



**UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO**

ENES Juriquilla

**Descripción de híbridos de
levaduras asociados al agave**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Licenciado en Ciencias Genómicas

P R E S E N T A:

Erick Santiago Garduño



**DIRECTORA DE TESIS:
Lucía Guadalupe Morales Reyes**

Juriquilla, Querétaro, 2024



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

A mis papás

Agradecimientos

Quiero agradecer a mi tutora y directora de tesis, la Dra. Lucía Morales Reyes. Gracias por tu mentoría, paciencia, comprensión y confianza que has tenido conmigo, aprecio todas las conversaciones académicas y no académicas que hemos tenido, gracias por apoyarme e impulsarme en tomar decisiones por mí y para mí. Trabajar contigo estos años has sido de las etapas que más he disfrutado de mi carrera universitaria, recordaré con mucho cariño esta etapa de mi formación.

Agradezco a los miembros de mi jurado conformado por la Dra. Alicia González Manjarrez, la Dra. Lina Raquel Riego Ruiz, la Dra. Cintia Gómez Muñoz y el Dr. Jesús Abraham Avelar Rivas por dedicar el tiempo para revisar mi escrito y discutir mi trabajo, gracias por sus observaciones y preguntas, las cuales enriquecieron mi proyecto. Quiero agradecer particularmente a Alicia por despertar mi interés en las levaduras en aquel seminario que impartió apasionadamente para mi generación en el quinto semestre, y a Abraham por toda la ayuda y atención que me brindó para elaborar este proyecto.

Agradezco el apoyo técnico del Laboratorio Internacional de Investigación sobre el Genoma Humano (LIIGH UNAM) Alejandra Castillo Carbajal, Carina Díaz Uribe y Jair García Sotelo. También agradezco a Luis Alberto Aguilar del Laboratorio Nacional de Visualización Científica Avanzada (LAVIS UNAM) con su apoyo en el uso del equipo de supercómputo.

Este proyecto fue llevado a cabo gracias a los donativos del Consejo Nacional de Humanidades Ciencias y Tecnología (Donativos CB-201-01284992 y CF-103000) y del Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (Donativos IN212524, IA201019 y IN209021, éste último siendo del cual recibí una beca durante mi titulación).

Agradezco a Margareta Boege por toda la atención que nos brinda a los alumnos de la licenciatura, gracias por escucharme y apoyarme cuando lo he necesitado, te agradezco

de corazón. De igual manera, quiero agradecer a los miembros del Laboratorio de Evolución del Genoma: a Maritrini, Mariana, Samuel, Fernanda, Ricardo e Iván; gracias por su compañía y por hacer del laboratorio un ambiente tan agradable y seguro.

Le doy las gracias a mi familia, especialmente a mis papás. Gracias por su amor, por su confianza y por brindarme los recursos para poder mudarme a otro estado y perseguir mis metas; gracias por todo lo que hacen por mí. Gracias a mi hermano Diego, mi compañero de vida.

Finalmente, gracias a mis amigos de toda la vida Giorgio y Mau; a mis amigas de generación Soph y Móni; a mi mejor amiga Tania y a mis queridxs amigxs Sofi, Marlon, Xóch y Victor. Sin ustedes la vida no sabría igual.

Gracias a mi familia y amigas por endulzarla.

Lista de figuras

Figura 1. Mapa genómico del híbrido de Alpechín. Esta figura fue modificada de D'Angiolo et al. (2019). El ADN está coloreado en función a su origen: *S. cerevisiae* (azul) y *S. paradoxus* (rojo). Se pueden observar las regiones de pérdida de heterocigosidad como segmentos del color opuesto intercalados en los subgenomas.

Figura 2. Gráficas de profundidad de cobertura nuclear y mitocondrial del aislado YMX005605. En el eje X se encuentra la referencia concatenada dividida en ventanas no superpuestas de 1 kb, cada color indica el organismo al que pertenece. En el eje Y se muestra la profundidad promedio de cada ventana. A) Gráfica de cobertura de 10 genomas nucleares de levaduras. B) Gráfica de cobertura de 10 genomas mitocondriales de levaduras.

Figura 3. Gráfica de cobertura a referencia mitocondrial concatenada *S. cerevisiae* y *S. paradoxus* del aislado YMX004626. A diferencia de la figura 1, estas gráficas corresponden al mapeo del genoma mitocondrial del híbrido YMX004626 a una referencia concatenada de los genomas mitocondriales únicamente de *S. cerevisiae* y *S. paradoxus*. En este ejemplo, la mitocondria observada es mayoritariamente del parental *S. cerevisiae*.

Figura 4. Gráficas para contrastar el contenido mitocondrial con el nuclear de cada híbrido. Estas gráficas ejemplifican al híbrido YMX506F06 con aportación nuclear de 66% *S. cerevisiae* y 33% *S. paradoxus*; y aportación mitocondrial prevalente de *S. cerevisiae*.

Figura 5. Base de datos de los aislados híbridos. A) Base de datos relacional con información de 64 híbridos *S. cerevisiae* x *S. paradoxus*. En las columnas se encuentran diferentes campos de información como: identificador (ID), sitio de colecta (estado, municipio, coordenadas geográficas, mezcalería), características del híbrido (grupo de híbrido, ploidía, contenido nuclear de *S. cerevisiae*, contenido nuclear de *S. paradoxus*),

etc. B) Vista de una entrada de la base de datos. Al seleccionar el ID de una muestra se despliegan, en una ventana adyacente, los metadatos de esa muestra junto con las gráficas de cobertura nuclear (superior) y mitocondrial (inferior).

Figura 6. Tres tipos de arquitecturas genómicas en los híbridos aislados ejemplificadas por los aislados A) YMX004626, B) YMX182c2 y C) YMX005623. A partir de las gráficas de cobertura se identificaron 3 tipos de arquitecturas genómicas. La diferencia en proporciones se puede apreciar mediante la profundidad promedio de cada sub genoma, como en los casos B) y C) donde uno de los genomas tiene el doble de profundidad que el otro.

Figura 7. Genomas mitocondriales híbridos divididos en tres categorías. Después de contabilizar las lecturas mapeadas a cada referencia mitocondrial y calcular sus proporciones para cada híbrido, clasificamos a las mitocondrias en tres categorías. 1) mayor proporción *S. cerevisiae* ejemplificada por YMX506C10, 2) mayor proporción *S. paradoxus* ejemplificada por YMX004593 y 3) recombinantes ejemplificada por YMX005605.

Figura 8. Origen de las mitocondrias recombinantes. En cada panel, las dos gráficas superiores pertenecen a las mitocondrias de los dos aislados puros más cercanos a los padres de un híbrido dado, en la fila inferior se encuentra la gráfica de la cobertura mitocondrial de dicho híbrido. A) Se observa que el híbrido conserva uno de los patrones parentales, sugiriendo que no hubo recombinación. B) En este caso, se puede observar que el patrón de cobertura mitocondrial presentado por el híbrido no corresponde a ninguno de los patrones parentales, sugiriendo que puede ser recombinante.

Figura 9. El origen del genoma mitocondrial prevalente no siempre coincide con el origen del genoma nuclear prevalente. En esta figura se muestran las proporciones de ADN de *S. cerevisiae* y *S. paradoxus* en el genoma mitocondrial y nuclear de un híbrido representativo para cada grupo. A) El híbrido YMX507A01 presenta genomas

nuclear y mitocondrial mayoritariamente de *S. cerevisiae*. B) El híbrido YMX005624 presenta genoma nuclear *S. cerevisiae* y mitocondrial *S. paradoxus*.

Figura 10. Árboles filogenéticos del híbrido XA020c1. Para cada uno de los híbridos se generaron dos filogenias de máxima verosimilitud: Una incluyendo las 199 cepas *S. cerevisiae* (panel A) y otra con las 84 cepas *S. paradoxus* (panel B). Para el caso del híbrido XA020c1, el número de sitios utilizados para hacer la filogenia *S. cerevisiae* es de 359,053 y la escala en el árbol es de 0.045 sustituciones por sitios. Para la filogenia *S. paradoxus* se utilizaron 190,490 sitios y la escala en el árbol es de 0.052 sustituciones por sitio. En ambos árboles se resalta la posición del híbrido XA020c1 con un nodo de color amarillo.

Figura 11. Densidad de las distancias filogenéticas de cada uno de los híbridos a las cepas más cercanas a los parentales *S. cerevisiae* y *S. paradoxus*. En rojo se muestra la distribución de *S. cerevisiae* y en azul la de *S. paradoxus*, las medianas de las distribuciones se indican con líneas sólidas. Las distancias se encuentran en las columnas 3 y 7 de la Tabla Suplementaria 4.

Figura 12. Gráfica de coordenadas paralelas de las distancias filogenéticas de los híbridos a las cepas más cercanas a los parentales *S. cerevisiae* y *S. paradoxus*.

La primera barra vertical representa la medida de distancia filogenética a la cepa más cercana a los parentales *S. cerevisiae*, la segunda indica la distancia a *S. paradoxus*. Cada recta relaciona las distancias de un híbrido a cada una de las especies parentales; varios híbridos presentan las mismas distancias, razón por la cual se aprecian menos puntos que híbridos.

Figura 13. Resumen de las ploidías y genomas mitocondriales presentes en las 18 familias de híbridos. En las filas se agrupan las familias por ploidía y proporciones de subgenomas parentales (*S. cerevisiae*, SACE; *S. paradoxus*, SAPA), por columnas se agrupan con base en el estado de la mitocondria y por el subgenoma parental predominante. Todas las familias diploides presentan subgenoma *S. cerevisiae*

predominante en la mitocondria. Interesantemente, todas las familias triploides de proporción 1 SACE: 2 SAPA con mitocondrias recombinantes tienen como cepa más cercana al aislado YMX005607.

