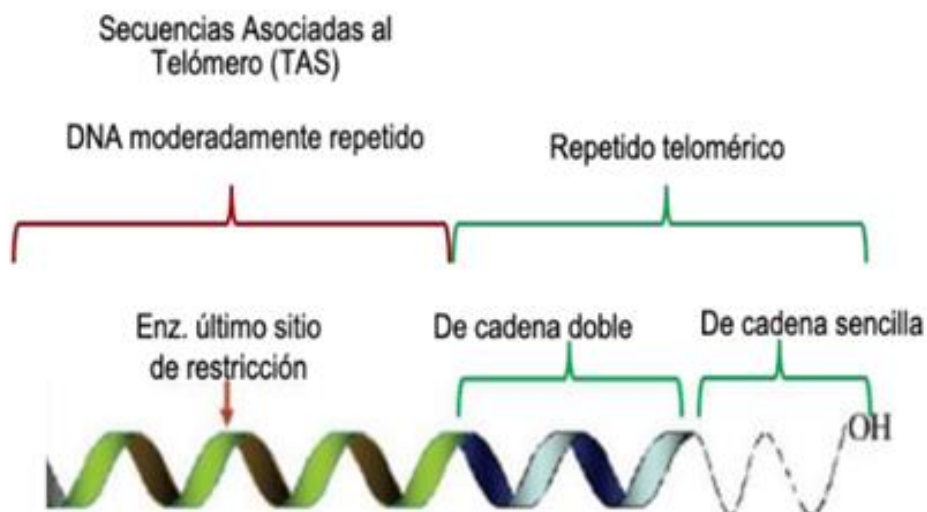


Figura. 1.- La región que se encuentra en el borde físico de los cromosomas se denomina dominio del repetido telomérico, y la región inmediatamente contigua del lado centrómero proximal se denomina secuencias asociadas al telómero o TAS. (McEachern, *et al* 2000; Pryde & Louis, 1997; Flint *et al.*, 1997).

I.1.1 El DNA telomérico

El DNA telomérico consta de motivos de secuencias repetidas en tándem, a menudo cortas, que normalmente contienen grupos de tres o cuatro guaninas (Schrumfová *et al.*, 2020). Estas repeticiones teloméricas son agregadas por la enzima telomerasa, un complejo de ribonucleoproteína que consta de una subunidad similar a la transcriptasa inversa (TERT) y una subunidad de RNA (TER) que contiene la plantilla para la repetición de DNA telomérico (Schrumfová *et al.*, 2020; Peska *et al.*, 2021). La hebra rica en guanina (G) se extiende en la dirección 3', formando un saliente monocatenario que actúa como sustrato para el alargamiento mediado por telomerasa. *In vitro*, estas hebras simples teloméricas ricas en G pueden plegarse en estructuras cuádruples de DNA tipo Hoogsteen (Brazda *et al.*, 2014), que son comunes en todos los organismos y pueden modular la actividad de la telomerasa en algunos organismos (Takahama *et al.*, 2013; Jansson *et al.*, 2019). También se pliegan sobre otras regiones del repetido telomérico, lo invaden y forman media estructura de Holliday y el asa T, o simplemente permiten el plegado del telómero sobre sí mismo (de Lange 2018, y de Bruin *et al.*, 2000).

I.1.2 Proteínas específicas del telómero y RNA telomérico

En mamíferos, las proteínas de unión al telómero son un complejo de 6 proteínas denominado telosoma o shelterina; estas son TRF1 (*Telomeric repeat binding factor 1*) y TRF2 (*Telomeric repeat binding factor 2*) las cuales reconocen exclusivamente al DNA telomérico de doble cadena; estas a su vez interactúan con RAP1 (*Repressor/activator protein 1*), y TIN2 (*T RF1 interacting protein 1*) (de Lange, 2018). Por su parte POT1 (*Protection of telomeres 1*) (Palm & de Lange, 2008) solo reconoce a csDNA rico en G, y TPP1 (*TINT1/P IP1/PTOP 1*) se une a ella para asociarse a las proteínas de unión al DNA telomérico de doble cadena, formando un complejo protector (de Lange, 2018). La proteína TRF2 es una de las proteínas más importantes en la formación del bucle -T y, al igual que TRF1, interviene en la estabilización de dicha estructura (de Lange, 2005).

En el telómero de eucariotes superiores también hay tramos de cromatina histónica. Algunas proteínas de protección del DNA también interactúan con el extremo cromosomal, entre las que se han reportado el complejo MRX/MRN, las helicasas WRB, BLN y el heterodímero Ku70/Ku80 (d'Adda di Fagagna *et al.*, 2003; Singh *et al.*, 2012)

Los complejos de unión a telómero generan un ambiente que impone la represión transcripcional que afecta las secuencias teloméricas y se extiende a las cercanas al telómero, en un fenómeno conocido como efecto de posición telomérica TPE (Stavenhagen & Zakian, 1994). Por ello se pensaba que los finales de los cromosomas no se transcribían; sin embargo, desde hace más de una década se demostró que tanto el telómero como las TAS se transcriben en familias de lncRNAs que incluyen a TERRA y el RNA anti sentido ARIA, así como a dos RNA subteloméricos complementarios denominados ARRET y α -ARRET también a subTERRA, lncRNA-TARE (Greenwood & Cooper, 2012, Bah *et al.*, 2012). Estos transcritos cuya función

comienza a dilucidarse conforman el transcriptoma telomérico (Kwapisz & Morillon, 2020). Los transcritos de la región telomérica se han encontrado formando parte del complejo nucleoproteico del telómero en organismos que van desde levaduras hasta mamíferos. Los indicios más recientes apuntan a la participación de estos transcritos en el mantenimiento de la función telomérica, en el silenciamiento transcripcional y a la regulación del ciclo celular.

TERRA es un RNA heterogéneo cuya alteración en la cantidad de transcripción conduce a anomalías teloméricas (Azzalin & Lingner, 2015). Acerca del mecanismo de transcripción y su regulación poco se sabe. En *Schizosaccharomyces pombe* la hebra telomérica rica en citocina (C) proporciona la plantilla para la transcripción de TERRA a partir de las regiones subteloméricas (Azzalin & Lingner, 2015). En el estudio de Nergadze *et al.* (2009) se identificó que en humanos las islas CpG que se encuentran en los extremos cromosomales promueven la transcripción de moléculas TERRA, esto ocurre mediante las islas de CpG que se encuentran río arriba de los sitios de inicio de transcripción de TERRA en los extremos cromosomales estas islas actúan como promotores y se unen a RNA polimerasa II activa in vivo.

I.1.3 Región subtelomérica

Las regiones subteloméricas son regiones que tienen diferentes funciones, terminan el mantenimiento de los telómeros, la formación de grupos teloméricos, la regulación del estado de la cromatina, la estabilidad cromosómica y la producción de una variedad de transcritos (Kwapisz & Morillon, 2020).

Adyacente al repetido telomérico se encuentra la región subtelomérica, la cual es extraordinariamente compleja, en *S. cerevisiae* las regiones subteloméricas son un mosaico de duplicaciones segmentarias y variaciones estructurales que varían de un telómero a otro (Chan *et al.*, 1983; Louis & Haber., 1992), estas regiones se pueden dividir en dos dominios: la primera es una región proximal del telómero que se encuentra inmediatamente adyacente y se denomina TAS (secuencias asociadas al telómero), la segunda es la región telomérica distal que es centrómero proximal (Pryde & Louis, 1997).

Como parte de las secuencias asociadas al telómero podemos encontrar algunos marcos de lectura, genes o restos de ellos; generalmente hacia el extremo centrómero proximal de las TAS se encuentran algunos genes denominados de contingencia cuya función es dispensable, que proporciona ventaja al organismo que los posee pues pueden servir en ambientes abruptamente cambiantes (Louis *et al.*, 1994; Barry *et al.*, 2003; Kwapisz & Morillon, 2020).

En *S. cerevisiae* existen dos elementos repetidos mayoritarios (X, Y'), como en la mayor parte de las regiones subteloméricas: el elemento Y' el cual se encuentra conservado y adyacente a los repetidos teloméricos, puede estar presente hasta cuatro veces o no estar presente, y el elemento X es una secuencia menos conservada, presenta diversos tamaños, este elemento tiene un parte central con una longitud aproximadamente de 475 pb (Chan *et al.*, 1983; Louis & Haber 1992; Louis, 1995).

En *Ustilago maydis*, nuestro modelo de estudio las secuencias asociadas al telómero también son mayoritariamente dos: *UTASa* y *UTASb* (Ustilago Telomere Associate Sequence) estos elementos repetidos de la región subtelomérica, son estructuralmente similares a los elementos

Y y X' que se encuentran en *S. cerevisiae*. En *U. maydis* como se mencionó anteriormente se han encontrado dos tipos principales de UTAS: *UTASa* el cual se está altamente conservado en la secuencia de nucleótidos, este se encuentra en los extremos cromosómicos y puede estar en algunos cromosomas, pero no en todos, codifica un marco de lectura para una helicasa relacionada con la familia de las RecQ helicasas, el otro elemento hallado es *UTASb* el cual está menos conservado y presente en todos los cromosomas. Tanto *UTASa* como *UTASb* contienen una región común de unos 300 pb situada inmediatamente al lado de las repeticiones de los telómeros que también se encuentran intercaladas en el genoma (Sánchez-Alonso & Guzmán, 1998).

I.1.4 Genes de contingencia.

En algunos eucariotas como *S. cerevisiae*, *S. pombe*, y *P. falciparum*, se han encontrado genes que se encuentran codificados cercanos a regiones subteloméricas proximales (Denayrolles *et al.*, 1997; Ness & Aigle, 1995), estos genes, conocidos como genes de contingencia no son esenciales, pero aportan características específicas que les confieren ventajas en ambientes que cambian abruptamente. En *P. falciparum* y *Trypanosoma brucei* la localización de los genes *var* (variación antigénica) en las regiones subteloméricas permite la evasión de la respuesta inmune a través de la expresión monoalélica de algún miembro de esta familia génica. Aunque el control de los cambios en la expresión transcripcional de estos genes de contingencia no se ha comprendido completamente; eventos de recombinación ectópica de genes silenciosos a sitios de expresión activa le permite la variación en la expresión de la proteína antigénica y evasión de la respuesta inmune (Deitsch *et al.*, 1997; Henderson *et al.*, 1999; Barry *et al.*, 2003). Existen algunos genes de contingencia de *S. cerevisiae* localizados en la región subtelomérica que se encargan del catabolismo del azúcar (Henderson *et al.*, 1999; Kwapisz & Morillon, 2020).

I.2 Mantenimiento del extremo cromosomal

La telomerasa es un complejo de múltiples subunidades en la que destacan dos componentes centrales: una subunidad proteica que tiene la función catalítica de transcriptasa reversa especializada en la síntesis de telómeros denominada TERT y una subunidad de RNA (TER) que porta el molde para la síntesis de los mismos. Estos componentes forman el centro catalítico de la enzima, los sustratos que utiliza en su reacción son: desoxinucleótido 5'-trifosfatos y el extremo 3' de cadena sencilla del telómero (en las pruebas *in vitro* es el DNA que contiene la secuencia correspondiente a las repeticiones teloméricas de los cromosomas), la telomerasa es capaz de adicionar varios repetidos teloméricos durante un solo ciclo de unión al sustrato (Kelleher *et al.*, 2002) como se muestra en la figura 2 (Zvereva *et al.*, 2010).

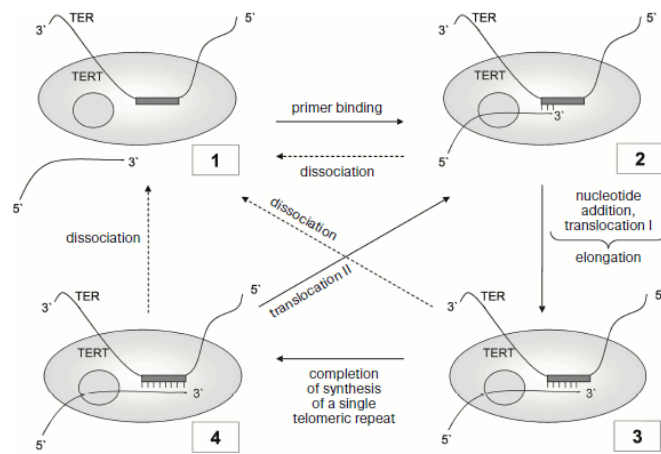


Figura 2.-El ciclo de reacción de la telomerasa *in vitro* se compone de varias etapas. En esta figura, podemos ver dos partes clave: TERT, que es la parte de la telomerasa responsable de catalizar la reacción (se muestra en gris). El círculo señala el lugar donde se une TER, que es una parte del ARN de la telomerasa que actúa como una especie de plantilla (se muestra como un rectángulo gris). Las etapas se indican en recuadros: En la primera etapa, la enzima no se une al cebador, en la segunda etapa, se inicia un emparejamiento, en la tercera etapa, ocurre el alargamiento, por último, en la cuarta etapa, se termina la síntesis de una sola repetición telomérica. Las flechas punteadas indican posibles procesos de disociación del cebador durante el funcionamiento de la enzima. Tomado de Zvereva et al. (2010).

Durante el proceso de extensión de los telómeros, la telomerasa utiliza repetidamente la región templete dentro de su fracción de RNA como molde para la síntesis de DNA, esto requiere una reacción cíclica. Aunque los mecanismos que impulsan este ciclo de reacción no se conocen del todo, está claro que los cambios en las propiedades bioquímicas de la telomerasa durante el ciclo de reacción se producen en las transiciones entre los diferentes estados, los cambios más evidentes son aquellos en los cuales se debe alcanzar el límite 5' de la plantilla, que completa una ronda de adición del repetido telomérico (Hammond & Cech, 1998).

Existen procesos de translocación en la vía de la telomerasa en experimentos *in vitro*; el proceso tipo I es la capacidad de la telomerasa para la translocación dúplex de RNA-DNA en el centro activo después de cada adición de nucleótidos en la etapa de elongación, el proceso tipo II es la capacidad de la telomerasa para la translocación relativa del cebador DNA unido después de la adición de una repetición telomérica, después de lo cual el cebador de nuevo tiene la capacidad de realizar alargamiento (Zvereva et al., 2010).

La telomerasa no es el único mecanismo que puede mantener el DNA telomérico mediante transcripción reversa. En *Drosophila*, la transposición de retrotransposones específicos de los telómeros es la vía principal para el mantenimiento de los telómeros (Biessmann et al., 1990).

I.2.1 Vías alternativas.

En organismos que normalmente dependen de la telomerasa, también existen mecanismos de mantenimiento de los telómeros independientes de ella (Lundblad & Blackburn, 1993).

Cuando *S. cerevisiae* carece del gen que codifica para alguno de los componentes centrales de la telomerasa, la mayoría de las células muere, debido al arresto celular producido por el acortamiento crítico del telómero (Lundblad & Szostak, 1989; Lundblad & Blackburn, 1993), sin embargo, entre la población de células en crisis surgen algunos supervivientes con relativa frecuencia. En la mutante *est2* de *S. cerevisiae*, se identificaron supervivientes que surgían después de aproximadamente 50 a 100 generaciones. Con base a este estudio se demostró que el alargamiento del extremo cromosomal puede ocurrir independiente de la actividad de telomerasa, y que la recombinación telómero-telómero es una vía de derivación eficiente para

el mantenimiento de los extremos cromosomales, que presentan una tasa de crecimiento menor al de las células de las cepas de tipo silvestre (Lundblad & Blackburn, 1993).

También se han descrito dos mecanismos independientes de telomerasa para el mantenimiento de telómeros en *S. pombe*: el primero es mediante la amplificación de sus repeticiones TA, probablemente por recombinación, el segundo es a través de la pérdida tanto de TA (secuencias asociadas al telómero) como del DNA telomérico seguido de fusiones de extremo a extremo para generar cromosomas circulares (Nakamura *et al.*, 1998).

II.- Antecedentes

U. maydis es dimórfico, y heterotálico; su sistema de apareamiento es tetrapolar, controlado por dos loci de apareamiento no ligados genéticamente.

Durante la infección, *U. maydis* experimenta un cambio en su patrón de crecimiento de un estado levaduriforme a un estado filamentososo el cual es desencadenado por la interacción de las moléculas ferohormona/receptor del apareamiento *a*, completada posteriormente por la interacción exitosa de los péptidos bE y bW codificados por los locus *b* del tipo de apareamiento que heterodimerizan cuando los apareantes son compatibles. El proceso de transición es crucialmente dependiente de la actividad del heterodímero bE/bW, y se ha demostrado de manera concluyente que la sobreexpresión del complejo bE/bW en células haploides es suficiente para inducir el crecimiento filamentososo. Los síntomas más distintivos de la enfermedad son los tumores en las partes aéreas de la planta, que resultan de las alteraciones en el crecimiento de los tejidos de estas (Hartmann *et al.*, 1996; Brachmann *et al.*, 2001; Matei & Doehlemann, 2016).

En las figuras 3 y 4 se muestran tumores producidos por *U. maydis* en la mazorca del maíz (izquierda) y los granos del maíz deformados por la invasión del hongo (derecha).



Figura 3.- Imagen Mazorca de maíz con tumores de *Ustilago maydis*.



Figura 4.- Imagen Tumores de *Ustilago maydis*.

El tamaño del genoma haploide de *U. maydis* es de ~20Mb, organizado en 23 cromosomas y contiene 6522 genes (Kämper *et al.*, 2006). Este tamaño y número de cromosomas es posible analizar en geles de campos pulsados, el cariotipo puede ser teñidos con bromuro de etidio, o se puede usar sondas para la localización de genes en autorradiografía.

Los extremos cromosomales de *U. maydis* al igual que en otros organismos, marcan los límites de cada uno de los elementos de ligamiento génico, es decir los cromosomas, estos están constituidos por repetidos de la secuencia TTAGGG, arreglados en tándem formando tramos de un tamaño promedio de 37 a 39 repetidos teloméricos (Guzmán & Sanchez, 1994; Sánchez-Alonso *et al.*, 1996), aunque recientemente se ha propuesto una cifra mayor de repetidos (Yu *et al.*, 2020). Hacia el extremo centrómero proximal de la región telomérica se encuentran dos tipos de elementos estructurales asociados al telómero, previamente denominados *UTASa* y *UTASb* (Sánchez-Alonso & Guzmán, 1998).

En *S. cerevisiae*, se ha visto que los elementos Y' sobre los diferentes cromosomas además de ser polimórficos también pueden ser dispensables sobre algunos de ellos (Tham & Zakian, 2002). Esto es importante debido al estudio que puede hacerse en cepas con diferente número de repetidos de TAS en cepas de otros organismos, como en *U. maydis*.

Después de un estudio extenso sobre la estructura del extremo cromosomal de cepas nativas de *U. maydis*, se identificó una cepa que tiene un número limitado de copias de *UTASa* caracterizada como PGA2.1, esta cepa contiene baja cantidad de *USHER*. Se ha mostrado un gran interés en identificar las secuencias de DNA que estructuran los elementos esenciales en cis de los cromosomas, así como el arreglo que tienen; de los telómeros de este hongo, aun se tienen que analizar, los elementos presentes en los extremos cromosomales.

En la figura 5 se muestran en los paneles de la izquierda y medio los patrones de hibridación de las secuencias rumT y UTASa utilizando las sondas rumT y UT4-a (fragmento de UTASa). El DNA total de varias cepas de *Ustilago maydis* fue digerido con PstI y resuelto por electroforesis en agarosa 0.8% en TBE. El DNA fue transferido a Nylon y sometido a hibridación. En el recuadro rojo se muestra que PGA2.1 presenta un bajo número de copias UTASa. En el panel de la derecha se muestra el cariotipo electroforético de dos de las cepas aquí estudiadas (521 y PGA2.1) así como una tercera (I2) no incluida en este estudio.

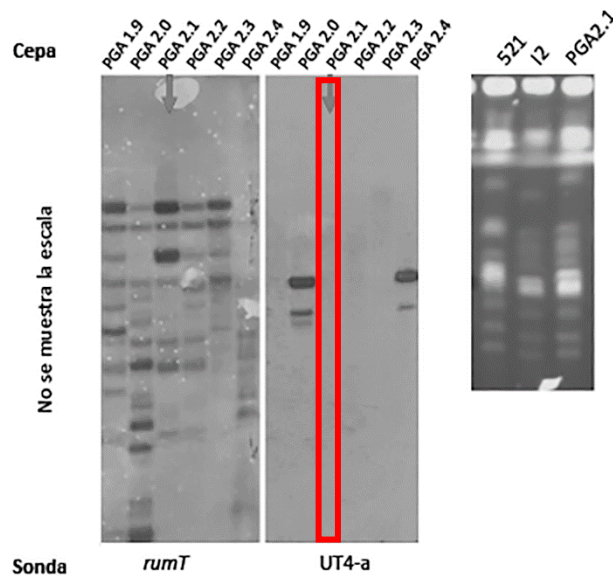


Figura 5.- Cromosomas de la Cepa (PGA 2.1) de *U. maydis* en la electroforesis.

III.- Justificación

La identificación y análisis de los elementos genéticos presentes en las secuencias subteloméricas de *U. maydis* silvestre y carente de actividad telomerasa serían de gran utilidad para comprender los mecanismos alternativos de mantenimiento del extremo cromosomal, así como la importancia de sus secuencias asociadas al telómero bajo estrés replicativo.

IV.- Hipótesis

El estudio de la estructura de las TAS en cepas con alto y bajo número de copias de estos elementos subteloméricos permite conocer el arreglo de secuencias en el extremo cromosomal de *U. maydis*.

V.- Objetivo general

Identificación de las secuencias del extremo cromosomal de *U. maydis* cepa PGA 2.1 y análisis de su arreglo.

V.1 Objetivos específicos

1. Obtener las secuencias del genoma de *U. maydis* 521, PGA2.1 y su derivada superviviente telomerasa-negativa mediante el uso de una plataforma de secuenciación tercera generación.
2. Ensamblar los genomas de las cepas *U. maydis* 521, PGA2.1 y su derivada superviviente telomerasa-negativa.
3. Analizar las secuencias de los extremos cromosomales, y elucidar del arreglo de las secuencias moderadamente repetidas asociadas al telómero (TAS) encontradas.

VI.- Esquema de trabajo

En el siguiente esquema se muestran las etapas de trabajo del presente proyecto. Inicialmente se realizó el cultivo de las cepas para extraer el DNA total, el cual se envió a secuenciar con tecnología PacBio. Posteriormente se evaluaron las lecturas obtenidas para realizar el ensamble y luego analizar los datos obtenidos del mismo.

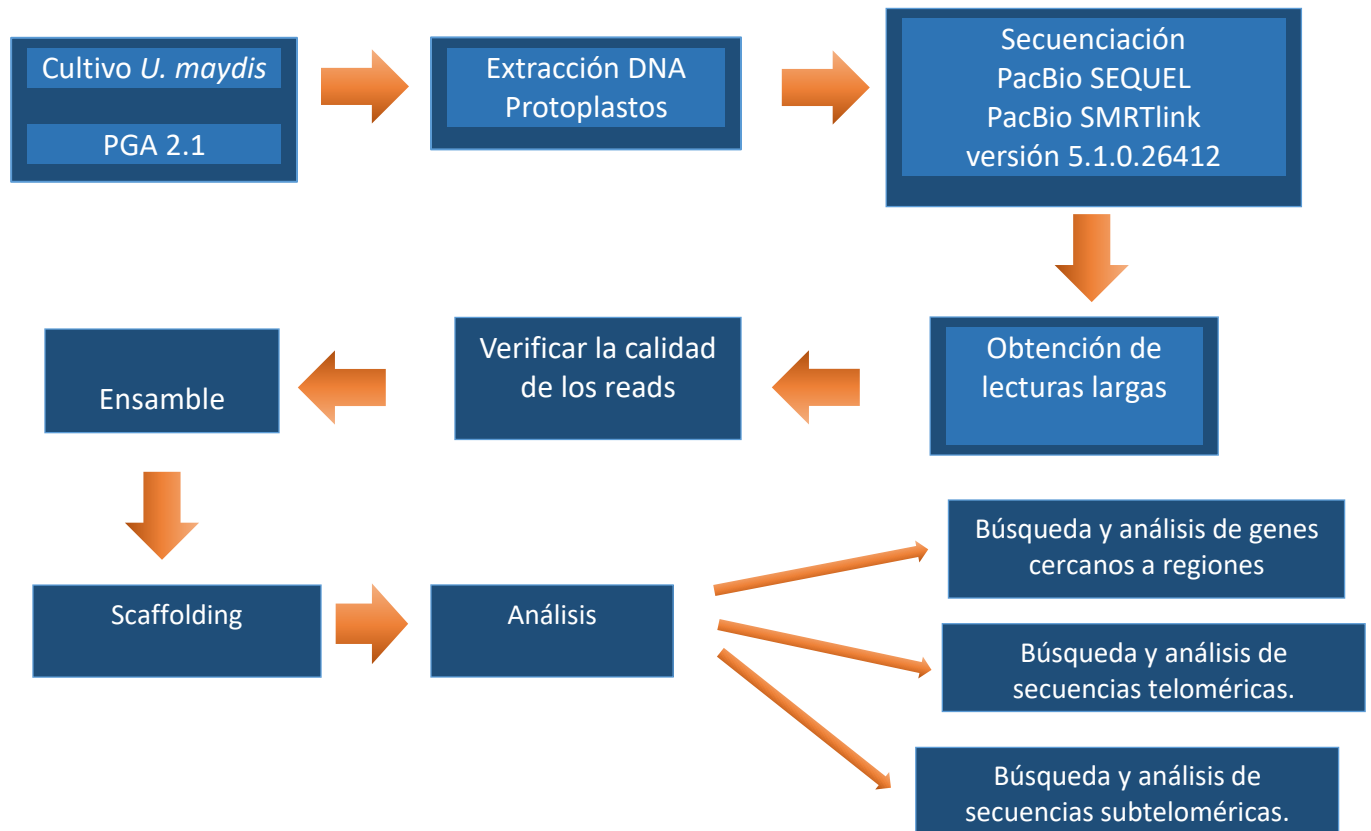


Figura 6.-Esquema de trabajo para la obtención de la secuencia de tres genomas y el análisis de los extremos cromosomales.

VII.- Materiales y métodos

VII.1 Materiales

Las cepas de *U. maydis* utilizadas en este trabajo fueron 521 (*a1, b1*) donada por el Dr. W.K. Holloman (Department of Microbiology y Immunology, Cornell University Weill Medical College, NY, USA), PGA2.1 un aislado silvestre de la colección del laboratorio de Microbiología Molecular y Celular (MMC) y, su derivada telomerasa negativa, PGAttr1-4, obtenida en el laboratorio CICM-MMC-ICUAP por el grupo de trabajo.

Las cepas de *U. maydis* se crecieron en medio YEPS (1% extracto de levadura, 2% peptona, 2% sacarosa) y agar PDA, en condiciones de temperatura de crecimiento entre 28 - 30 °C; se tomó un inóculo de una placa de cultivo fresco y posteriormente fue cultivado en matraces con aeración a 120 rpm.

VII.2 Métodos

VII.2.1 Secuenciación y ensamble

Tres cepas de *U. maydis* 521 (*a1, b1*), PGA2.1 (*a2, bx*) y su derivada PGAttr1-4 fueron secuenciadas en este trabajo. Para la secuenciación del genoma, se extrajo el DNA de alto peso molecular de las cepas y se enviaron a secuenciar en la plataforma PacBio Sequel II en la empresa DNALink Sequencing Lab. Los Angeles CA, USA. La calidad de las lecturas obtenidas se analizó usando el programa FastQC (Andrews,2010), Los datos evaluados incluyeron distribución de calidad de bases, longitud de lecturas, calidad phred, contenido de GC y presencia de adaptadores. Para el ensamblado de las secuencias de las cepas 521 y PGA2.1, y PGA2.1trt1 se utilizó el programa Canu 1.5 (Koren et al., 2017), llevando a cabo tres etapas, en la primera se realizó una evaluación de la calidad, en la segunda, etapa de recorte, se eliminaron adaptadores, en la última etapa se realizó el ensamblaje de las secuencias, en esta etapa se probaron diversos tamaños de k-meros y se eligieron tamaños óptimos para los genomas. Para evaluar la calidad de los ensamblados generados, se utilizó el programa Mauve (Darling, Mau, Blattner, & Perna, 2004) utilizando las secuencias de las cepas 521 y PGA2.1, y PGA2.1trt1. Como referencia se utilizó el genoma de la cepa 521 obtenida por Kämper et al. (2006) GenBank No. (GCA_000328475.2). Las secuencias de referencia se descargaron de Gen Bank (GCA_000328475.2), disponible 21 de febrero 2022 en https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000328475.2/. Conexión 21 de febrero 2022, de Microbiología Molecular y Celular (ICUAP).

VII.2.2 Construcción de catálogo de genes cercanos a región telomérica

En el genoma de las cepas 521 (*a1, b1*), PGA2.1 se identificaron a los genes situados en las secuencias del 50 kb centrómero-proximal de la región telomérica con el uso de BLASTX

(Altschul *et al.*, 1990), y a partir de este resultado se fueron identificando y etiquetando cada uno de los genes encontrados en los 23 cromosomas de ambas cepas.

VII.2.3 Identificación de secuencias teloméricas en el genoma

La búsqueda de las secuencias teloméricas en los cromosomas de *U. maydis* 521, PGA2.1 y PGAttr1 se realizó utilizando los programas SnapGene (GSL Biotech, available at snapgene.com) y UniproUGENE (Okonechnikov *et al.*, 2012), identificando la secuencia "TTAGGG" en cada uno de los 23 contigs reconstruidos que corresponden a cada uno de los cromosomas.

VII.2.4 Identificación de secuencias asociadas al telómero *UTASa* y *UTASb* en el genoma.

La identificación de las secuencias *UTAS* en los cromosomas del genoma de *U. maydis* 521 y PGA2.1 se realizó con la ayuda del programa BLASTX para la localización, arreglo y distribución de estas secuencias por cromosoma (AF030886.2, AF030887.1, AF030888.1, AF030889.1.) (Altschul *et al.*, 1990).