Figura 14. Co-filogenia de aislados parentales *S. cerevisiae* y *S. paradoxus*. En la izquierda se encuentra la filogenia de máxima verosimilitud de 199 cepas *S. cerevisiae* y en la derecha la filogenia de 84 cepas *S. paradoxus*. Cada línea que une dos nodos representa la cruce de una cepa *S. cerevisiae* con una *S. paradoxus*. Sobre la línea hay una figura (círculo o triángulo dependiendo de la ploidía) que representa un híbrido (o varios con la misma arquitectura genómica), el color externo del triángulo indica el lugar de aislamiento, mientras que el color interno corresponde a la proporción de lecturas nucleares mapeadas a las referencias *S. cerevisiae* y *S. paradoxus*. La figura pequeña contigua a cada híbrido representa su mitocondria y su color indica la proporción de lecturas mitocondriales que mapean a cada una de las referencias parentales.

Figura 15. Co-filogenia de aislados parentales *S. cerevisiae* y *S. paradoxus* clasificados en familias. Al igual que la figura anterior, cada línea que une dos nodos representa la cruce de una cepa *S. cerevisiae* con una *S. paradoxus*. Sobre la línea hay un círculo o triángulo dependiendo de la ploidía, coloreado dependiendo del grupo.

Figura 16. Gráfica de dispersión de porcentajes nucleares y mitocondriales. El eje X indica el porcentaje de lecturas mapeadas al genoma de referencia nuclear *S. cerevisiae*, el eje Y el porcentaje de lecturas mitocondriales mapeadas al genoma mitocondrial de *S. cerevisiae* para cada híbrido. En ambos ejes se resaltan los porcentajes 30, 50 y 66.

FIGURA 17. Correlación Spearman entre los porcentajes de contenido mitocondrial y nuclear. El eje X indica el porcentaje de lecturas mapeadas al genoma de referencia nuclear *S. cerevisiae*, el eje Y el porcentaje de lecturas mitocondriales mapeadas al genoma mitocondrial de *S. cerevisiae*.

Lista de tablas

Tabla 1. Combinaciones encontradas de subgenomas en el núcleo y mitocondria.

Tabla 2. Resumen de los 18 grupos híbridos clasificados en este trabajo. Integrando la información obtenida hasta ahora (pares de cepas más cercanas, ploidía, proporción de lecturas nucleares y mitocondriales mapeadas, y estado de aislamiento) se definieron 18 familias de híbridos. Las proporciones de lecturas se muestran en rojo para *S. cerevisiae* y en azul para *S. paradoxus*.

Tabla 3. Combinaciones encontradas de subgenomas en el núcleo y mitocondria en familias de híbridos.

Tabla 4. Resumen de los aislados parentales más cercanos a los híbridos. En rojo se muestran las cepas más cercanas a los parentales *S. cerevisiae* y en azul las cepas más cercanas a los parentales *S. paradoxus*. En cada celda se indica el nombre del grupo de híbridos con esos parentales (negritas), el número de híbridos en ese grupo, la ploidía, la proporción del sub genoma nuclear (Nuc.) o mitocondrial (Mic.) *S. cerevisiae* (rojo) o *S. paradoxus* (azul) y la distancia filogenética del híbrido a cada parental putativo (Dist.).

Tabla Suplementaria 1. Códigos utilizados y/o desarrollados en este trabajo.

Esta tabla muestra los programas empleados en este trabajo, cada renglón corresponde a un script. Los códigos que modifiqué o desarrollé específicamente para este trabajo se indican con una celda de título de script color verde.

Tabla Suplementaria 2. Aislados empleados en la filogenia *S. paradoxus*.

Ochenta y cuatro cepas pertenecientes a los clados de agave mexicano y norteamericanos (SpB y SpC).

Tabla Suplementaria 3. Aislados empleados en la filogenia *S. cerevisiae*.

Ciento noventa y nueve cepas de los clados de agave mexicano, sudamericano y Guayana francesa.

Tabla Suplementaria 4. Aislados parentales más cercanos a los híbridos.

En esta tabla se despliegan, para cada híbrido, los IDs de las cepas identificadas como más cercanas a los parentales junto con la distancia filogenética a su híbrido más cercano, el número de SNPs usados en la filogenia y número de SNPs compartidos con su híbrido más cercano. La información de la cepa más cercana al parental *S. cerevisiae* se resalta en rojo y naranja, la de *S. paradoxus* en azul; la intensidad del color en la columna de distancias indica la magnitud del valor. Los híbridos con grupo NatHyb corresponden a dos híbridos aislados de ambientes naturales en México y todavía no se cuenta con la información suficiente para clasificarlos bajo los criterios utilizados en este trabajo.

Tabla Suplementaria 5. Distancias filogenéticas entre los aislados parentales *S. cerevisiae*.

Esta tabla muestra las distancias filogenéticas (sustituciones por sitio) entre las 12 cepas *S. cerevisiae* identificadas como más cercanas.

Tabla Suplementaria 6. Distancias filogenéticas entre los aislados parentales *S. paradoxus*.

Esta tabla muestra las distancias filogenéticas (sustituciones por sitio) entre las 8 cepas *S. paradoxus* identificadas como más cercanas.

Índice

Agradecimientos	iii
Lista de figuras.....	v
Lista de tablas.....	ix
Índice	xi
Resumen	1
1. Introducción	2
1.1 Definición de híbrido y su rol evolutivo.....	2
1.2 El género <i>Saccharomyces</i>	3
1.3 El genoma mitocondrial de <i>Saccharomyces</i>	4
1.4 Genomas de referencia y genomas disponibles de la familia <i>Saccharomyces</i>	5
1.5 <i>S. cerevisiae</i> y <i>S. paradoxus</i>	6
1.6 Impacto de la hibridación en la familia <i>Saccharomyces</i>	7
1.7 ¿Cómo se identifica el genoma de un híbrido de levadura?.....	9
1.8 La fermentación de jugo de agave para la producción de mezcal podría ser un proceso que promueve hibridaciones.....	10
1.9 Muestreo de tinas de fermentación de jugo agave por el consorcio YeastGenomesMX ..	11
2. Objetivos.....	12
3. Métodos	12
3.1 Creación de base de datos de aislados híbridos.....	13
3.2 Gráficas de cobertura y extracción de coberturas mitocondriales.	13
3.3 Generación de árboles filogenéticos.....	16
3.4 Clasificación de los aislados en familias	17
4. Resultados.....	18
4.1 Base de datos relacional de los aislados híbridos.....	18
4.2 Los genomas mitocondriales híbridos se dividen en tres categorías.....	21
4.3 En la mayoría de los híbridos, los subgenomas nuclear y mitocondrial mayoritarios provienen de la misma especie	23
4.4 Los aislados parentales más cercanos a los híbridos fueron buscados por medio de filogenias	25

4.5 La proporción del contenido mitocondrial es similar dentro de los individuos de cada grupo.....	38
5. Discusión	44
6. Conclusiones	46
7. Material Suplementario	48
8. Glosario	73
9. Referencias.....	75

Descripción de híbridos de levaduras asociados al agave

Resumen

El género *Saccharomyces* está conformado por ocho especies claramente definidas filogenéticamente, a pesar del gran tiempo de divergencia estas especies son capaces de formar híbridos entre ellas. Se ha demostrado que las hibridaciones en este clado son abundantes y juegan un papel importante en su evolución, ya que habilitan el flujo génico entre especies, aumentan su variabilidad genética y posiblemente permiten adaptaciones a nuevos ambientes. Se definió un conjunto de 64 aislados híbridos (62 mexicanos y 2 brasileños) entre las levaduras *Saccharomyces cerevisiae* y *Saccharomyces paradoxus* los cuales fueron secuenciados y clasificados filogenéticamente. Dicha clasificación permitió identificar 18 grupos definidos por: i) las cepas secuenciadas más cercanas a los parentales, ii) su ploidía y finalmente iii) la proporción del genoma asociada a cada especie parental. En este trabajo diseñé e implementé una base de datos relacional de los aislados híbridos, la cual facilita la consulta de metadatos y gráficas de coberturas nucleares y mitocondriales de cada uno de sus genomas secuenciados. También identificamos a los aislados más cercanos a los parentales de cada uno de los híbridos y, con base en los resultados de los análisis filogenéticos, se lograron asignar a alguno de los 18 grupos. Adicionalmente, se realizó un mapeo competitivo a una referencia concatenada de los genomas mitocondriales de *S. cerevisiae* y *S. paradoxus* para contabilizar las lecturas mapeadas a cada referencia y obtener una aproximación de la contribución mitocondrial de cada progenitor. Nuestros resultados resaltan la diversidad genómica de los híbridos presentes en tinas de fermentación de agave; aportan evidencia de una historia de hibridación compleja con, posiblemente, múltiples eventos de hibridación y sugieren que algunas de las mitocondrias retenidas son recombinantes en diferentes proporciones.

1. Introducción

1.1 Definición de híbrido y su rol evolutivo

La definición biológica de una especie propone que ésta constituye un conjunto de individuos que son capaces de entrecruzarse y producir descendencia fértil. Bajo esta misma definición, los miembros de una especie están reproductivamente aislados de los miembros que conforman otra especie, debido a que presentan barreras geográficas y/o genéticas que impiden la producción de descendencia fértil. Sin embargo, se han descrito varios casos en los que individuos de especies diferentes se cruzan (Bendixen, et al., 2022; Gabaldón, 2020; Mixão y Gabaldón, 2020; Samarasinghe y Xu, 2018). Para este trabajo, definiremos a un híbrido como el resultado viable de la cruce entre individuos pertenecientes a especies diferentes.

La formación de híbridos interespecie viables puede tener varias consecuencias. Los híbridos permiten la mezcla de los acervos genéticos de las poblaciones parentales. Este fenómeno aumenta la variabilidad genética de la población, lo cual a su vez podría permitir la adaptación a nuevos nichos y ambientes (Stelkens, R y Bendixsen, D, 2022).

La formación de nuevos linajes estables es un fenómeno posible si la población híbrida ha desarrollado la adecuación suficiente para proliferar en algún ambiente específico. En el caso de organismos con reproducción exclusivamente sexual, la formación de nuevos linajes a través de la hibridación es un fenómeno poco común ya que requiere recuperar la fertilidad en el híbrido. Este no es el caso de organismos con reproducción asexual, ya que éstos conservan su capacidad de reproducirse mitóticamente de forma indefinida, así que su reproducción asexual no está condicionada a la recuperación de su fertilidad.

Los híbridos interespecie han sido identificados con facilidad en organismos macroscópicos, dado que presentan características morfológicas de ambas especies. Un ejemplo muy conocido de un híbrido interespecie es la mula, que proviene de la cruce entre una yegua y un burro; en este caso, el híbrido presenta características que permiten distinguirla de

sus progenitores: la cabeza y orejas son similares a las de un burro mientras que la dentadura y resto de su cuerpo son más parecidos a las de un caballo (Mercado, 2022). Por otro lado, en microorganismos como las levaduras, la identificación de híbridos a partir de sus características morfológicas no es una tarea fácil, ya que los híbridos no necesariamente presentan caracteres visibles diferentes; esto aunado a que, incluso entre cepas de la misma especie, puede haber una extensa variedad de morfologías (Gabaldón, 2020).

Los híbridos de levaduras son relevantes no sólo por ser un fenómeno interesante *per se* ni por su impacto evolutivo únicamente, sino también por su utilidad industrial. Es bien conocido que la levadura encargada de la fermentación de cerveza lager (proceso caracterizado por sus bajas temperaturas de fermentación) es un híbrido interspecie con alta criotolerancia conocido como *Saccharomyces pastorianus*; el hallazgo de la naturaleza híbrida de esta levadura fue posible gracias al descubrimiento y estudio de las especies que le dieron origen (Libkind et al., 2011).

1.2 El género *Saccharomyces*

Las levaduras son organismos eucariontes unicelulares pertenecientes al reino fungi, generalmente están asociadas con procesos fermentativos; sin embargo, éstas abarcan un amplio espectro de nichos y hábitats (Peter, et al., 2018). Hasta la fecha, se han identificado más de 1500 especies de levaduras (Kurtzman, et al, 2011) y se estima que este número de manera proporcional al muestreo en ambientes naturales, se amplíen los métodos de cultivo y mejoren las técnicas moleculares de identificación.

Entre los grupos mejor estudiados de levaduras se encuentra el género *Saccharomyces*, conformado por ocho especies bien definidas filogenéticamente a las que se les conoce como *Saccharomyces sensu stricto*: *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. jurei*, *S. kudriavzevii*, *S. arboricola*, *S. eubayanus* y *S. uvarum* (Alsammar y Delneri, 2020). Las levaduras de este género derivan de un evento de duplicación del genoma completo (Whole Genome Duplication o WGD, por sus siglas en inglés), que actualmente se reconoce como resultado de un evento de hibridación interspecie (Wolfe, 2015; Marcet-

Houben y Gabaldón, 2015). Las ocho especies presentan alta colinealidad entre sus genomas, así como reproducción sexual y asexual por gemación. Adicionalmente, las levaduras de esta familia presentan una estrategia muy particular en el consumo de fuentes de carbono conocida como Efecto Crabtree. Este fenómeno consiste en reprimir el metabolismo respiratorio y obtener energía a través de la fermentación de glucosa aún en la presencia de oxígeno, con la finalidad de consumir rápidamente las fuentes de carbono e inhibir el crecimiento de posibles competidores cercanos a través del etanol producido durante la fermentación.

La especie más estudiada de este género, *Saccharomyces cerevisiae*, ha acompañado a la humanidad durante miles de años en la fermentación de alimentos y bebidas, interacción que favoreció su domesticación. Aun cuando la presencia de *S. cerevisiae* se asocia principalmente con ambientes antropogénicos relacionados con la fermentación, su presencia en un ambiente dado no está necesariamente restringido a las prácticas humanas. Derivado de estudios poblacionales de la especie sabemos que existen poblaciones de *S. cerevisiae* y otras levaduras *Saccharomyces* en ambientes naturales no antropogénicos. De hecho, *S. mikatae*, *S. jurei*, *S. arboricola* y *S. eubayanus* son especies que, hasta ahora, se han clasificado como exclusivamente de ambientes naturales, mientras que *S. cerevisiae*, *S. paradoxus*, *S. kudriavzevii* y *S. uvarum* abarcan ambientes antropogénicos y naturales (Alsammar, 2019; Gallegos-Casillas, 2023).

1.3 El genoma mitocondrial de *Saccharomyces*

A diferencia del genoma mitocondrial de los animales, los genomas mitocondriales de *Saccharomyces* son ricos en intrones, en secuencias no codificantes y en elementos repetidos (Solieri, 2010), además sus 85 Kb los hacen significativamente más extensos que el genoma mitocondrial humano que tan solo cuenta con 17 Kb. Las mitocondrias cuentan con su propia maquinaria de traducción, sin embargo, gran parte de su proteoma es codificado en el núcleo y posteriormente importado al espacio mitocondrial. Alrededor del 12% de los genes nucleares en *S. cerevisiae* están destinados a codificar el proteoma mitocondrial (Merz, 2019) y como se mencionó anteriormente, las levaduras

Saccharomyces son capaces de vivir sin ADN mitocondrial debido a que llevan a cabo su producción energética mediante el proceso de fermentación (Malina, et al., 2018).

Los genomas mitocondriales de *Saccharomyces* pueden recombinar entre ellos. Después del apareamiento, la nueva célula se encuentra en un estado de heteroplasma, i.e., presencia de más de un tipo de genoma mitocondrial dentro de una misma célula, durante el cual los genomas mitocondriales pueden recombinar. La resolución del estado de heteroplasma puede fijar un haplotipo parental o un haplotipo recombinante que es heredado por medio de segregación vegetativa (Hénault, et al., 2022). Se sabe que la recombinación mitocondrial es importante para la correcta herencia del genoma mitocondrial (Fritsch, et al., 2014) y que tiene consecuencias en su variación genética (Peris, et al., 2017).

A pesar de la íntima interacción entre el genoma nuclear y el mitocondrial, éstos tienen procesos de flujo génico y evolución diferentes. De Chiara y colaboradores (2020) analizaron la variación del genoma mitocondrial de *S. cerevisiae* de la colección de los 1011 genomas (Peter, et al. 2018) y encontraron que el coeficiente de variación (π) en secuencias codificantes de genomas nucleares es mucho menor ($\pi \sim 0.003$) que aquel de las secuencias codificantes de los genomas mitocondriales ($\pi \sim 0.0085$). También observaron, comparando filogenias nucleares y mitocondriales, que los genomas mitocondriales presentan niveles mayores de mezcla (*admixture*, en inglés) entre poblaciones de *S. cerevisiae* que los genomas nucleares.

1.4 Genomas de referencia y genomas disponibles de la familia *Saccharomyces*

La abundancia de recursos genómicos, debida a la reducción en costos de secuenciación, ha permitido una definición más clara de los integrantes de la familia *Saccharomyces*. Actualmente, las ocho especies del género *Saccharomyces* cuentan con genomas de referencia de muy alta calidad públicamente disponibles. Además, se han secuenciado más de 3000 genomas de diversos aislados pertenecientes a este

género, de los cuales más del 76% pertenecen a *S. cerevisiae*, 23% a las especies *S. paradoxus*, *S. eubayanus* y *S. uvarum*, mientras que menos del 1% se reparte entre *S. mikatae*, *S. jurei*, *S. kudriavzevii* y *S. arboricola*. Este sesgo se puede atribuir a un conjunto de razones: el descubrimiento reciente de algunas de estas especies, a su limitación geográfica, a los métodos de aislamiento, y a la utilidad científica e industrial que se le atribuye a cada especie (Bendixsen, et al., 2022).

1.5 *S. cerevisiae* y *S. paradoxus*

La levadura *Saccharomyces cerevisiae* es un organismo ampliamente estudiado gracias a su relativa simplicidad como organismo eucarionte y por su utilidad en la producción de alimentos y bebidas. Su genoma fue el primer genoma eucariota completamente secuenciado en 1996 (Goffeau, et al., 1996), el cual consta de 16 cromosomas lineales con longitudes en un rango de 235 a 1,400 kb. Cuenta con múltiples copias de un genoma mitocondrial de ~85 kb y frecuentemente presenta un elemento extra cromosomal conocido como 2 μ m de 6 kb de longitud (Futcher, 1988).

Canónicamente, el ciclo de vida de *S. cerevisiae* consta de una fase diploide y una haploide, en donde las células diploides proliferan mitóticamente en la presencia de un ambiente rico en nutrientes y pasan por meiosis cuando la disponibilidad de alimentos es reducida para producir asci compuestos de hasta cuatro esporas haploides (Fischer, et al., 2021). Aunque el ciclo de vida de *S. cerevisiae* está descrito estrictamente como cambios entre la fase haploide y diploide, se ha observado la presencia de poblaciones diploides con alta heterocigosidad y poliploides (Fisher, et al., 2021; Peter, et al., 2018), lo cual invita a pensar que el ciclo de vida de *S. cerevisiae* es más complejo de lo que se conoce hasta ahora.