VIII. Resultados

VIII. 1 Análisis de calidad de las secuencias

Para determinar la calidad de las secuencias, se utilizó el software FastQC, que realiza un análisis de diversas métricas, como la calidad de las bases, la distribución de la longitud de las lecturas y la presencia de adaptadores. En las figuras 7,8 y 9 se muestran los resultados del análisis en los que se observa que la calidad general de las secuencias fue alta, con valores de puntuación de calidad (phred scores) de 33 en más del 90% de las bases, también se observa la distribución de la longitud de las secuencias, de donde se deduce una alta confiabilidad de bases por secuencia.

La evaluación de calidad de las secuencias y posterior filtrado proporcionó una base sólida para el siguiente proceso de ensamblaje. El enfoque riguroso en la calidad de las secuencias garantiza la precisión y confiabilidad en los resultados del ensamblaje, y su posterior análisis bioinformático.

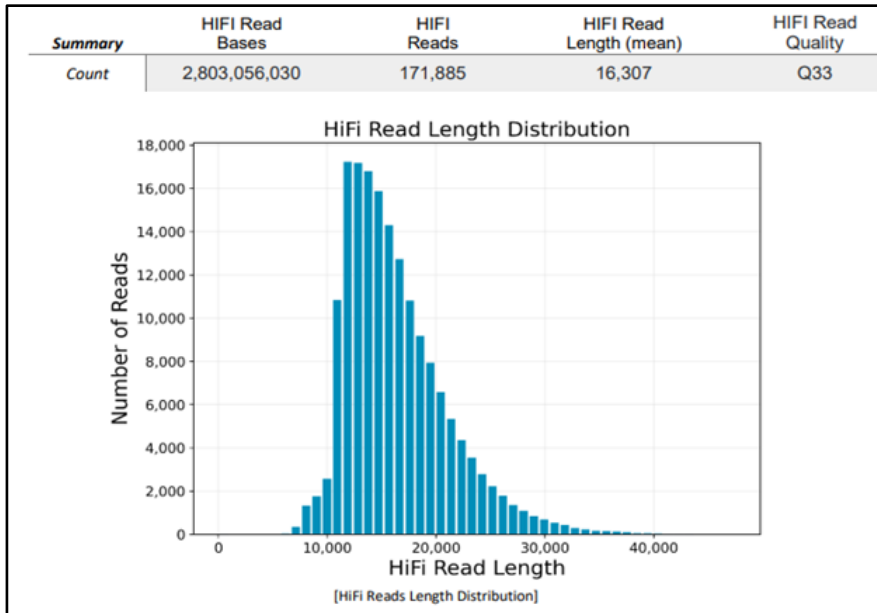


Figura 7. - Gráfica de distribución de secuencias largas tipo PacBio de la cepa *U. maydis* (521).

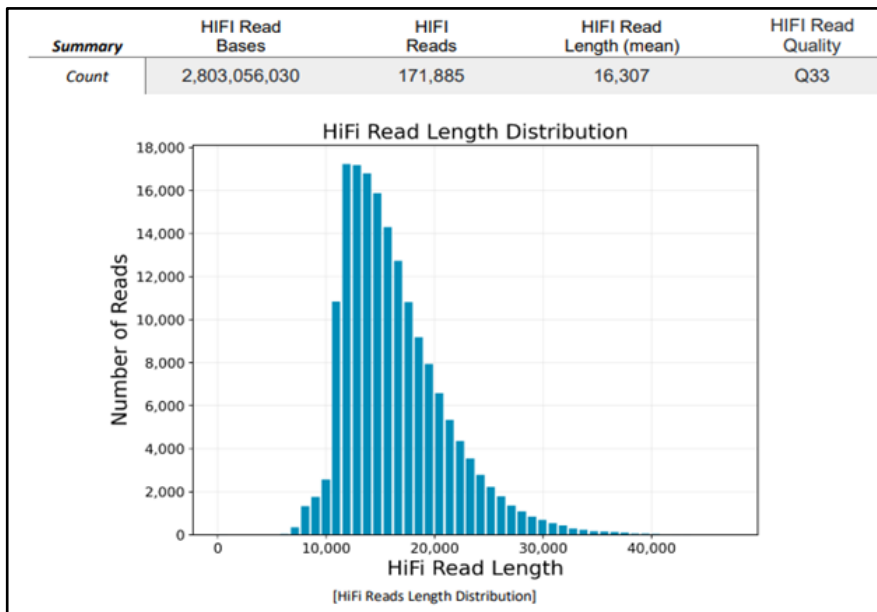


Figura 8. - Gráfica de distribución de secuencias largas tipo PacBio de la cepa *U. maydis* (PGA2.1).

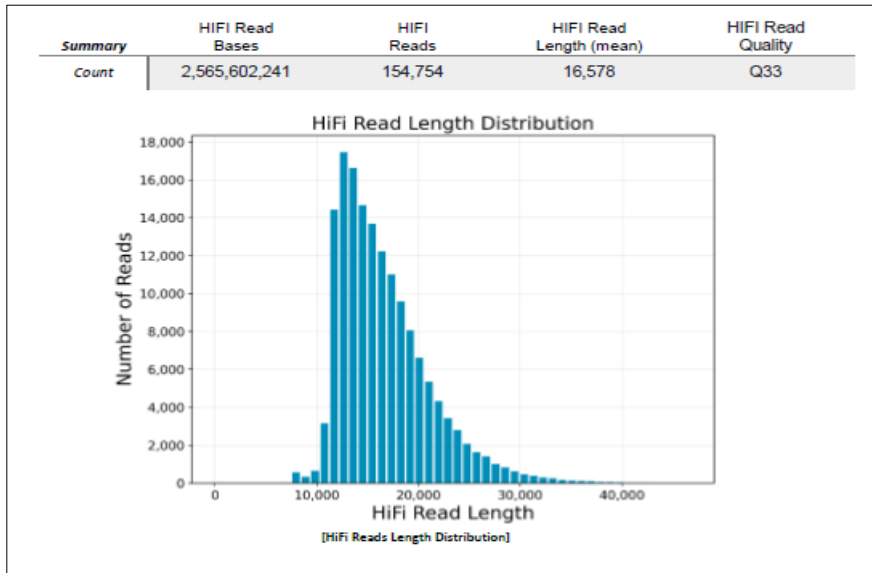


Figura 9. - Gráfica de distribución de secuencias largas tipo PacBio de la cepa *U. maydis* (PGAttr1-4).

VIII. 2 Ensamble de secuencias

Los ensambles de las cepas *U. maydis* 521 (a1, b1), PGA2.1 y PGAttr1-4 se realizaron con el programa Canu (versión 1.5) a partir de las secuencias de lecturas generadas mediante la tecnología de secuenciación PacBio Sequel II (**DNA Link USA**).

El proceso de ensamblaje se realizó con los datos de secuenciación que correspondieron a una cobertura de 126x, con lecturas de longitud promedio de 16.3 kbp (521), 16.68 kbp (PGA2.1) y 16.57 kbp y (PGAttr1-4), para la ejecución de Canu se tomaron en cuenta parámetros como el tamaño de los genomas, longitud de lecturas y la cobertura calculada, con ello se obtuvo un ensamblaje de 24 contigs para cada una de las cepas.

En la tabla 1 se encuentran los parámetros bioinformáticos de los ensambles, así como el número de contigs, lo cual fue un punto a considerar debido a que en estos se obtuvieron 24 contigs que, se relacionaron con 23 cromosomas, que coinciden con lo reportado en GenBank para el genoma., El obtener menor cantidad de contigs nos indica una menor fragmentación del genoma que para nuestro trabajo es primordial debido a que se busca mantener secuencias más largas e intactas para facilitar el ensamblaje de regiones teloméricas y subteloméricas.

Parámetro	Resultado <i>U. maydis</i> 521	Resultado <i>U. maydis</i> PGA2.1	Resultado <i>U. maydis</i> PGAttr1
# Contigs	24	24	24
Largest contig	2,485,961	2,485,571	2,478,813
Total Length	20,239,569	19,662,490	19,735,011
% CG	53.90	53.91	53.92
N50	919,653	921,999	888,049
L50	7	7	7

Tabla 1. - Parámetros bioinformáticos de ensamble cepas *U. maydis* (521, PGA2.1 y PGAttr1).

El tamaño total de los genomas ensamblados fue de aproximadamente 20 Mbp (521), 19 Mbp (PGA2.1) y 19 Mbp (PGAtr1-4), siendo el tamaño de las cepas PGA2.1 y PGAtr1-4 muy similar en los ensamblajes obtenidos.

VIII. 2 .1 Calidad de ensamble de secuencias *U. maydis*

La métrica del número total de contigs reveló un total de 24 contigs como resultado del proceso de ensamblaje mediante QUAST y Mauve.

La longitud del ensamblaje alcanzó un tamaño total de 19, 662,490 bases, que se asemeja al tamaño estimado del genoma de referencia (GCA_000328475.2). El valor de N50 de 921,999 bases y L50 de 7 contigs sugiere parámetros buenos de reconstrucción en el análisis del ensamble.

Una observación en la cepa PGA2.1, es la longitud de los contigs generados en el ensamble en donde, el contig de menor longitud es de 334 ,946 pb y el de mayor longitud es de 2,482 ,571 pb de los 23 contigs generados considerando la longitud promedio de los cromosomas de la cepa 521 representa un tamaño adecuado considerando el tamaño del genoma.

VIII. 2 .2 Resultados de la evaluación de ensamblajes con la herramienta Mauve.

La evaluación de los ensamblajes genómicos mediante MAUVE proporcionó una visión de la calidad, la conservación y la estructura de los genomas ensamblados con relación al genoma de referencia. En las figuras 10,11 y 12, se presentan los resultados obtenidos de esta evaluación.

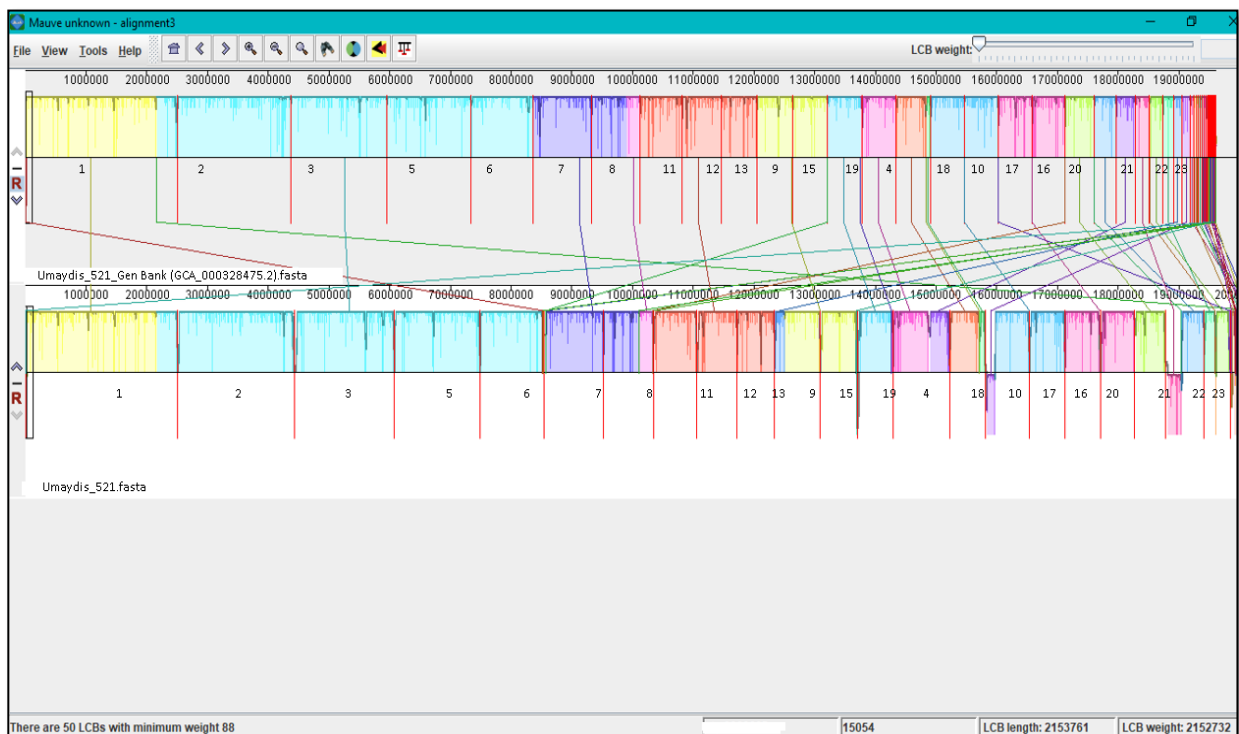


Figura 10. - Alineamiento del genoma de referencia *U. maydis* 521 GenBank comparado con el ensamble CANU *U. maydis* 521 utilizando el programa MAUVE. Los 50 bloques localmente colineales (LCBs) encontrados por MAUVE relacionan regiones de secuencia de ambos genomas con las líneas diagonales, cada bloque corresponde a una secuencia en particular, los números debajo de los bloques coloreados indican los bloques asociados (cromosomas o parte de ellos) entre las cepas. 20

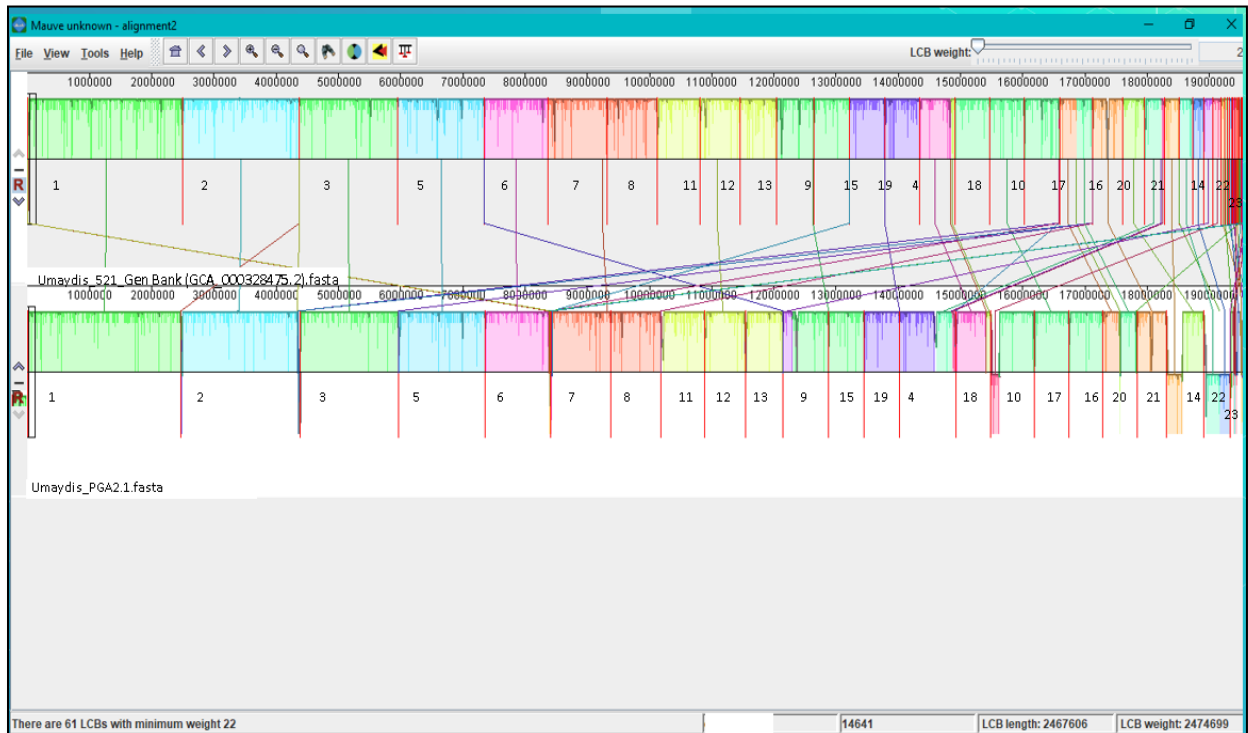


Figura 11. - Alineamiento del genoma de referencia *U. maydis* 521 GenBank comparado con el ensamble CANU *U. maydis* PGA2.1 utilizando el programa MAUVE. Los 61 bloques localmente colineales (LCBs) encontrados por MAUVE relacionan regiones de secuencia de ambos genomas con las líneas diagonales, cada bloque corresponde a una secuencia en particular, los números de bajo de los bloques coloreados indican los bloques asociados (cromosomas o parte de ellos) entre las cepas.