Saccharomyces paradoxus es la especie filogenéticamente más cercana a *S. cerevisiae*. Al igual que *S. cerevisiae*, cuenta con 16 cromosomas lineales con longitudes entre 230 a 1,500 kb y un genoma mitocondrial de 85 kb. Su ciclo de vida canónico es prácticamente idéntico al de *S. cerevisiae* (Tsai, et al., 2008), con una fase diploide y una haploide con la producción de esporas. *S. paradoxus* es considerada como una especie

mayoritariamente silvestre debido a su frecuencia de aislamiento en ambientes naturales. A diferencia de su especie hermana, la presencia de *S. paradoxus* en fermentaciones es poco común, probablemente debido a que se encuentra en desventaja competitiva con *S. cerevisiae*. Esta suposición está basada en la abundancia de *S. paradoxus* en la naturaleza y su escasez en ambientes fermentativos (Nikulín, et al, 2020). Sin embargo, sus capacidades en la producción de bebidas han sido estudiadas (Majdak, 2020) y los resultados han sido prometedores.

Estas dos especies divergieron aproximadamente hace 5 millones de años, presentan colinealidad o sintenia entre sus genomas con escasos rearrreglos (Fischer, et al., 2000) y tienen en promedio 12% de divergencia nucleotídica a lo largo del genoma. Ambas especies son capaces de aparearse entre ellas y formar híbridos, este fenómeno parece ser común en la naturaleza de estas levaduras (Barbosa, et al., 2016).

1.6 Impacto de la hibridación en la familia *Saccharomyces*

El clado *Saccharomyces* cuenta con un gran historial de hibridaciones. Existe evidencia de un evento de duplicación del genoma completo precedente al origen del clado *Saccharomyces* el cual resultó de un evento de hibridación entre dos especies ancestrales (Marcet-Houben y Gabaldón, 2015; Wolfe, 2015), que también afectó a las especies de los géneros *Saccharomyces sensu lato*, *Candida* y algunas *Kluyveromyces* (Knop, 2006). Dentro del propio clado *Saccharomyces* las hibridaciones son recurrentes. Gracias a la disponibilidad de recursos genómicos se pudieron identificar como híbridos a muchas cepas que anteriormente habían sido clasificadas como especies diferentes con base en evidencia morfológica o metabólica o a secuenciación de genes ribosomales. Uno de los ejemplos más conocidos es *S. pastorianus*, resultado de la cruce entre *S. cerevisiae* y *S. eubayanus*, este híbrido es ampliamente utilizado en la producción de cerveza lager gracias a su criotolerancia y a su eficiencia fermentativa, las cuales heredó de *S. eubayanus* y *S. cerevisiae*, respectivamente (Alsammar, 2019).

Inmediatamente después del evento de hibridación, los híbridos presentan dos subgenomas divergentes dentro de un solo núcleo, lo cual implica altos niveles de

heterocigosidad e incompatibilidades entre ambos subgenomas (Gabaldón, 2020) (Mixao y Gabaldón, 2020). En respuesta, el genoma híbrido pasa por un proceso de estabilización por medio de varios mecanismos como: duplicación completa o parcial del genoma, aneuploidías, rearrreglos, pérdidas de genes, conversión de secuencias homeólogas mediada por recombinación (Mixao y Gabaldón, 2018), que ultimadamente resultan en pérdida de heterocigosidad (generalmente referido como Loss of heterozygosity, *LOH* por sus siglas en inglés) (Mixao y Gabaldón, 2018). Las regiones de pérdida de heterocigosidad en el genoma híbrido son sometidas a presión selectiva para reducir el número de incompatibilidades entre los genomas parentales u otorgan alguna ventaja adaptativa al híbrido (Gabaldón, 2020). Viéndolo en su totalidad, el genoma de un híbrido *Saccharomyces* puede, dependiendo de su historia demográfica, de recombinación y de selección, ser un mosaico aneuploide muy diferente a la estructura genómica de las especies parentales o puede estar tan conservado y mantener la estructura genómica de la primera generación del evento de hibridación, como en el caso del híbrido de Alpechín (D'Angiolo, 2019). Véase Figura 1.

Niveles altos de introgresión sugieren la ocurrencia de hibridaciones. El clado *S. cerevisiae* de Alpechín se originó a partir de una hibridación ancestral, la población híbrida resultante recobró fertilidad gracias a la pérdida de heterocigosidad resultante de la inestabilidad genómica y permitió la acumulación de introgresiones en la población de *S. cerevisiae* (D'angiolo, 2019). El origen del clado de Alpechín fue elucidado gracias al aislamiento de un descendiente clonal de la hibridación ancestral entre *S. cerevisiae* y *S. paradoxus*, el cual coexiste con la población *S. cerevisiae* introgresada (Figura 1).

La existencia de híbridos interspecie fértiles habilita el flujo génico de los acervos genéticos de las especies parentales por medio de retrocruzas entre la población híbrida y la población de sus progenitores. Dichas cruzas resultan en introgresiones las cuales pueden introducir variantes adaptativas en la población. Por definición, una introgresión es la transferencia de variación genética entre especies por medio de hibridación y retrocruzas recurrentes (Martin y Jiggins, 2017).

Algunos ejemplos de poblaciones de *S. cerevisiae* con gran cantidad de introgresiones provenientes de *S. paradoxus* son las constituidas por aislados de Alpechín, Bioetanol Brasileño, Agave Mexicano y Guyana Francesa (D'angiolo, 2019)

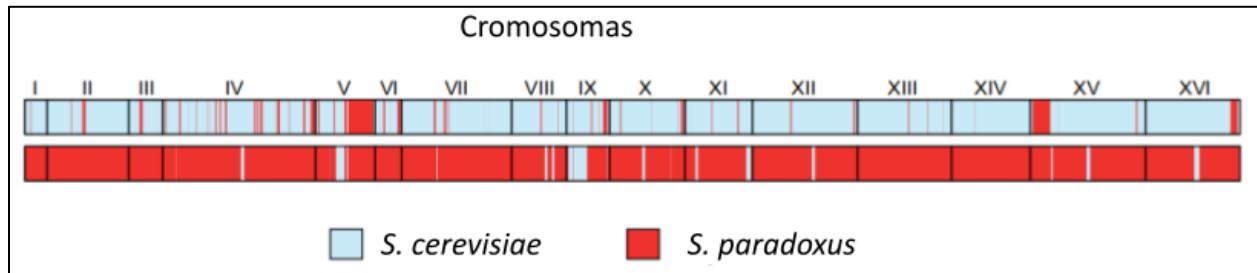


Figura 1. Mapa genómico del híbrido de Alpechín.

Esta figura fue modificada de D'Angiolo et al. (2019). El ADN está coloreado en función a su origen: *S. cerevisiae* (azul) y *S. paradoxus* (rojo). Se pueden observar las regiones de pérdida de heterocigosidad como segmentos del color opuesto intercalados en los subgenomas.

1.7 ¿Cómo se identifica el genoma de un híbrido de levadura?

Previo al advenimiento de la genómica, la identificación de híbridos *Saccharomyces* consistía en el uso de técnicas moleculares como cariotipados, polimorfismo de longitud de fragmentos de restricción, hibridación ADN-ADN, entre otras; estas técnicas resultan útiles para la identificación de híbridos interespecie, no obstante, no son capaces de caracterizar por completo los genomas híbridos (Morales y Dujon, 2012). El auge de la genómica, mediante la secuenciación de genomas completos y las herramientas bioinformáticas para analizarlo, permitió obtener un nivel de resolución mucho mayor de los genomas híbridos. Ahora podemos, por ejemplo, inferir ploidía por medio de cobertura, identificar el origen de los subgenomas del híbrido mediante mapeos competitivos, ubicar los puntos en el genoma donde ocurrió recombinación e identificar a las cepas más cercanas filogenéticamente (ver métodos).

Se han secuenciado más de 200 genomas de híbridos interespecie del género *Saccharomyces*, de los cuales la mitad pertenecen a híbridos entre las especies *S. cerevisiae* y *S. eubayanus*, pero también incluyen, en menor proporción, híbridos *S. cerevisiae* x *S. kudriavzevii*, *S. cerevisiae* x *S. paradoxus*, *S. cerevisiae* x *S. uvarum*, *S.*

cerevisiae x *S. mikatae* y *S. eubayanus* x *S. uvarum* (Bendixsen, et al., 2022). Gran parte de los genomas híbridos secuenciados fueron obtenidos de ambientes fermentativos de vino, cerveza y pan, razón por la que casi todos tienen contribución genómica de *S. cerevisiae*; a pesar de este sesgo, también se han reportado híbridos de ambientes naturales de dos subpoblaciones de *S. paradoxus* (Leducq, et al, 2016).

1.8 La fermentación de jugo de agave para la producción de mezcal podría ser un proceso que promueve hibridaciones.

La elaboración de mezcal es un proceso artesanal que emplea fermentaciones abiertas, implicando que es un proceso semi controlado en donde los microorganismos presentes durante la fermentación son diversos y variables a lo largo del proceso. Las levaduras *S. cerevisiae* presentes en las tinajas de fermentación de jugo de agave tienen un alto grado de introgresiones, en promedio cuentan con 137.5 genes provenientes de su especie hermana *S. paradoxus*, a diferencia de las cepas *S. cerevisiae* del resto del mundo (sin incluir Alpechín) del estudio de Peter y colaboradores de 2018, las cuales cuentan con un promedio menor a 50 genes introgresados (Sedeño, 2023). El muestreo en México aisló una gran cantidad de híbridos *S. cerevisiae* y *S. paradoxus* que coexisten con sus especies parentales. Dichos híbridos fueron aislados independientemente en diferentes estados del país y presentan gran diversidad genómica entre ellos (ver métodos, sección 1).