La comparación entre el ensamble del genoma de *U. maydis* (521) realizado con CANU y el genoma de referencia de GenBank reveló una concordancia sustancial., En la figura 10 se puede observar en bloques de colores las regiones de los genomas de 521(GenBank) y 521(CANU) indicado la alineación de esas regiones con los bloques presentes en los genomas, las líneas de colores cruzadas que se observan entre los bloques son translocaciones que se presentaron en el genoma 521 ensamblado con CANU, estas presentan un tamaño promedio de 80 a 100 pb, a pesar de estas translocaciones, el alineamiento de los genomas fue más del 99.19 % ,esto considerando algunas diferencias por la plataforma de secuenciación utilizada

En la figura 11 se muestra el resultado del alineamiento de los genomas de *U. maydis* PGA2.1 ensamblado con CANU y el genoma de la cepa de referencia 521 de GenBank; patrones de colinealidad en la mayoría de los bloques es buena considerando que es un genoma de una cepa diferente, las líneas diagonales de colores que conectan los bloques indican translocaciones en el genoma PGA2.1. A pesar de las translocaciones, el alineamiento arroja una similitud 96 % estructural, la cual es significativa a pesar de provenir de diferentes cepas.

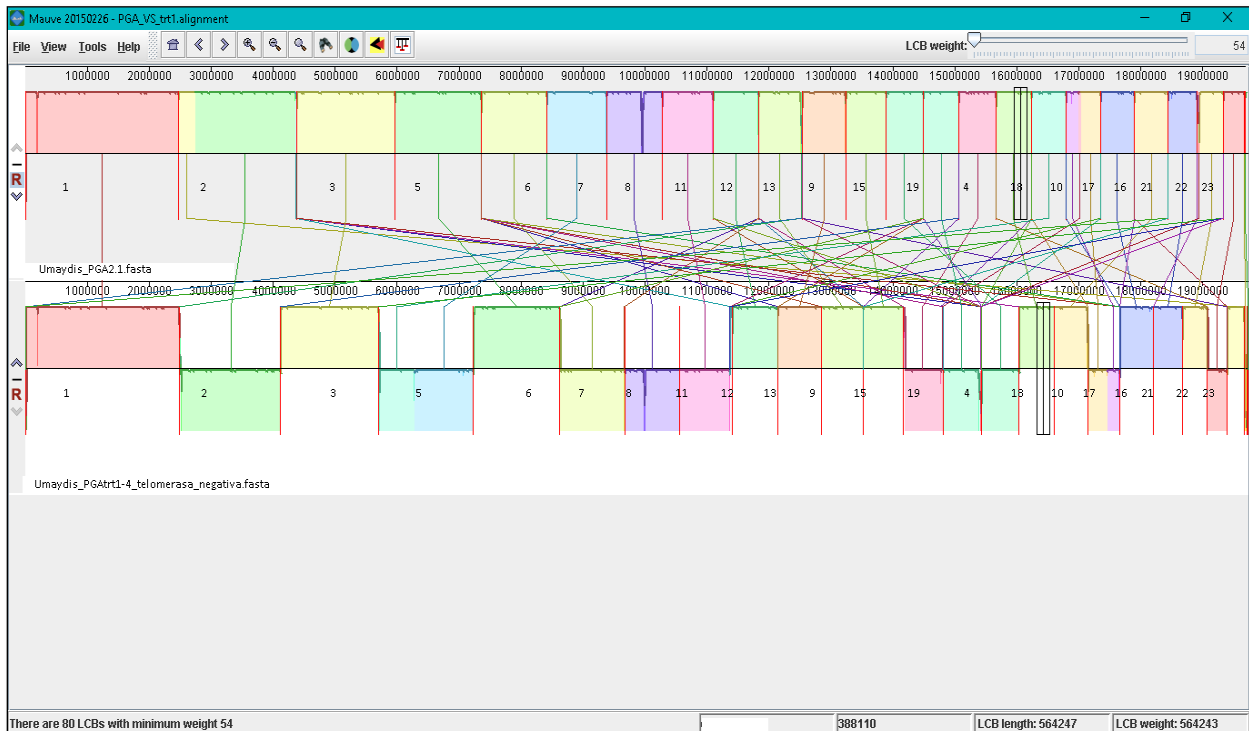


Figura 12. – Alineamiento del genoma de *U. maydis* PGA2.1 comparado con el ensamble CANU *U. maydis* PGA2.1tert1 utilizando el programa MAUVE. Se encontraron 80 bloques localmente colineales (LCBs) por MAUVE. Los bloques de secuencia de ambos genomas que se encuentran alejados son unidos con líneas diagonales e indican posibles translocaciones, cada bloque corresponde a una secuencia en particular, los números en los bloques coloreados identifican los bloques que se asocian (cromosomas o parte de ellos) entre las cepas.

En la figura 12 se muestra el patrón de colinealidad de las cepas PGA2.1 y PGAttr1, con patrones de colinealidad complicados que los alineamientos anteriores, el número de interacciones encontradas para el alineamiento es de 88 predominando en la parte central del alineamiento de los genomas; el porcentaje de similitud arrojado por el ensamble fue de 90%. Este resultado es interesante porque posiblemente refleja la inestabilidad genómica de la cepa después de la pérdida de la telomerasa. Un análisis informático más detallado podría generar los mapas de los sitios de posibles translocaciones y sería interesante determinar si ocurre en el extremo cromosomal.

En las figuras 11 y 12 muestran que a pesar de la similitud de los genomas de PGA2.1 y de su derivada, los eventos recombinacionales que han llevado a la supervivencia de la cepa PGA2.1trt1 también han generado una cantidad muy grande de rearrregos (posiblemente translocaciones, inversiones, deleciones, duplicaciones, etc.) similar a las que exhiben algunos cánceres, aunque en este caso son cepas telomerasa negativas, y en la mayoría de los cánceres ocurre una reactivación de la enzima. Sería interesante conocer si los rearrregos preceden a la reactivación de la telomerasa en humano.

VIII. 2.3 Resultados de la evaluación de ensamblajes con la herramienta QUASt (Quality Assessment Tool) para los ensamblajes genómicos.

Parámetros bioinformáticos	<i>U. maydis</i> 521	<i>U. maydis</i> PGA2.1	<i>U. maydis</i> PGAttr1-4
# contigs (>= 0 bp)	24	24	24
# contigs (>= 1000 bp)	24	24	24
# contigs (>= 5000 bp)	24	24	24
# contigs (>= 10000 bp)	24	24	24
# contigs (>= 25000 bp)	24	24	24
# contigs (>= 50000 bp)	24	24	24
Total length (>= 0 bp)	19,766,249	19,786,289	19,775,658
Total length (>= 1000 bp)	19,766,249	19,786,289	19,775,658
Total length (>= 5000 bp)	19,766,249	19,786,289	19,775,658
Total length (>= 10000 bp)	19,766,249	19,786,289	19,775,658
Total length (>= 25000 bp)	19,766,249	19,786,289	19,775,658
Total length (>= 50000 bp)	19,766,249	19,786,289	19,775,658
# contigs	24	24	24
Largest contig	2,485, 961	2,485,571	2,482,571
Total length	20,239,569	19,735,249	19,735,011
GC (%)	53.90	53.91	53.92
N50	919,653	921,999	888,049
N75	600,785	600,655	600,668
L50	7	7	7
L75	14	14	14
# N's per 100 kbp	0.00	0.00	0.00

Tabla 2. – Resultados de análisis QUASt para el ensamblaje de las cepas *U. maydis* 521, PGA2.1 Y PGAttr1.

Los resultados de la evaluación de los ensamblajes genómicos mediante la herramienta QUASt proporcionó un análisis cuantitativo y cualitativo de la calidad de los genomas reconstruidos en comparación con la referencia del genoma *U. maydis* 521 (Kämper et al., 2006).

En la tabla 2 están los resultados del genoma ensamblado de la cepa 521 tiene una longitud total de 20,239,569 pares de bases (pb), se encuentra dividido en 24 contigs; de estos contigs, al menos la mitad, es decir, 12 contigs, tienen una longitud promedio de 919,653 pb, lo que indica una buena continuidad en la representación del genoma. Además, se observa que los siete contigs más largos en el ensamblaje, que representan el 50% del genoma ensamblado, contribuyen significativamente a la cobertura y continuidad del ensamblaje. Estos resultados

sugieren un ensamblaje con una considerable representación y continuidad del genoma de la cepa 521.

El genoma ensamblado de la cepa PGA2.1 tiene una longitud total de 19,735,249 pares de bases (pb), contenido en 24 contigs, de estos al menos la mitad tienen una longitud promedio de 921,999 pb, indicando buena continuidad del mismo, muestra un L50 de 7 que indica el número de contigs que representan el 50% del genoma ensamblado. Para cepa PGAttr1 se muestra un total de ensamble de 19,735 ,011 pares de bases (pb), en donde se encuentra constituido por 24 contigs, teniendo al menos la mitad de estos una longitud promedio de 888, 049 pb y un L50 al igual que los otros ensambles de siete contigs.

VIII.3 Genes cercanos a los extremos teloméricos, y secuencias teloméricas en los cromosomas de *U. maydis*.

Con el objetivo de explorar la influencia de las regiones teloméricas en los genes cercanos a los extremos teloméricos, se realizó la búsqueda de genes de los 50 kb posteriores a cada región telomérica en ambas cepas (521 y PGA2.1). Identificando genes codificantes de proteínas en estas regiones, analizadas en los contigs.

Cromosoma <i>U. maydis</i> 521 GenBank	No de contig ensamble <i>U. maydis</i> 521	No de contig ensamble <i>U. maydis</i> PGA2.1
1	contig 0001	tig00000001
2	contig 0003	tig00000003
3	contig 0006	tig00000006
4	contig 00014	tig00000014
5	contig 00008	tig00000008
6	contig 00010	tig00000010
7	contig 00012	tig00000012
8	contig 00018	tig00000018
9	contig 00021	tig00000021
10	contig 00016	tig00000016
11	contig 00019	tig00000019
12	contig 00024	tig00000024
13	contig 00038	tig00000038
14	contig 00026	tig00000026
15	contig 00036	tig00000036
16	contig 00043	tig00000043
17	contig 00028	tig00000028
18	contig 00032	tig00000032
19	contig 00030	tig00000030
20	contig 00040	tig00000040
21	contig 00034	tig00000034

22	contig 00045	tig00000045
23	contig 00047	tig00000047
No asociado con algún cromosoma.	contig 00049	contig 00049

Tabla 3. - Tabla Contigs de las cepas *U. maydis* 521 y PGA2.1 relacionados con cromosomas de la cepa *U. maydis* referencia GCA_000328475.2.

A continuación, en las figuras 13,14,15 y 16 se muestran la presencia de una variedad de genes en las secuencias cercanas a las regiones teloméricas.

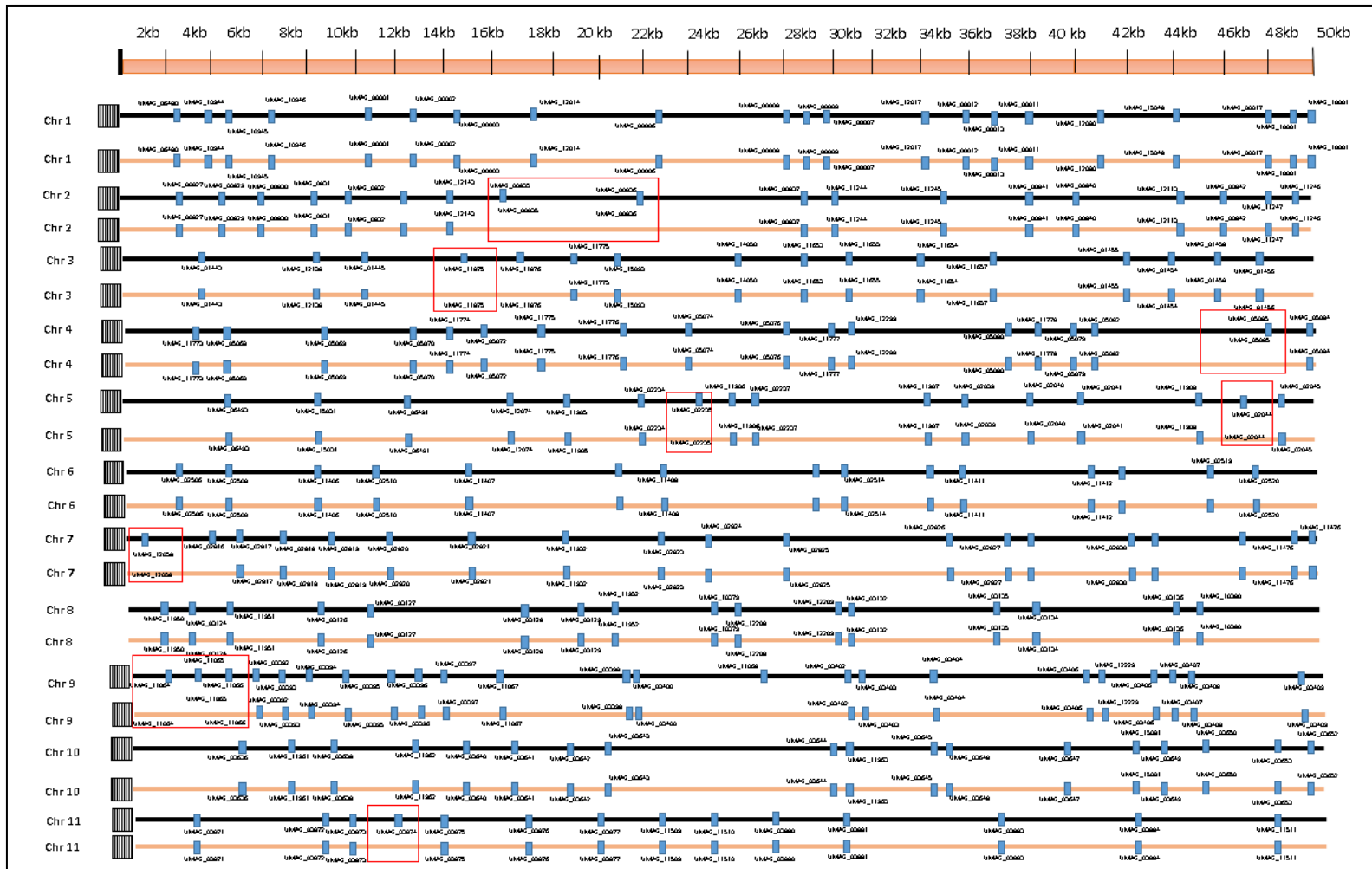


Figura 13.- Representación de genes que se encuentran en los 50 kb cercanos al extremo telomérico izquierdo en las cepas *U. maydis* 521 (línea negra) y *U. maydis* PGA 2.1 (línea rosada) en los cromosomas 1- 11, los recuadros rojos en los extremos de los cromosomas muestran donde se presenta la ausencia del gen.