La frecuencia con la que han sido aisladas las cepas híbridas y la diversidad de arquitecturas genómicas que presentan nos invitan a preguntarnos si estos híbridos son resultado de uno o múltiples eventos de hibridación. La ocurrencia de múltiples eventos de hibridación independientes es un fenómeno comúnmente reportado en plantas, sin embargo, también ha sido reportado en levaduras, como es el caso de *Candida orthosilopsis*, una levadura patogénica asociada a humanos (Schröder, et al., 2016). Para dilucidar los múltiples orígenes de esta levadura, Schröder y colaboradores analizaron polimorfismos nucleares, los loci MTL y los genomas mitocondriales e infirieron al menos cuatro eventos de hibridación independientes entre dos especies parentales.

1.9 Muestreo de tinas de fermentación de jugo agave por el consorcio YeastGenomesMX

El consorcio YeastGenomesMX tiene como propósito describir y analizar la diversidad genómica y funcional de levaduras en México (Gallegos-Casillas, et al., 2023). Para lograrlo, hicieron una colección de levaduras mexicanas la cual incluye muestras aisladas de tinas de fermentación de agave de 68 destilerías distribuidas en 13 estados de la república: Colima, Durango, Estado de México, Guanajuato, Guerrero, Jalisco, Michoacán, Oaxaca, Puebla, San Luis Potosí, Sonora, Tamaulipas, y Zacatecas (Lopez-Gallegos, et al., 2023). Más recientemente, se muestrearon tanto destilerías como sitios naturales circundantes en tres regiones (Lopez-Gallegos, en prep.). De esta colección, se secuenciaron más de 300 genomas completos de levaduras *Saccharomyces* con la tecnología DNBSEQ de BGI Genomics, los cuales se encuentran disponibles para uso del consorcio.

Hasta ahora, se han encontrado 62 aislados híbridos los cuales fueron confirmados mediante análisis de cobertura de mapeos competitivos. Los híbridos fueron aislados de 8 estados diferentes (Sonora, Durango, Nuevo León, San Luis Potosí, Guanajuato, Michoacán, Estado de México y Oaxaca) y presentan arquitecturas genómicas variables. Tomando en cuenta diferentes niveles de información como filogenias, ploidía, contenido nuclear y el estado de aislamiento, 44 de estos híbridos fueron previamente clasificados en al menos 13 grupos diferentes por el consorcio YeastGenomesMX. A pesar de esta agrupación, aún es importante clasificar al resto de los híbridos y mejorar los análisis previos.

Con la finalidad de mejorar la clasificación previa, añadí información sobre los genomas mitocondriales de los híbridos, e implementé de forma sistematizada la inferencia de la relación filogenética entre los híbridos y sus especies parentales considerando un conjunto de datos más apropiado (ver métodos). Este tipo de inferencia consiste en construir filogenias para cada subgenoma incluyendo un conjunto de secuencias de los posibles parentales para determinar el origen de los subgenomas presentes en los

híbridos (D'angiolo, et al., 2020). En la clasificación realizada por el consorcio, el análisis filogenético se realizó integrando un conjunto de secuencias de posibles parentales que incluía cepas de híbridos y del resto del mundo, no sólo americanas; este enfoque presenta dos principales limitaciones: La presencia de otros subgenomas híbridos dentro de la filogenia puede sesgar el posicionamiento del subgenoma en cuestión y la inclusión de cepas del resto del mundo puede alterar la estructura del árbol.

2. Objetivos

El objetivo general de este trabajo es definir y describir el conjunto de aislados híbridos secuenciados de la colección de YeastGenomesMX.

Los objetivos específicos son:

1. Crear una base de datos para el conjunto de aislados híbridos secuenciados.
2. Identificar las cepas secuenciadas más cercanas a los parentales de cada híbrido.
3. Describir el genoma mitocondrial de cada híbrido y contrastarlo con su genoma nuclear.

3. Métodos

Los análisis se llevaron a cabo en el clúster “DNA” del Laboratorio Nacional de Visualización Científica Avanzada (LAVIS) de la UNAM campus Juriquilla. Dentro de este clúster se pueden encontrar los genomas de referencia utilizados y lecturas de los aislados de mezcal. Los códigos que utilicé para este trabajo se encuentran disponibles dentro del repositorio de Github <https://github.com/esantiago24/DescripcionDeHibridosDeLevadurasAsociadosAlAgave> y se enlistan en la Tabla Suplementaria 1 indicando los códigos que modifiqué o creé específicamente para este trabajo.

3.1 Creación de base de datos de aislados híbridos

Notion es una aplicación web que ofrece una variedad de herramientas de organización y colaboración de proyectos en tiempo real (Notion Labs Inc, 2016). Dentro de ellas, Notion ofrece una herramienta en línea de creación de bases de datos relacionales de fácil manipulación, acceso y amplia personalización; por estas razones elegimos este software para esta sección del proyecto.

La información incluida en este nuevo recurso proviene de datos directos producidos por el laboratorio derivados del muestreo descrito en Gallegos-Casillas, et al. (2023). Primero se realizó un filtrado para seleccionar aquellas muestras que fueron categorizadas como híbridos *S. cerevisiae* x *S. paradoxus*, dicho subconjunto fue descargado como formato separado por comas y posteriormente fue adjuntado al nuevo recurso en Notion.

Estos datos cuentan con más de 10 campos de información acerca de los aislados, entre ellas: identificador (ID), sitio de colecta (estado, municipio, coordenadas geográficas, mezcalería), características del híbrido (grupo, ploidía, contenido nuclear de *S. cerevisiae*, contenido nuclear de *S. paradoxus*), etc. Se revisó manualmente que todos los campos de información estuvieran completos para cada aislado. En algunos casos, la información faltante se obtuvo de los resultados de citometría de flujo para la ploidía y de un registro de las localizaciones muestreadas, ambos proveídos por el consorcio YeastGenomesMX.

3.2 Gráficas de cobertura y extracción de coberturas mitocondriales.

Como parte del *pipeline* que emplea el consorcio YeastGenomesMX en el procesamiento de las muestras, se generan gráficas de cobertura de genoma nuclear (Figura 2 A) y genoma mitocondrial (Figura 2 B) para identificar la especie a la que pertenecen los genomas nucleares y mitocondriales de cada aislado. Este alineamiento se realiza con una referencia concatenada de diez genomas que incluye a las ocho especies del género *Saccharomyces* y a las levaduras *K. marxianus* y *P. kluyveri*. Dichas gráficas muestran

la profundidad de cada ventana de 1kb de la referencia en el eje Y y en el eje X la referencia coloreada en función a cada uno de los 10 genomas concatenados. Se confirmó la presencia de híbridos *S. cerevisiae* x *S. paradoxus* empleando gráficas como la que se muestra en la Figura 2, en donde se puede observar que el aislado al cual pertenecen las gráficas de cobertura es un híbrido ya que la profundidad es mayor en las referencias pertenecientes a los genomas de *S. cerevisiae* y *S. paradoxus* (indicadas como los picos de cobertura verde y morado, respectivamente). Para algunos de los híbridos en la nueva base de datos, fue necesario generar dichas gráficas ya que no se contaba con ellas; para crearlas, se utilizó el *script* de python *coverage_plot.py* (véase apartado 3. Métodos) desarrollado por el consorcio YeastGenomesMX.

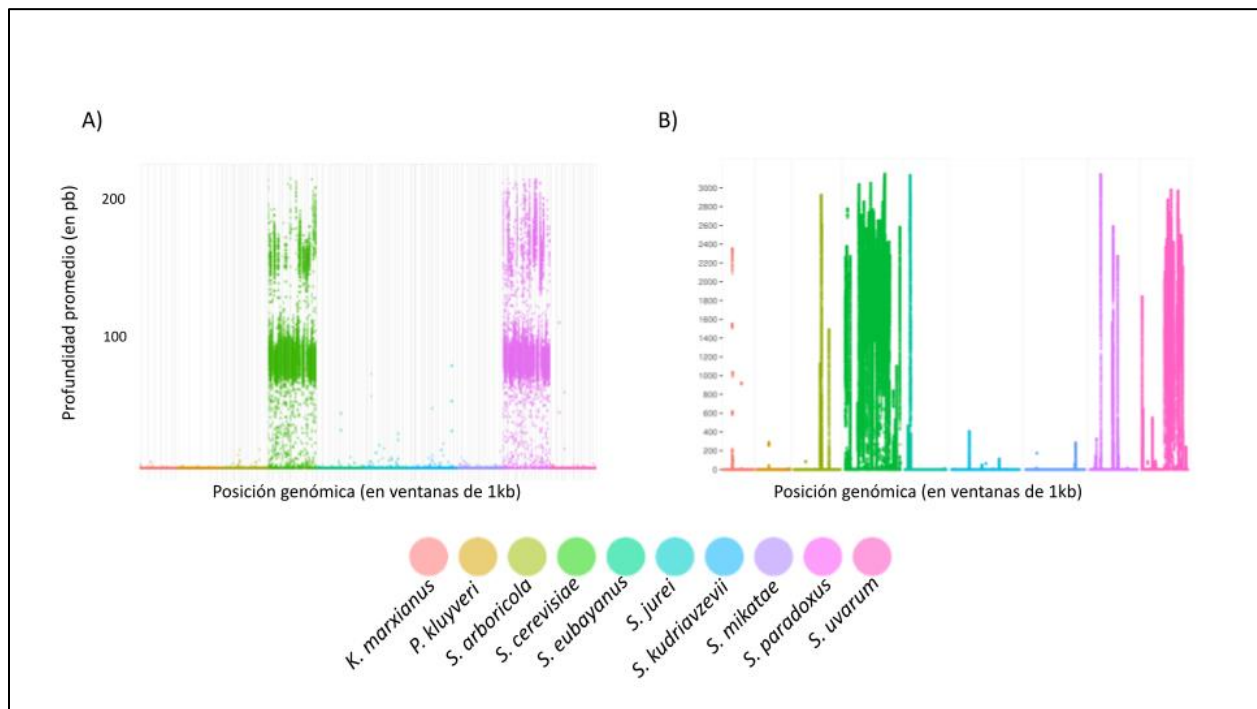


Figura 2. Gráficas de profundidad de cobertura nuclear y mitocondrial del aislado YMX005605.

En el eje X se encuentra la referencia concatenada dividida en ventanas no sobrepuestas de 1 kb, cada color indica el organismo al que pertenece. En el eje Y se muestra la profundidad promedio de cada ventana. A) Gráfica de cobertura de 10 genomas nucleares de levaduras. B) Gráfica de cobertura de 10 genomas mitocondriales de levaduras.

Se realizó un mapeo competitivo de las lecturas de cada aislado híbrido a una referencia concatenada, la cual incluye únicamente las mitocondrias de *S. cerevisiae* y *S. paradoxus* (Figura 3). El mapeo fue realizado con *bwa* (version 0.7.4) con parámetros *mem -t 10 -M*, ordenado e indexado con *samtools* (versión 1.9). La contabilización del total de lecturas mapeadas fue hecha con *samtools idxstats* y posteriormente se calculó la proporción de lecturas mapeadas a cada una de las referencias.

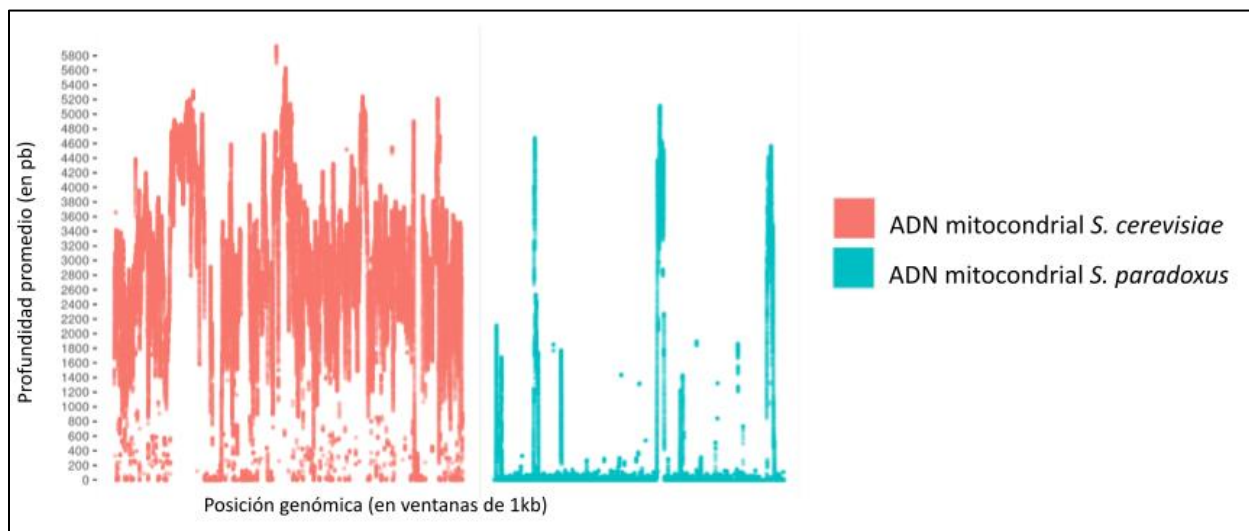


Figura 3. Gráfica de cobertura a referencia mitocondrial concatenada *S. cerevisiae* y *S. paradoxus* del aislado YMX004626.

A diferencia de la figura 1, estas gráficas corresponden al mapeo del genoma mitocondrial del híbrido YMX004626 a una referencia concatenada de los genomas mitocondriales únicamente de *S. cerevisiae* y *S. paradoxus*. En este ejemplo, la mitocondria observada es mayoritariamente del parental *S. cerevisiae*.

Con el fin de visualizar la proporción del contenido mitocondrial y nuclear de los híbridos, desarrollé un código en R (*NuclearAndMitPlots.R*, Tabla Suplementaria 1) capaz de generar gráficas de dona representando las proporciones parentales en el genoma mitocondrial y gráficas de barra para representar el contenido nuclear de cada uno de ellos (Figura 4). A manera de ejemplo, en la Figura 4 se observa que el genoma nuclear del híbrido YMX506F06 tiene alrededor de 66% de aporte del subgenoma de *S. cerevisiae* (componente rojo de la barra) y alrededor de 33% de aporte del subgenoma

de *S. paradoxus* (componente azul de la barra), mientras que en el genoma mitocondrial el aporte de *S. paradoxus* es apenas de 10% (azul).

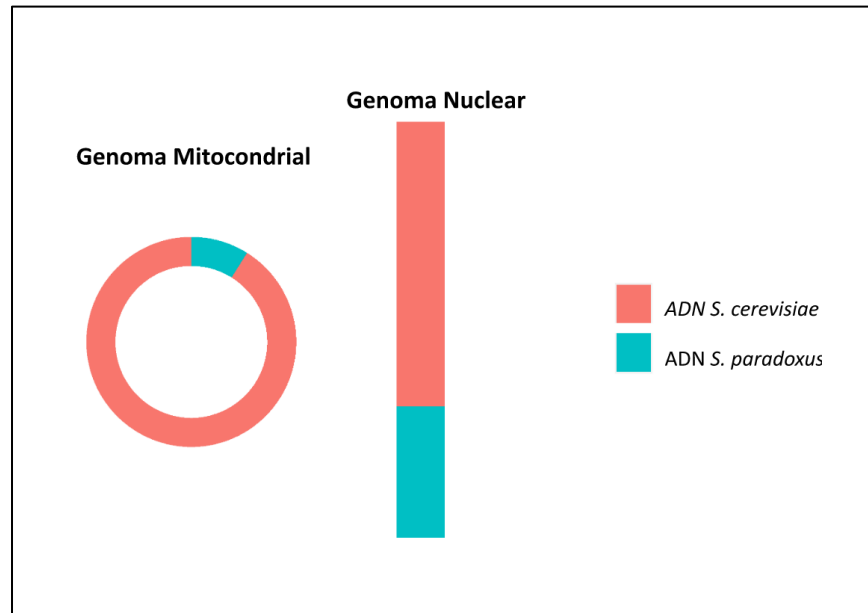


Figura 4. Gráficas para contrastar el contenido mitocondrial con el nuclear de cada híbrido. Estas gráficas ejemplifican al híbrido YMX506F06 con aportación nuclear de 66% *S. cerevisiae* y 33% *S. paradoxus*; y aportación mitocondrial prevalente de *S. cerevisiae*.

3.3 Generación de árboles filogenéticos

Para identificar a los aislados de *S. paradoxus* más cercanos a los híbridos (o posibles parentales), se seleccionaron 84 cepas pertenecientes a los clados de agave mexicano y norteamericano (SpB y SpC; Tabla Suplementaria 2). Con la finalidad de maximizar la información disponible en la secuencia del híbrido, se desarrolló un *script* en Python (*reduce_phylip.py*, Tabla Suplementaria 1) que elimina los sitios de ambigüedad (N) y *gaps* (-) del híbrido en el alineamiento múltiple utilizado para elaborar la filogenia (archivo phylip). De esta manera, se asegura que todos los sitios empleados en la generación del árbol sean informativos para el híbrido y así contar con una precisión mayor para posicionarlo dentro de la filogenia.

Se generaron sistemáticamente filogenias de máxima verosimilitud a partir de SNPs bialélicos con el programa RAxML versión 8.2.12 (`raxml HPC-PHREADS-AVX2 -f a -x 12345 -p 12345 -N 100 -T 8 -m GTRGAMMA -O`) incluyendo al set de 84 cepas y un híbrido por árbol, resultando en un total de 64 filogenias diferentes. Se incluyó únicamente a un híbrido por filogenia para evitar alteraciones en la estructura del árbol causada por la presencia de otros híbridos. Para realizar el mismo proceso en *S. cerevisiae*, se seleccionaron 199 cepas de los clados de agave mexicano, sudamérica y Guayana francesa (Tabla Suplementaria 3), el proceso de generación de filogenias fue el mismo que con las cepas de *S. paradoxus* y se incluyeron las 199 cepas y un híbrido por árbol.

Se calculó la distancia filogenética (promedio de sustituciones por sitio) entre los subgenomas de cada híbrido y la cepa secuenciada más cercana a su parental con la función `cophenetic.phylo` de la librería `ape` (5.7-1) de R, esta función obtiene las distancias entre todos los pares de nodos terminales de un árbol filogenético usando la longitud de las ramas como unidades. Los árboles de cada híbrido fueron visualizados usando `Microreact`.

Adicionalmente, se generaron filogenias de los sets *S. cerevisiae* y *S. paradoxus* puros con la finalidad de visualizar ambos árboles simultáneamente y representar a los híbridos uniendo mediante líneas los nodos de aquellas cepas identificadas como posibles parentales (Figura 14). La filogenia pareada se creó con la función `cophylo` del paquete `ape` (5.7-1) y fue modificada con el editor de gráficos vectoriales `Inkscape`.