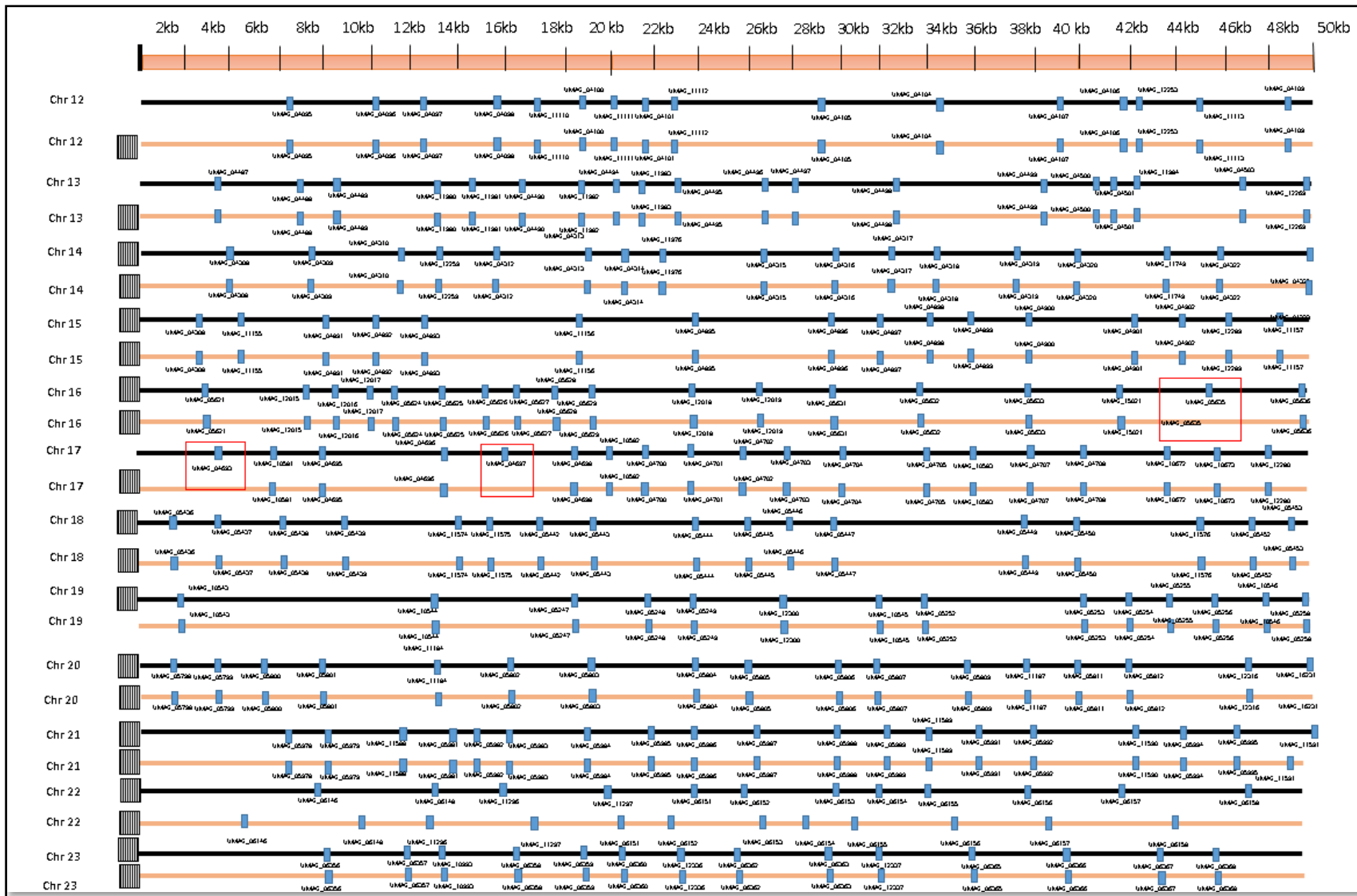


Figura 14.- Representación de genes que se encuentran en los 50 kb cercanas al extremo telomérico izquierdo en las cepas *U. maydis* 521 (línea negra) y *U. maydis* PGA 2.1 (línea rosada) en los cromosomas 12-23, los recuadros rojos en los extremos de los cromosomas muestran donde se presenta la ausencia del gen.

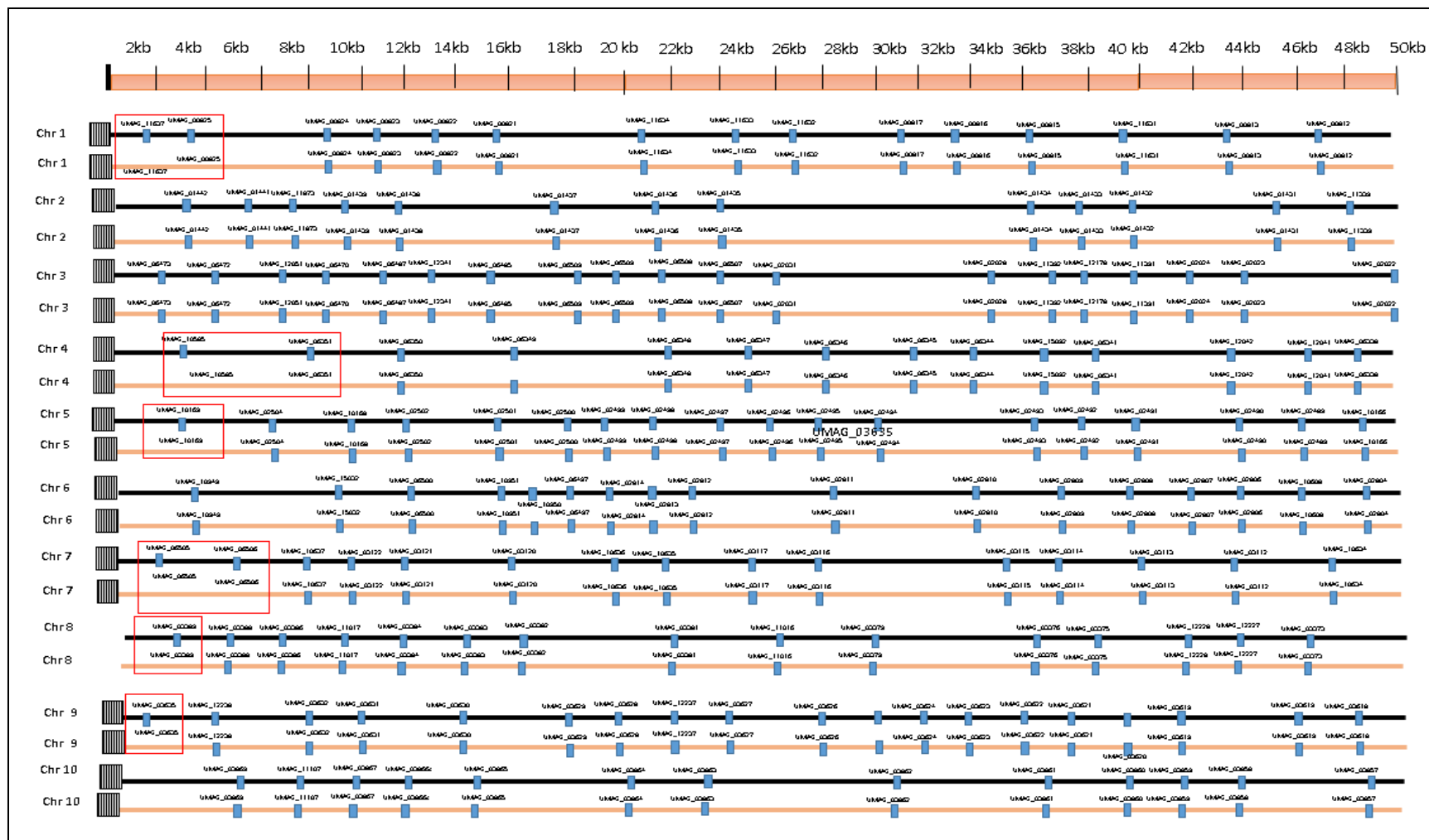


Figura 15.- Representación de genes que se encuentran en los 50 kb cercanos al extremo telomérico derecho en las cepas *U. maydis* 521 (línea negra) y *U. maydis* PGA 2.1 (línea rosada) en los cromosomas 1- 10, los recuadros rojos en los extremos de los cromosomas muestran donde se presenta la ausencia del gen.

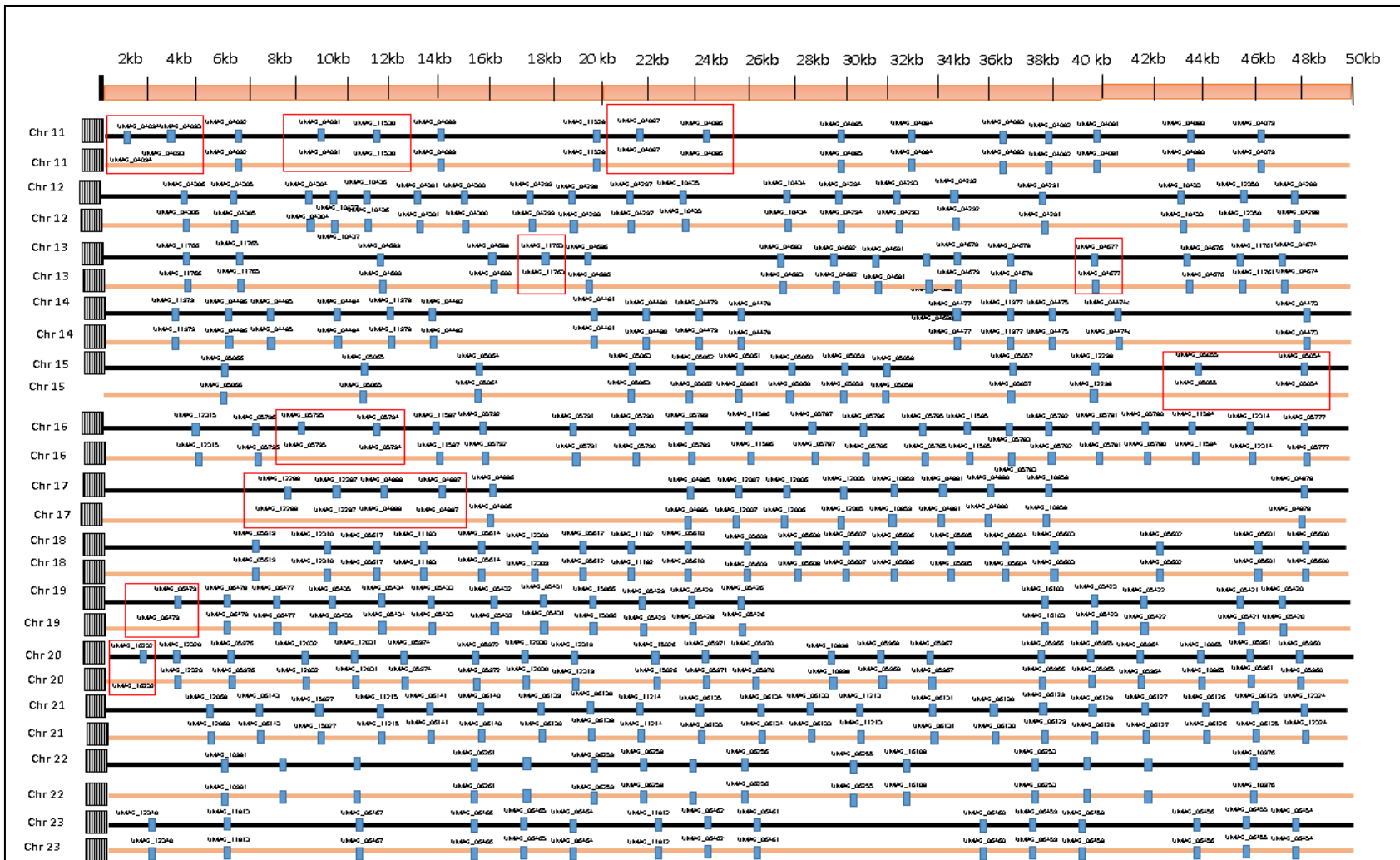


Figura 16.- Representación de genes que se encuentran en los 50 kb cercanas al extremo telomérico derecho en las cepas *U. maydis* 521 (línea negra) y *U. maydis* PGA 2.1 (línea rosada) en los cromosomas 11- 23, los recuadros rojos en los extremos de los cromosomas muestran donde se presenta la ausencia del gen.

En las figuras mostradas anteriormente (13,14,15 y 16) se pueden observar los genes presentes cercanos a la región telomérica, estos genes no forman parte de los genes esenciales para U. maydis, la mayoría de estos se encuentran presentes en ambas cepas (521 y PGA2.1), las regiones con recuadros rojos son genes que no se encuentran en la cepa PGA2.1.

En la tabla 4 se muestran los genes de la cepa 521 que no se encuentran en los cromosomas (indicados a la izquierda) de la cepa PGA2.1, así como la proteína que sintetiza dicho gen y la relación con alguna proteína del banco de datos.

Chr	Locus	Descripción
1	UMAG_00825	Proteína no caracterizada
1	UMAG_11637	Proteína hipotética
2	UMAG_00835	Proteína no caracterizada. DEXDc; DEAD-like helicases superfamily
2	UMAG_00836	Proteína no caracterizada. Mak10; Mak10 subunit, NatC N(alpha)-terminal acetyltransferase
3	UMAG_11876	Proteína no caracterizada. Dynactin_p62; Dynactin p62 family
4	UMAG_06351	Proteína no caracterizada
4	UMAG_10585	Proteína no caracterizada
4	UMAG_05085	E3 ubiquitin-protein ligase RAD18
5	UMAG_10169	Proteína no caracterizada. BRCT domain-containing protein 1
5	UMAG_02035	Proteína no caracterizada. $\alpha\beta$ -hydrolase_1; alpha/beta hydrolase fold
5	UMAG_02044	Putative methionyl-tRNA synthetase
7	UMAG_03121	Proteína no caracterizada. LamG; Laminin G domain; Laminin G-like domains are usually Ca ⁺⁺ mediated receptors that can have binding sites for steroids, beta1 integrins, heparin, sulfatides, fibulin-1, and alpha-dystroglycans.
7	UMAG_06505	Proteína no caracterizada
7	UMAG_06506	Proteína no caracterizada: Sitios de unión C2H2 dedos de zinc
7	UMAG_12058	Proteína no caracterizada
7	UMAG_02816	Proteína no caracterizada
8	UMAG_03389	Proteína no caracterizada
9	UMAG_03635	Proteína no caracterizada
9	UMAG_11064	Proteína no caracterizada. GIY-YIG_SF; GIY-YIG nuclease domain superfamily
9	UMAG_11065	Proteína no caracterizada: RecQ-like Helicase

9	UMAG_11066	Proteína no caracterizada. Similar a acyl-CoA oxidase
9	UMAG_11109	Proteína no caracterizada, conservada
9	UMAG_11068	Proteína no caracterizada. tynA; tyramine oxidase; Provisional
11	UMAG_04086	Proteína no caracterizada* pequeña
11	UMAG_04087	Proteína no caracterizada pequeña
11	UMAG_11528	Proteína no caracterizada. Pyridine nucleotide-disulphide oxidoreductase
11	UMAG_04089	Proteína no caracterizada. HNH_2; HNH endonuclease
11	UMAG_11530	Proteína no caracterizada
11	UMAG_04091	Proteína no caracterizada. RT_like_1; RT_like_1: A subfamily of reverse transcriptases (RTs).
11	UMAG_04093	Hypothetical protein
11	UMAG_04094	Proteína no caracterizada RecQ-Like OrsD domain
11	UMAG_03874	Proteína no caracterizada. Leucine Rich repeats (2 copies) interacción Prot-Prot ¿Adenilato ciclasa?
13	UMAG_11763	Proteína no caracterizada. $\alpha\beta$ -hydrolase; alpha/beta hydrolases
13	UMAG_04677	Proteína no caracterizada. Shugoshin_C; Shugoshin C terminus. Centromere protection
15	UMAG_05054	Proteína no caracterizada. Forkhead; Forkhead domain
15	UMAG_05055	Proteína no caracterizada. PHA03247; large tegument protein UL36; Provisional
16	UMAG_05794	Putative ATP-binding cassette glutathione S-conjugate transporter MRP_assoc_pro; multi drug resistance-associated protein (MRP)
16	UMAG_05795	Proteína no caracterizada. Hipotético major facilitator superfamily (MFS)
16	UMAG_05635	Proteína no caracterizada. PHA03307; transcriptional regulator ICP4; Provisional
17	UMAG_04887	Proteína no caracterizada
17	UMAG_04888	Proteína no caracterizada pequeña
17	UMAG_12287	Proteína no caracterizada. Retrotransposon_gag_2; gag-polyptide of LTR copia-type

17	UMAG_12288	RVT_2; Reverse transcriptase (RNA-dependent DNA polymerase)
17	UMAG_04693	Proteína no caracterizada similar a UMAG_04887
17	UMAG_04697	Proteína no caracterizada. WCAG; Nucleoside-diphosphate-sugar epimerase [Cell wall/membrane/envelope biogenesis]
19	UMAG_06479	Proteína no caracterizada similar a UMAG_04887 BRCT DOMAIN-CONTAINING
20	UMAG_16231	5S ribosomal RNA
20	UMAG_16232	5S ribosomal RNA
20	UMAG_05977	Proteína de función desconocida. Sitio de unión dedos Zinc C2H2
22	UMAG_16108	tRNA-Ala
22	UMAG_06256	Proteína no caracterizada. Fungal_TF_MHR; fungal transcription factor regulatory middle homology region. GAL4; GAL4-like Zn(II)2Cys6 (or C6 zinc) binuclear cluster DNA-binding domain
22	UMAG_06156	putative DNA-directed RNA polymerase III core subunit RPO31

Tabla 4. -Genes que no se encontraron en cromosomas de la cepa *U. maydis* PGA2.1 en el catálogo de genes de los 50 kb cercanos a las regiones teloméricas. En rojo se indican proteínas importantes para el funcionamiento óptimo de la célula y que requieren buscarse mediante técnicas experimentales.

En la tabla 4 se enlistan los genes que no se encontraron al final (50 kb) de los cromosomas de la cepa PGA2.1 respecto a la cepa 521; al menos en 17 cromosomas de los 23 cromosomas se encuentra ausente algún gen en la región analizada, dando así una ausencia de 52 genes en la cepa PGA2.1 cuyas funciones no son esenciales en otros organismos donde se encontró una similitud con estos.

Previamente localizadas las regiones teloméricas se procedió al análisis de estas en cada ensamble de las cepas. Utilizando el programa SnapGene se identificaron regiones en los extremos cromosómicos de *U. maydis* que presentaron repetidos teloméricos (TTAGGG). Estas secuencias se encontraron en las tres cepas analizadas.

Se determinó la longitud de las secuencias teloméricas y se observó una variabilidad en la longitud de estas secuencias entre diferentes cromosomas y dentro de la misma población de *U. maydis*. Algunas secuencias teloméricas eran más cortas en comparación que otras. Con base en la medición de la longitud de los telómeros se identificó el tamaño promedio de los

telómeros en las cepas analizadas. En las figuras 17, 18 y 19 se encuentran los gráficos de los valores de tamaño de cromosomas reportados, así como la variabilidad de la longitud en los telómeros de las cepas 521, PGA2.1, PGAttr1-4.

En la figura 17 se puede observar la longitud promedio (717 pb) de los telómeros de la cepa 521, encontrando el telómero de menor (402 pb) en el cromosoma 14 y el de mayor longitud (1142 pb) en el cromosoma 19.

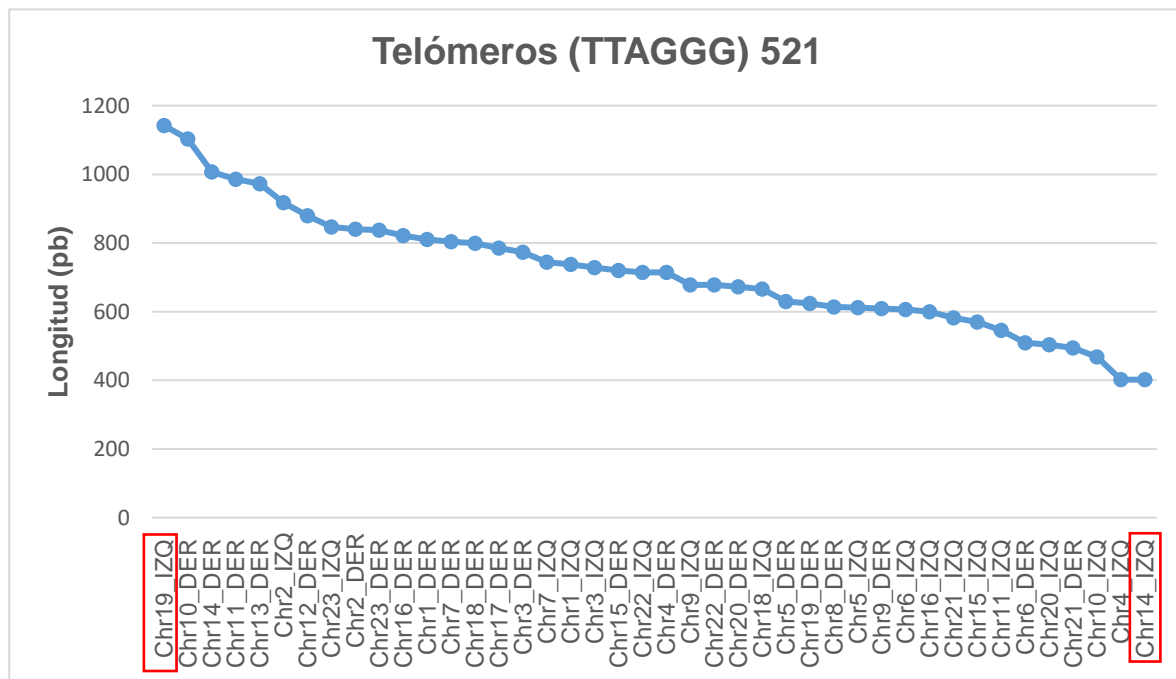


Figura 17. - Gráfica Longitud de telómeros de cepa *U. maydis* 521 en base a la longitud telomérica.

En la figura 18 se puede observar la longitud promedio (872 pb) de los telómeros de la cepa PGA2.1, encontrando el telómero de menor longitud (474 pb) y mayor longitud (1590 pb) ambos identificados en el cromosoma 20.

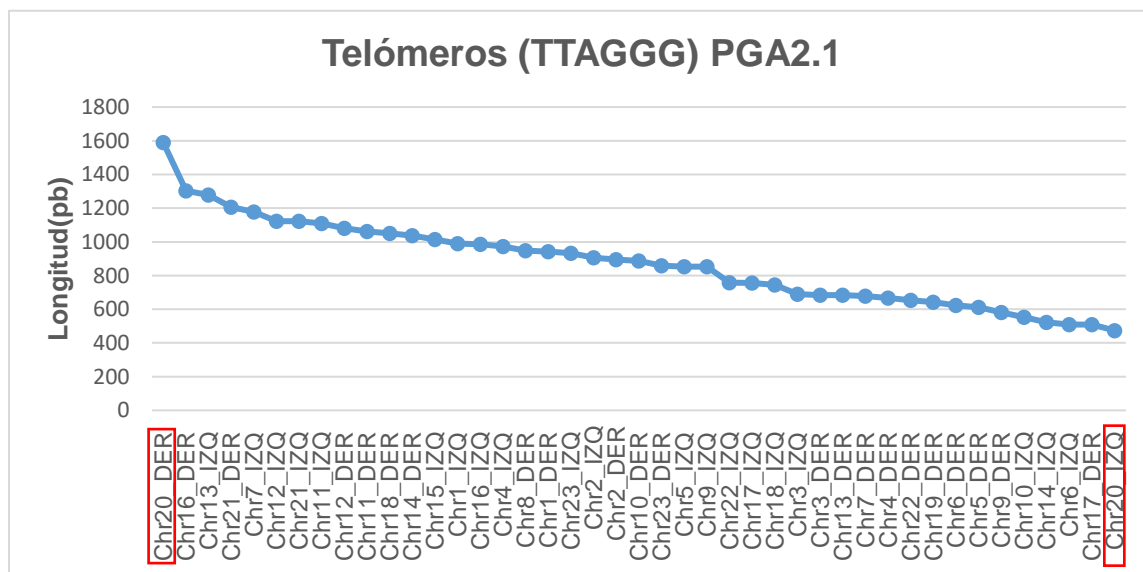


Figura 18. - Gráfica Longitud de telómeros de cepa *U. maydis* PGA2.1 en base a la longitud telomérica.

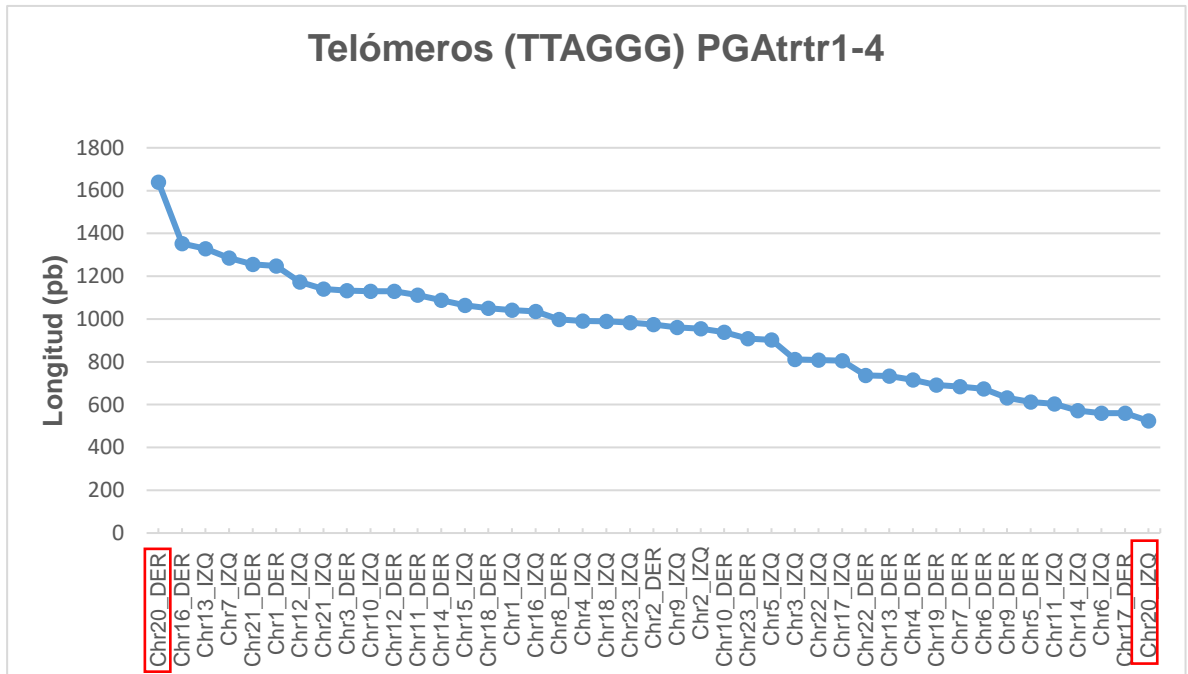


Figura 19. - Gráfica de la longitud de telómeros de cepa *U. maydis* PGAttr1-4.

En la figura 19 se muestra la distribución de la longitud de los telómeros en los que el cromosoma 20 presenta un extremo con la longitud más grande (1640 pb) y el otro con la de menor longitud (524 pb), para los telómeros de la cepa PGAttr1-4.

De las figuras anteriores se deduce que la cepa que presenta mayor longitud promedio de los telómeros es la cepa telomerasa negativa (PGAttr1-4), y la que presenta telómeros de menor longitud es la cepa 521. La cepa PGA2.1 así como la telomerasa negativa presentaron en el mismo cromosoma 20 el telómero de menor y mayor longitud.

VIII. 3.1 Identificación de secuencias asociadas al telómero *UTASa*, *UTASb* y *rumT*, así como otras secuencias en el genoma.

Se identificaron las secuencias con similitud a *UTASa* y *UTASb* previamente reportadas (Sánchez-Alonso y Guzmán, 1998) así también las del borde del repetido telomérico *rumT* secuencias que se encuentran en la región subtelomérica en los cromosomas de los genomas ensamblados de las cepas 521 y PGA2.1.