3.4 Clasificación de los aislados en familias

Una vez completada la extracción de las coberturas mitocondriales y la identificación de las cepas parentales más cercanas para cada híbrido, se integró la información obtenida con los metadatos existentes de cada aislado (ploidía y lugar de aislamiento). Con esta información, se agruparon en familias considerando el siguiente criterio en el orden enlistado:

1. Mismo par de cepas parentales
2. Ploidía
3. Porcentajes de lecturas nucleares mapeadas a las referencias *S. cerevisiae* y *S. paradoxus*
4. Porcentajes de lecturas mitocondriales mapeadas a las referencias *S. cerevisiae* y *S. paradoxus*
5. Lugar de aislamiento

4. Resultados

4.1 Base de datos relacional de los aislados híbridos

Con la finalidad de contar con un recurso de fácil acceso a los metadatos de los híbridos, creamos una base de datos relacional en la plataforma Notion (Figura 5). Esta nueva base de datos nos permite 1) contar con un recurso curado, actualizable y de fácil acceso dedicado a los aislados híbridos secuenciados, y 2) relacionar visualmente gráficas de profundidades de cobertura nucleares y mitocondriales de cada híbrido (Figura 5B). Cabe mencionar que ninguna de las opciones anteriores se podía hacer directamente con las herramientas implementadas hasta ahora por el consorcio YeastGenomesMX.

Enlace de acceso:

<https://www.notion.so/8135c05ebd21460f991beaf7bd94abae?v=948b016e365944828cb057af4fb9d6b0&pvs=4>

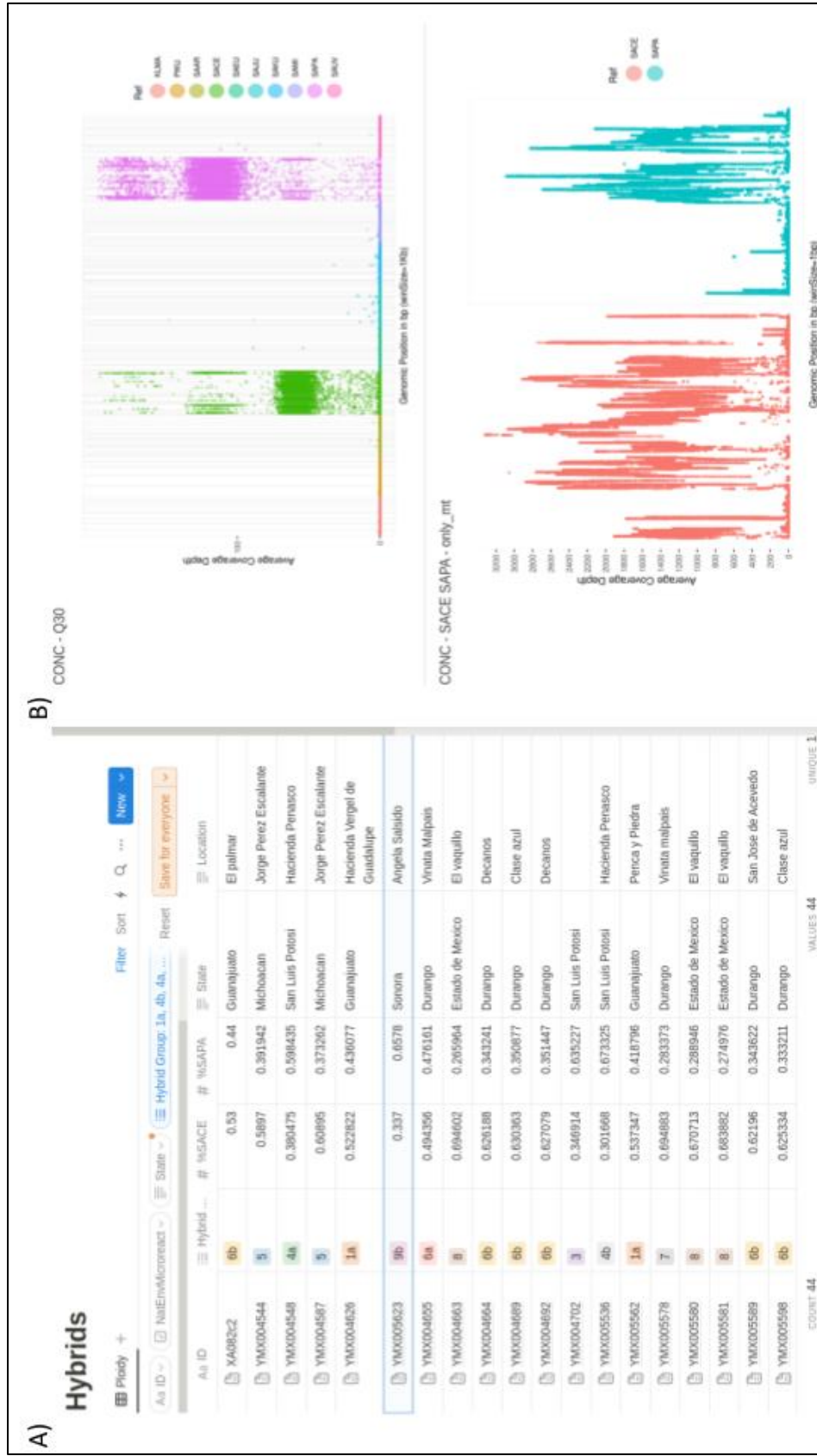


Figura 5. Base de datos de los aislados híbridos.

A) Base de datos relacional con información de 64 híbridos *S. cerevisiae* x *S. paradoxus*. En las columnas se encuentran diferentes campos de información como: identificador (ID), sitio de colecta (estado, municipio, coordenadas geográficas, mezcalería), características del híbrido (grupo de híbrido, ploidía, contenido nuclear de *S. cerevisiae*, contenido nuclear de *S. paradoxus*), etc. B) Vista de una entrada de la base de datos. Al seleccionar el ID de una muestra se despliegan, en una ventana adyacente, los metadatos de esa muestra junto con las gráficas de cobertura de cobertura nuclear (superior) y mitocondrial (inferior).

El conjunto seleccionado se compone de 64 híbridos de las especies *S. cerevisiae* (SACE) y *S. paradoxus* (SAPA), la mayoría de ellos provenientes de tinajas de fermentación de agave, a excepción de dos muestras de híbridos de Brasil previamente reportados por Barbosa y colaboradores en 2016. Los 62 híbridos mexicanos se distribuyen en 8 estados y presentan, en conjunto, al menos 3 arquitecturas genómicas diferentes: diploides con una copia de cada sub genoma parental [proporción 1 SACE : 1 SAPA] (Figura 6 A), triploides con dos copias del sub genoma *S. cerevisiae* y una *S. paradoxus* [proporción 2 SACE : 1 SAPA] (Figura 6 B) y triploides con dos copias del sub genoma *S. paradoxus* y una *S. cerevisiae* [proporción 1 SACE : 2 SAPA] (Figura 6 C). El consorcio YeastGenomesMX únicamente había clasificado 44 de estos híbridos en 13 grupos con base en su ploidía, arquitectura genómica y su posición en filogenias. En este trabajo se clasificaron a 64 híbridos *S. cerevisiae* x *S. paradoxus* (62 mexicanos y 2 brasileños) en 18 familias como se describe a continuación.

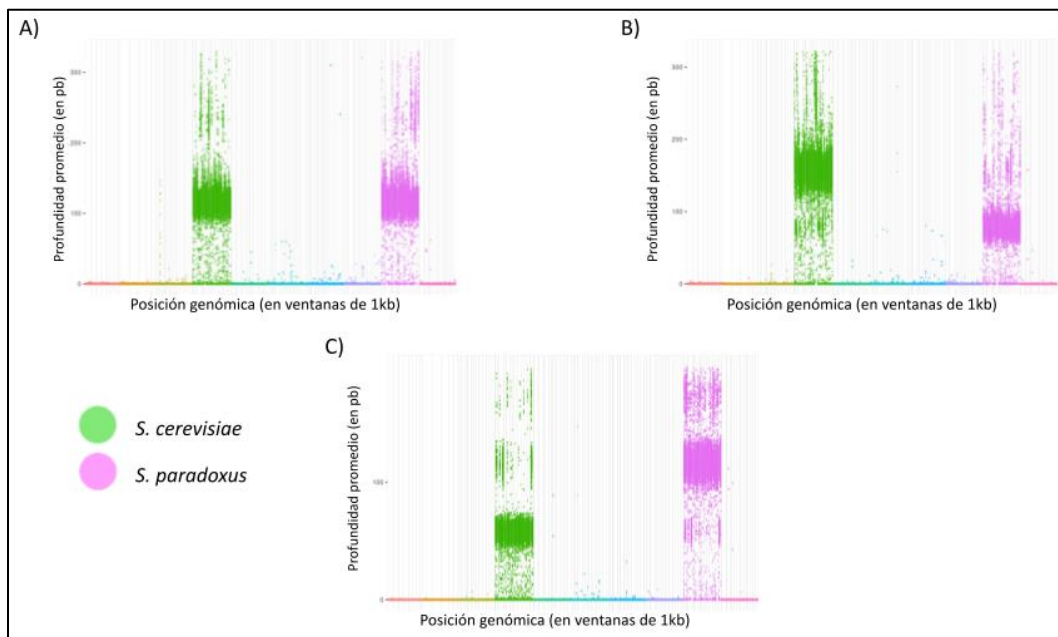


Figura 6. Tres tipos de arquitecturas genómicas en los híbridos aislados ejemplificadas por los aislados A) YMX004626, B) YMX182c2 y C) YMX005623

A partir de las gráficas de cobertura se identificaron 3 tipos de arquitecturas genómicas. La diferencia en proporciones se puede apreciar mediante la profundidad promedio de cada sub genoma, como en los casos B) y C) donde uno de los genomas tiene el doble de profundidad que el otro.

4.2 Los genomas mitocondriales híbridos se dividen en tres categorías

Interesados por conocer la mitocondria parental retenida por los híbridos, realicé mapeos competitivos de las lecturas mitocondriales de los híbridos (Figura 5 B: Gráfica de cobertura mitocondrial) y cuantifiqué el número de lecturas alineadas a cada referencia para calcular sus proporciones.