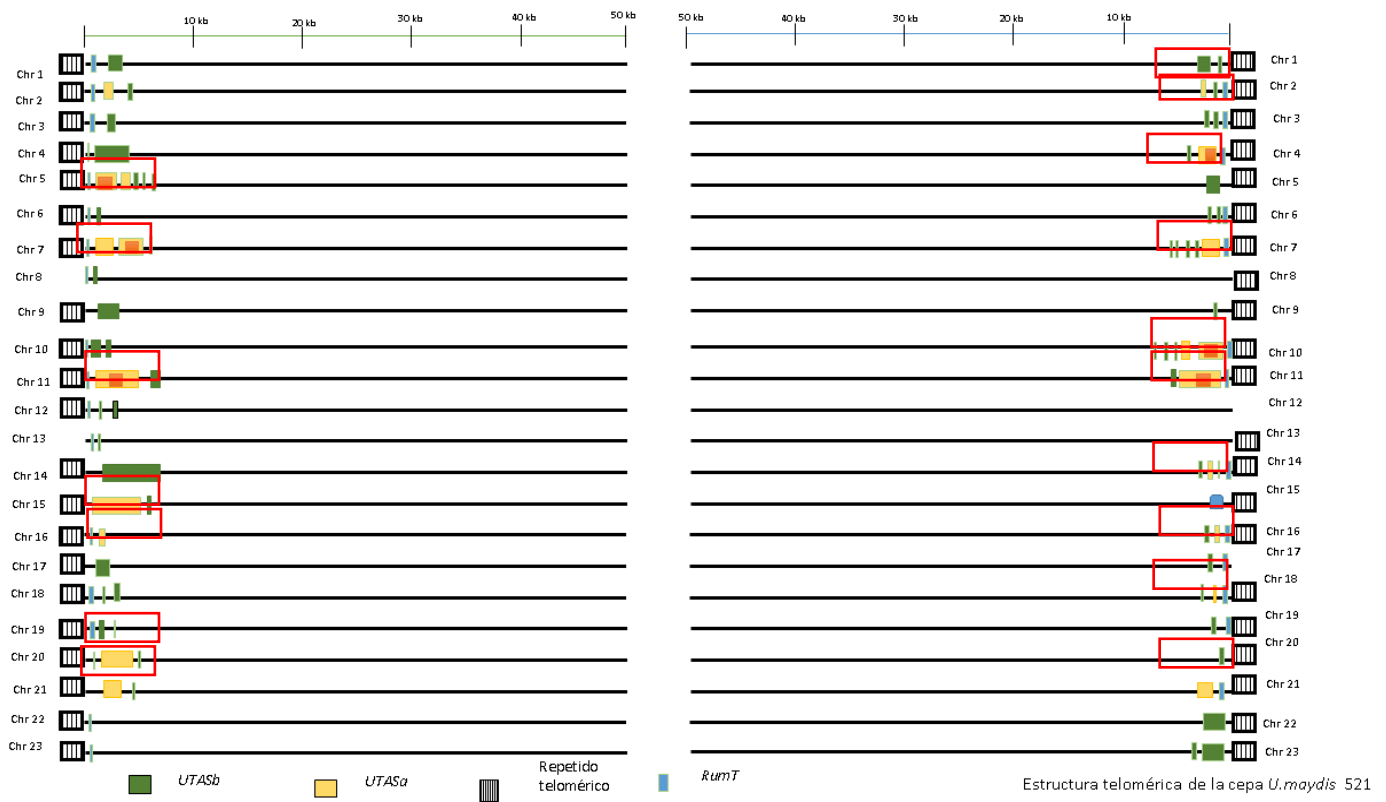


Figura 20.- Organización de secuencias en el extremo cromosomal de *U. maydis* 521.

Los recuadros de color rojo son regiones subtelerómicas en donde *rumT*, *UTASa* o *UTASb* se encuentran ausentes.

En la figura 20 se representa el diagrama de las secuencias subtelerómicas *UTASa* y *UTASb* de la cepa 521; se muestra que no en todos los extremos cromosomales se encontraron elementos *UTASb* mediante el análisis bioinformático. Previamente, mediante el uso de ensayos de Southern blot del cariotipo electroforético se encontró que todos los cromosomas tuvieron señal de hibridación positiva. De igual manera, la secuencia *rumT* que se encontró experimentalmente adyacente al repetido telomérico en la organización propuesta previamente, no se encontró en todos los cromosomas ensamblados, resultado que se discutirá adelante.

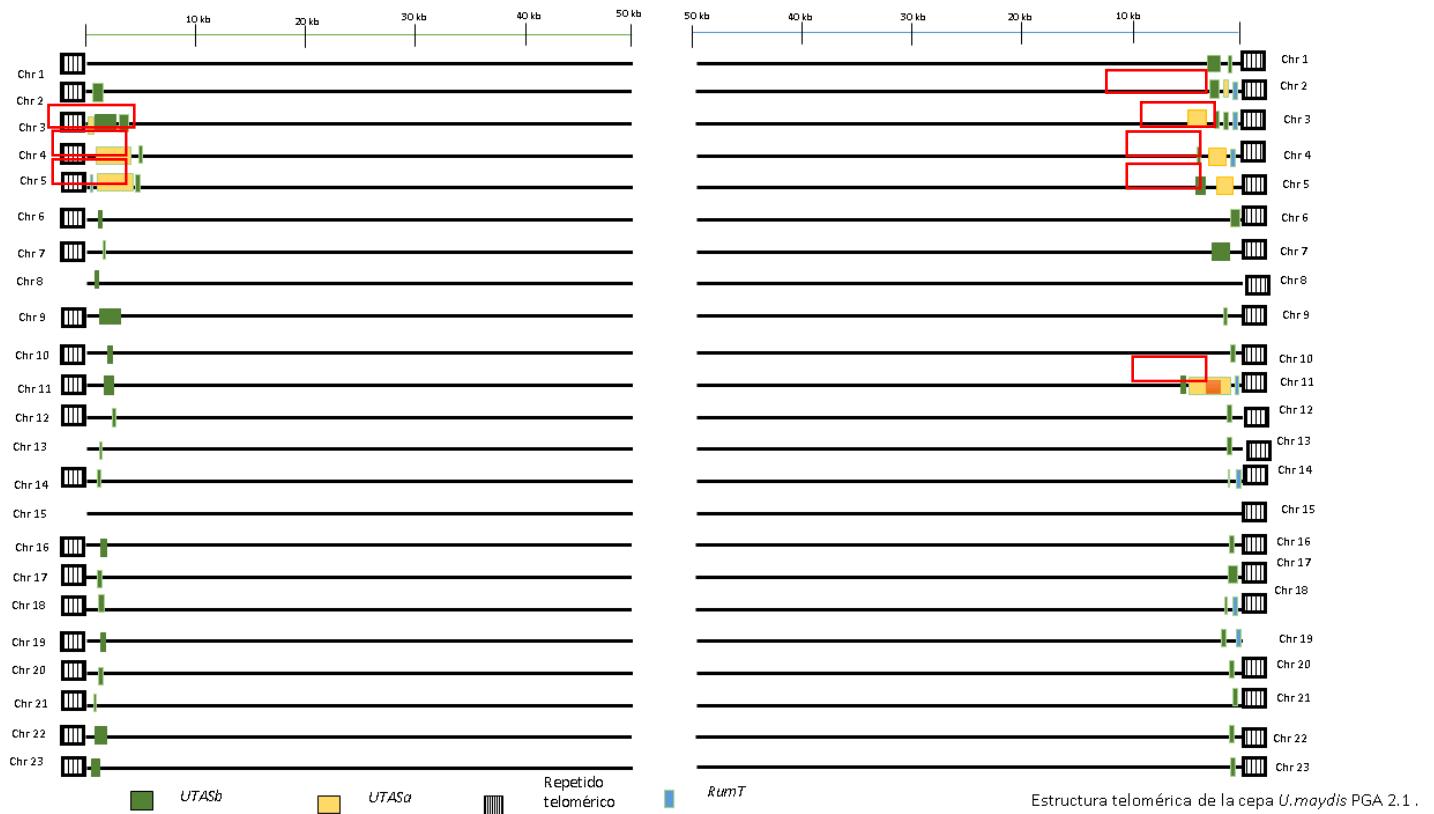


Figura 21- Organización de secuencias en el extremo cromosomal de *U. maydis* PGA2.1.

Los recuadros de color rojo son regiones subteloméricas en donde se encuentra la ausencia de *rumT*, *UTASa* o *UTASb*.

En la figura 21 se muestra la organización propuesta para los extremos cromosomales de la cepa PGA2.1, la cual es de interés, por tener una cantidad baja de copias *UTASa* en sus cromosomas, y la utilidad que puede tener en la caracterización de eventos recombinacionales que ocurren en ausencia de la enzima telomerasa.

Al analizar el número de copias de las secuencias *UTASa* y *UTASb* en la región subtelomérica de los cromosomas de la cepa PGA2.1, al igual que en la cepa 521 se presentaron algunas discrepancias en el número de elementos en el ensamble: la cantidad de copias *UTASa* en el ensamble se encuentra una cantidad mayor de la que se encuentra mediante ensayos semicuantitativos de PCR. Las secuencias *UTASb* que están reportadas en trabajos previos como abundantes los extremos cromosomales no se detecta, de esta manera en todos los extremos de los cromosomas. Otra secuencia analizada en este ensamble fue *rumT* una secuencia que se encuentra adyacente al repetido telomérico en todos los extremos cromosomales de la cepa 521 y abundante en ensayos de Southern blot de digeridos de DNA total de la cepa no se encontró distribuida de esa manera, ya que solo se identificó en algunos cromosomas.

Con base a lo encontrado en las figuras 20 y 21 los resultados difieren en cuanto a la cantidad de secuencias presentes en la región subtelomérica, resultados que se discuten adelante.

IX. Discusión

En este estudio, se llevó a cabo el análisis de calidad de las secuencias del genoma de tres cepas de *U. maydis* obtenidas mediante una plataforma de secuenciación de tercera generación. Asimismo, se llevó a cabo el ensamblaje del genoma y un análisis bioinformático de las secuencias de DNA que estructuran los extremos cromosomales en *Ustilago maydis* cepas 521, PGA 2.1 y PGA2.1trt1 para asegurar la calidad de los ensamblajes se realizaron algunos análisis con programas como Mauve que en las figuras de los alineamientos se muestra la precisión y similitud estructural, mediante este análisis visual y colinealidad de genomas se respaldan la calidad de los mismos. En este trabajo se logró proponer un patrón estructural de las secuencias *UTASa*, *UTASb* y *rumT*, estos siendo identificados en 23 contigs resultantes de los ensamblajes siendo estos contigs asociados con los cromosomas de la cepa caracterizada y dando un valor de calidad al ensamblaje. Sin embargo, el número de repetidos de las secuencias subteloméricas no corresponde al previamente determinado (Sánchez-Alonso & Guzmán, 1998). Esto se debe a que *UTASa*, es una secuencia muy conservada, con al menos un 90% de identidad entre las secuencias a la fecha encontradas, y en *UTASb* existe un centro o core también conservado (Sánchez-Alonso & Guzmán, 1998). Las copias de este elemento se encuentran en forma de duplicaciones segmentales, en un número hasta ahora poco conocido, y durante el ensamblaje los programas utilizados pudieron haber sobrelapado las diferentes copias; por lo que la estructura predicha tiene que verificarse experimentalmente.

Aunque en algunos grupos de trabajo de interés en regiones teloméricas, subteloméricas y centroméricas proponen el uso de múltiples plataformas de secuenciación para resolver problemas de sobrelapamiento entre las lecturas al ensamblarse esto para tratar de resolver estas regiones de los extremos cromosomales, se requiere del etiquetamiento de los extremos cromosomales con genes o secuencias heterólogas en diferentes cepas para lograr discernir la estructura de la región subtelomérica. Otros intentos realizados por el grupo de trabajo, y no mostrados aquí se basaron en la cuantificación del número de elementos *UTASa*, *UTASb*, y *rumT* mediante qPCR, y los datos indicaron alrededor de 4 copias del elemento por cromosoma, aunque su distribución no puede discernirse por esta técnica. La hibridación del *UTASa* en el cariotipo electroforético de la cepa 521 muestra la posible carencia del elemento en un par de cromosomas, pero no puede predecirse en número de elementos por cromosoma, aun utilizando técnicas de densitometría de la autorradiografía (Harley et al., 1990) o mediante FISH (Hernández-Ibarra et al., 2008). Posiblemente la extensión con polimerasas de muy alta fidelidad pudiera completar los mapas aquí propuestos

Otro punto desfavorable fue la imposibilidad de encontrar el extremo cromosomal de algunos cromosomas como el 8,9,15. Pese a esto, logramos ensamblar más del 90% de los repetidos del genoma de PGA2.1, también y llevar a cabo un análisis comparativo respecto a la longitud telomérica. Aunque a primera vista puede parecer lógico encontrar diferencias en la longitud promedio de los extremos cromosomales de esta manera, la preparación del DNA de alto peso molecular implica la extracción de DNA total de más de 10^7 células. La literatura predice un alargamiento y acortamiento dinámico de los telómeros a lo largo de las consecutivas rondas de replicación de cada célula (Hug and Lingner, 2006). Si no hubiera un control, o perpetuación

de un tamaño promedio posiblemente el registro del tamaño hubiera sido igual para todos los extremos cromosomales. Curiosamente existe un tamaño promedio para los diferentes repetidos teloméricos de los cromosomas de las cepas estudiadas, pero más sorprendente es el hallazgo de una relación de perpetuación entre la cepa PGA2.1 y la cepa PGA2.1trt1, la cual carece de telomerasa y mantiene la longitud del extremo cromosomal mediante mecanismos propensos al error basados en recombinación; esto ha sido reportado previamente para humanos (Graakjaer *et al.*, 2003). Posiblemente la medición de la longitud promedio de los telómeros de cada cepa mediante técnicas de PCR pudieran corroborar los hallazgos de la longitud promedio del telómero de PGA2.1 (872 pb) en comparación con la cepa 521(717 pb) y la longitud de la cepa telomerasa negativa PGA2.1trt1 la cual podría ser aún mayor de los > 900 pb aquí reportados (Cawthon, 2002).

En cuanto a los genes albergados en las 50 kb de DNA adyacente de la región telomérica, se encontró que la cepa PGA2.1 carece de 52 genes que se encuentra en la región subteloamérica de la cepa 521. Algunos de estos genes, aunque no parecen ser esenciales, tienen importancia en el metabolismo óptimo de las células, y dado que se presenta una serie de translocaciones entre ambas cepas, es posible que estos se encuentren localizados en otras partes del genoma, Sería importante hacer la búsqueda mediante el alineamiento de secuencias de ambas cepas y experimentalmente ya sea por amplificación de secuencias o mediante análisis Southern blot.

Otro hallazgo importante fueron las translocaciones que ocurre en el genoma de la cepa telomerasa negativa, hasta ahora se ha pensado que la propensión a eventos recombinacionales para el mantenimiento y sanamiento de los telómeros se limitaría a esta región. Sin embargo, los mapas de MAUVE muestran una cantidad de translocaciones entre las cepas PGA2.1 y su derivada PGA2.1trt1 que es mayor que la de 521 respecto a PGA2.1. Esto correlaciona con la cantidad de barridos en el cariotipo electroforético de las cepas telomerasa negativas estudiadas en nuestro laboratorio. Posiblemente una serie de entrecruzamientos que se originan de las vías alternativas de mantenimiento del extremo cromosomal propendan a la fusión de cromosomas y a la generación de ciclos fusión-puente-ruptura que afectan la integridad del tamaño de los cromosomas.

Futuros trabajos sobre el tema pudieran ahondar en los mecanismos que generan estas diferencias en el extremo cromosomal de cepas de la misma especie.

X. Conclusiones

- Se logró la secuenciación de los 23 cromosomas que contienen las cepas de *Ustilago maydis* 521 (a1, b1), PGA2.1 (a2, bx) y la derivada PGAttr1-4 521, con una alta resolución.
- Se logró la secuenciación de la mayoría de los extremos cromosomales, llegando a observar el repetido telomérico "TTAGGG" en un rango de presencia entre 1700 y 400X. Siendo más cortas las secuencias con repetidos en la cepa 521 y más largas en la mutante PGAttr1-4 521.
- El análisis de 50 kb de los extremos cromosomales mostró un interesante mosaicismo génico que sugiere frecuentes procesos de recombinación homóloga con reemplazos de regiones específicas.
- El análisis de localización de secuencias *UTASa*, *UTASb*, *rumT* mostró una variada presencia de estos elementos en los extremos cromosomales que hoy se ven resueltos con mucha precisión, pero que deben ser valorados con más estudios para conocer su grado de significancia funcional.
- En el 17% de los extremos cromosomales de PGA2.1 se observó la presencia de *UTASa* ligado a *UTASb*, mientras que el 91% de extremos cromosomales contiene sólo *UTASb*.
- En el 36% de los extremos cromosomales de 521 se observó la presencia de *UTASa* ligado a *UTASb*, mientras que el 95% de extremos cromosomales contiene sólo *UTASb*.
- En el 9% de los extremos cromosomales de PGA2.1 y en el 6% de los extremos cromosomales de 521 se observó una aparente erosión de los extremos cromosomales, cuya relevancia en el rearrreglo de los extremos cromosomales debe ser investigada.

XI. Perspectivas

- Complementación de análisis del banco de datos en algunas regiones teloméricas y subteloméricas de cepas (*Ustilago maydis* 521 (a1, b1), PGA2.1 (a2, bx) y la derivada PGAttr1-4 para reducir algunos faltantes de estas.
- Análisis de la cepa derivada PGAttr1-4 521 respecto a los genes cercanos a los extremos cromosomales.
- Refinar el análisis de los extremos cromosomales específicos por medio de técnicas de biología molecular como PFGE, Southern blot, PCR y qPCR para corroborar y complementar los resultados de la secuenciación.

XII. Referencias

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215(3), 403-410.
- Andrews, S. (2010). FastQC: una herramienta de control de calidad para datos de secuencia de alto rendimiento. In.
- Azzalin, C. M., & Lingner, J. (2015). Telomere functions grounding on TERRA firma. *Trends in cell biology*, 25(1), 29-36.
- Bah, A., Wischnewski, H., Shchepachev, V., & Azzalin, C. M. (2012). The telomeric transcriptome of *Schizosaccharomyces pombe*. *Nucleic acids research*, 40(7), 2995-3005.
- Barry, J., Ginger, M. L., Burton, P., & McCulloch, R. (2003). Why are parasite contingency genes often associated with telomeres? *International journal for parasitology*, 33(1), 29-45.
- Biessmann, H., Mason, J. M., Ferry, K., d'Hulst, M., Valgeirsdottir, K., Traverse, K. L., & Pardue, M.-L. (1990). Addition of telomere-associated HeT DNA sequences "heals" broken chromosome ends in *Drosophila*. *Cell*, 61(4), 663-673.
- Blackburn, E. H. (1994). Telomeres: no end in sight. *Cell*, 77(5), 621-623. [https://doi.org/10.1016/0092-8674\(94\)90046-9](https://doi.org/10.1016/0092-8674(94)90046-9)
- Brachmann, A., Weinzierl, G., Kämper, J., & Kahmann, R. (2001). Identification of genes in the bW/bE regulatory cascade in *Ustilago maydis*. *Molecular microbiology*, 42(4), 1047-1063.
- Brazda, V., Haronikova, L., Liao, J. C., & Fojta, M. (2014). DNA and RNA quadruplex-binding proteins. *Int J Mol Sci*, 15(10), 17493-17517. doi:10.3390/ijms151017493.
- Cawthon, R. M. (2002). Telomere measurement by quantitative PCR. *Nucleic Acids Res*, 30(10), e47. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12000852>
- Chan, C. S., Tye, B.-K., & Herskowitz, I. (1983). A family of *Saccharomyces cerevisiae* repetitive autonomously replicating sequences that have very similar genomic environments. *Journal of molecular biology*, 168(3), 505-523.
- d'Adda di Fagagna, F. d. A., Teo, S.-H., & Jackson, S. P. (2004). Functional links between telomeres and proteins of the DNA-damage response. *Genes & development*, 18(15), 1781-1799.
- Darling, A. C., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome research*, 14(7), 1394-1403.
- de Bruin, D., Kantrow, S. M., Liberatore, R. A., & Zakian, V. A. (2000). Telomere folding is required for the stable maintenance of telomere position effects in yeast. *Molecular and cellular biology*, 20(21), 7991-8000.

- de Lange, T. (2005). Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes & development*, 19(18), 2100-2110.
- de Lange, T. (2018). Shelterin-mediated telomere protection. *Annual review of genetics*, 52, 223-247.
- Deitsch, K. W., Moxon, E. R., & Wellems, T. E. (1997). Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections. *Microbiology and Molecular Biology Reviews*, 61(3), 281.
- Denayrolles, M., de Villechenon, E. P., Lonvaud-Funel, A., & Aigle, M. (1997). Incidence of SUC-RTM telomeric repeated genes in brewing and wild wine strains of *Saccharomyces*. *Current genetics*, 31(6), 457-461.
- Flint, J., Bates, G., Clark, K., Dorman, A., Willingham, D., Roe, B., . . . Louis, E. (1997). Sequence comparison of human and yeast telomeres identifies structurally distinct subtelomeric domains. *Human molecular genetics*, 6(8), 1305-1314.
- Graakjaer, J., Bischoff, C., Korsholm, L., Holstebro, S., Vach, W., Bohr, V. A., . . . Kølvrå, S. (2003). The pattern of chromosome-specific variations in telomere length in humans is determined by inherited, telomere-near factors and is maintained throughout life. *Mech Ageing Dev*, 124(5), 629-640. doi:10.1016/s0047-6374(03)00081-2
- Greenwood, J., & Cooper, J. P. (2012). Non-coding telomeric and subtelomeric transcripts are differentially regulated by telomeric and heterochromatin assembly factors in fission yeast. *Nucleic acids research*, 40(7), 2956-2963.
- Greider, C. W. (1991). Telomeres. *Curr Opin Cell Biol*, 3(3), 444-451. [https://doi.org/10.1016/0955-0674\(91\)90072-7](https://doi.org/10.1016/0955-0674(91)90072-7)
- Grenon, M., Gilbert, C., & Lowndes, N. F. (2001). Checkpoint activation in response to double-strand breaks requires the Mre11/Rad50/Xrs2 complex. *Nature cell biology*, 3(9), 844-847.
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29(8), 1072-1075.
- Guzmán, P. A., & Sanchez, J. G. (1994). Characterization of telomeric regions from *Ustilago maydis*. *Microbiology*, 140(3), 551-557.
- Hammond, P. W., & Cech, T. R. (1998). Euplotes telomerase: evidence for limited base-pairing during primer elongation and dGTP as an effector of translocation. *Biochemistry*, 37(15), 5162-5172. <https://doi.org/10.1021/bi972988o>
- Harley, C. B., Futcher, A. B., & Greider, C. W. (1990). Telomeres shorten during ageing of human fibroblasts. *Nature*, 345(6274), 458-460.
- Hartmann, H. A., Krüger, J., Lottspeich, F., & Kahmann, R. (1999). Environmental signals controlling sexual development of the corn smut fungus *Ustilago maydis* through the transcriptional regulator Prf1. *The Plant Cell*, 11(7), 1293-1305.

- Henderson, I. R., Owen, P., & Nataro, J. P. (1999). Molecular switches—the ON and OFF of bacterial phase variation. *Molecular microbiology*, 33(5), 919-932.
- Hernandez-Ibarra, N. K., Leitch, A. R., Cruz, P., & Ibarra, A. M. (2008). Fluorescent in situ hybridization and characterization of the Sall family of satellite repeats in the *Haliotis L.* species (abalone) of the Northeast Pacific. *Genome*, 51(8), 570-579. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18650947
- Hug, N., & Lingner, J. (2006). Telomere length homeostasis. *Chromosoma*, 115(6), 413-425. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16741708
- Jansson, L. I., Hentschel, J., Parks, J. W., Chang, T. R., Lu, C., Baral, R., . . . Stone, M. D. (2019). Telomere DNA G-quadruplex folding within actively extending human telomerase. *Proceedings of the National Academy of Sciences*, 116(19), 9350-9359.
- Kelleher, C., Teixeira, M. T., Förstemann, K., & Lingner, J. (2002). Telomerase: biochemical considerations for enzyme and substrate. *Trends in biochemical sciences*, 27(11), 572-579.
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome research*, 27(5), 722-736.
- Kwapisz, M., & Morillon, A. (2020). Subtelomeric transcription and its regulation. *Journal of molecular biology*, 432(15), 4199-4219.
- Kämper, J., Kahmann, R., Bölker, M., Ma, L. J., Brefort, T., Saville, B. J., . . . Birren, B. W. (2006). Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature*, 444(7115), 97-101. <https://doi.org/10.1038/nature05248>
- König, P., Giraldo, R., Chapman, L., & Rhodes, D. (1996). The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA. *Cell*, 85(1), 125-136.
- Louis, E. J., & Haber, J. E. (1992). The structure and evolution of subtelomeric Y' repeats in *Saccharomyces cerevisiae*. *Genetics*, 131(3), 559-574.
- Louis, E., Naumova, E., Lee, A., Naumov, G., & Haber, J. (1994). The chromosome end in yeast: its mosaic nature and influence on recombinational dynamics. *Genetics*, 136(3), 789-802.
- Louis, E. J. (1995). The chromosome ends of *Saccharomyces cerevisiae*. *Yeast*, 11(16), 1553-1573. <https://doi.org/10.1002/yea.320111604>
- Lundblad, V., & Blackburn, E. H. (1993). An alternative pathway for yeast telomere maintenance rescues est1- senescence. *Cell*, 73(2), 347-360.
- Lundblad, V., & Szostak, J. W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell*, 57(4), 633-643.

- Matei, A., & Doehlemann, G. (2016). Cell biology of corn smut disease—*Ustilago maydis* as a model for biotrophic interactions. *Current Opinion in Microbiology*, *34*, 60-66.
- McEachern, M. J., Krauskopf, A., & Blackburn, E. H. (2000). Telomeres and their control. *Annual review of genetics*, *34*(1), 331-358.
- Mendgen, K., & Hahn, M. (2002). Plant infection and the establishment of fungal biotrophy. *Trends in plant science*, *7*(8), 352-356.
- Nakamura, T. M., Cooper, J. P., & Cech, T. R. (1998). Two modes of survival of fission yeast without telomerase. *Science*, *282*(5388), 493-496.
- Nergadze, S. G., Farnung, B. O., Wischniewski, H., Khorauli, L., Vitelli, V., Chawla, R., . . . Azzalin, C. M. (2009). CpG-island promoters drive transcription of human telomeres. *Rna*, *15*(12), 2186-2194.
- Ness, F., & Aigle, M. (1995). RTM1: a member of a new family of telomeric repeated genes in yeast. *Genetics*, *140*(3), 945-956.
- Olovnikov, A. M. (1972). [The immune response and the process of marginotomy in lymphoid cells]. *Vestn Akad Med Nauk SSSR*, *27*(12), 85-87.
- Okonechnikov, K., Golosova, O., Fursov, M., & Team, U. (2012a). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, *28*(8), 1166-1167.
- Palm, W., & de Lange, T. (2008). How shelterin protects mammalian telomeres. *Annual review of genetics*, *42*, 301-334.
- Peska, V., Fajkus, P., Bubeník, M., Brázda, V., Bohálová, N., Dvořáček, V., . . . Garcia, S. (2021). Extraordinary diversity of telomeres, telomerase RNAs and their template regions in Saccharomycetaceae. *Scientific reports*, *11*(1), 1-11.
- Pryde, F., & Louis, E. (1997). *Saccharomyces cerevisiae* telomeres. A review. *Biochemistry-New York-English Translation of Biokhimiya*, *62*(11), 1232-1241.
- Schrumpfová, P. P., & Fajkus, J. (2020). Composition and function of telomerase—a polymerase associated with the origin of eukaryotes. *Biomolecules*, *10*(10), 1425.
- Sánchez-Alonso, P., & Guzmán, P. (1998). Organization of chromosome ends in *Ustilago maydis*. RecQ-like helicase motifs at telomeric regions. *Genetics*, *148*(3), 1043-1054.
- Sánchez-Alonso, P., Valverde, M. E., Paredes-López, O., & Guzmán, P. (1996). Detection of genetic variation in *Ustilago maydis* strains by probes derived from telomeric sequences. *Microbiology*, *142*(10), 2931-2936.
- Singh, D. K., Ghosh, A. K., Croteau, D. L., & Bohr, V. A. (2012). RecQ helicases in DNA double strand break repair and telomere maintenance. *Mutat Res*, *736*(1-2), 15-24. doi:10.1016/j.mrfmmm.2011.06.002

- Stavenhagen, J. B., & Zakian, V. A. (1994). Internal tracts of telomeric DNA act as silencers in *Saccharomyces cerevisiae*. *Genes Dev*, 8(12), 1411-1422. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7926741>
- Takahama, K., Takada, A., Tada, S., Shimizu, M., Sayama, K., Kurokawa, R., & Oyoshi, T. (2013). Regulation of telomere length by G-quadruplex telomere DNA- and TERRA-binding protein TLS/FUS. *Chem Biol*, 20(3), 341-350. doi:10.1016/j.chembiol.2013.02.013
- Tham, W. H., & Zakian, V. A. (2002). Transcriptional silencing at *Saccharomyces* telomeres: implications for other organisms. *Oncogene*, 21(4), 512-521. <https://doi.org/10.1038/sj.onc.1205078>
- Yu, E. Y., Zahid, S. S., Ganduri, S., Sutherland, J. H., Hsu, M., Holloman, W. K., & Lue, N. F. (2020). Structurally distinct telomere-binding proteins in *Ustilago maydis* execute non-overlapping functions in telomere replication, recombination, and protection. *Commun Biol*, 3(1), 777. <https://doi.org/10.1038/s42003-020-01505-z>
- Zakian, V. A. (1996). Telomere functions: lessons from yeast. *Trends in cell biology*, 6(1), 29-33.
- Zvereva, M., Shcherbakova, D., & Dontsova, O. (2010). Telomerase: structure, functions, and activity regulation. *Biochemistry (Moscow)*, 75(13), 1563-1583.