Observé que los genomas mitocondriales se dividían en tres categorías principales: 1) mayor proporción *S. cerevisiae*, 2) mayor proporción *S. paradoxus* y 3) recombinantes (Figura 7).

Para clasificar a una mitocondria como recombinante consideré que los subgenomas mitocondriales se encontraran en proporción 3:1 y además inspeccioné las gráficas de cobertura mitocondriales buscando aquellas que tuvieran un patrón de cobertura con regiones altamente cubiertas en ambos subgenomas, como el que se muestra en la gráfica de cobertura del híbrido YMX005605 (Figura 8 B).

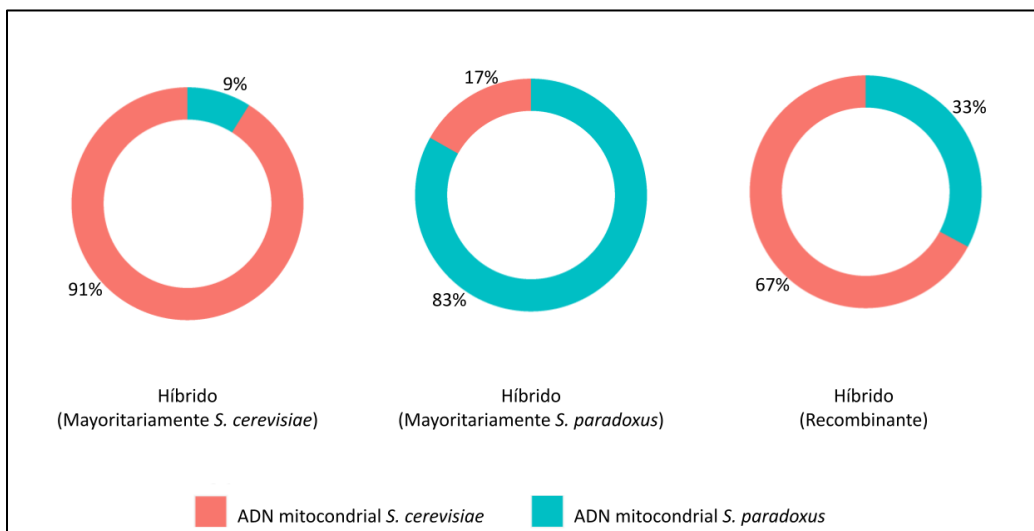


Figura 7. Genomas mitocondriales híbridos divididos en tres categorías.

Después de contabilizar las lecturas mapeadas a cada referencia mitocondrial y calcular sus proporciones para cada híbrido, clasificamos a las mitocondrias en tres categorías. 1) mayor proporción *S. cerevisiae* ejemplificada por YMX506C10, 2) mayor proporción *S. paradoxus* ejemplificada por YMX004593 y 3) recombinantes ejemplificada por YMX005605.

Para entender el origen del genoma mitocondrial en los híbridos de la tercera categoría (recombinantes), comparé las gráficas de cobertura mitocondrial de cada híbrido con las de los aislados más cercanos a la cepa parental (Figura 8 B).

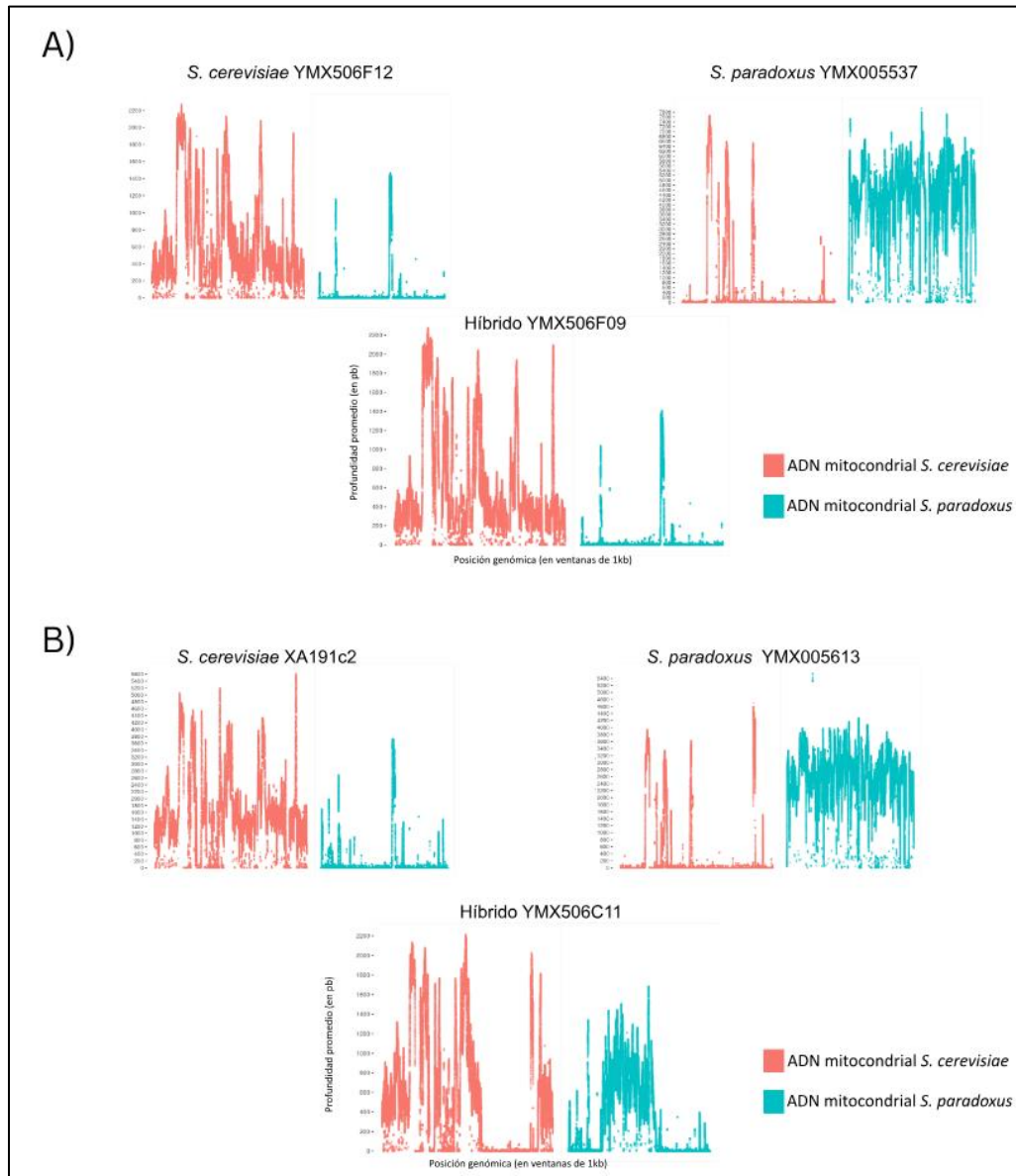


Figura 8. Origen de las mitocondrias recombinantes.

En cada panel, las dos gráficas superiores pertenecen a las mitocondrias de los dos aislados puros más cercanos a los padres de un híbrido dado, en la fila inferior se encuentra la gráfica de la cobertura mitocondrial de dicho híbrido. A) Se observa que el híbrido conserva uno de los patrones parentales, sugiriendo que no hubo recombinación. B) En este caso, se puede observar que el patrón de cobertura mitocondrial presentado por el híbrido no corresponde a ninguno de los patrones parentales, sugiriendo que puede ser recombinante.

En el panel A se muestra la cobertura mitocondrial del híbrido YMX506F09, así como de las cepas de *S. cerevisiae* (YMX506F12) y *S. paradoxus* (YMX005537) más cercanas a las cepas parentales. Se puede observar que el patrón de cobertura mitocondrial en el híbrido es idéntico al del parental putativo *S. cerevisiae*. A diferencia del caso anterior, el patrón de cobertura del genoma mitocondrial del híbrido YMX506C11 mostrado en el panel B es diferente tanto al de *S. cerevisiae* como de *S. paradoxus*, lo que sugiere que: i) uno o más eventos de recombinación pudieron haber ocurrido y que eventualmente este nuevo haplotipo se fijó en el híbrido; o ii) que en las cepas de nuestra colección no existe la cepa parental más cercana a este híbrido y por ello su patrón de cobertura es diferente al de los putativos parentales.

Treinta y seis aislados híbridos secuenciados presentan una mitocondria recombinante. En 5 de ellos, uno de sus padres ya tenía mitocondria recombinante. Para los 31 restantes, las cepas más cercanas a sus parentales no tenían un patrón de cobertura mitocondrial que sugiriera recombinación.

4.3 En la mayoría de los híbridos, los subgenomas nuclear y mitocondrial mayoritarios provienen de la misma especie

Nos preguntamos si la prevalencia del contenido mitocondrial en alguna de las dos especies parentales está relacionada con el contenido predominante en el genoma nuclear. Para ello contrastamos el contenido de *S. cerevisiae* y *S. paradoxus* en el genoma mitocondrial (gráficas de dona) y nuclear (barras) de cada híbrido.

Observamos que en 56 de 64 (87%) de los híbridos el subgenoma mitocondrial y nuclear predominantes provienen de la misma especie, para ejemplificar estas proporciones se muestra al híbrido YMX507A01 (Figura 9 A). De este porcentaje de aislados con el mismo subgenoma predominante en el núcleo y la mitocondria (n=56), en 47 de 56 *S. cerevisiae* es prevalente, y solo en 9 *S. paradoxus* es prevalente.

En el resto de los híbridos (8/64, 13%), el origen del subgenoma mitocondrial predominante no coincide con el subgenoma nuclear predominante, un ejemplo de este caso es el híbrido YMX005624 que se muestra en la Figura 9 B. Véase Tabla 1

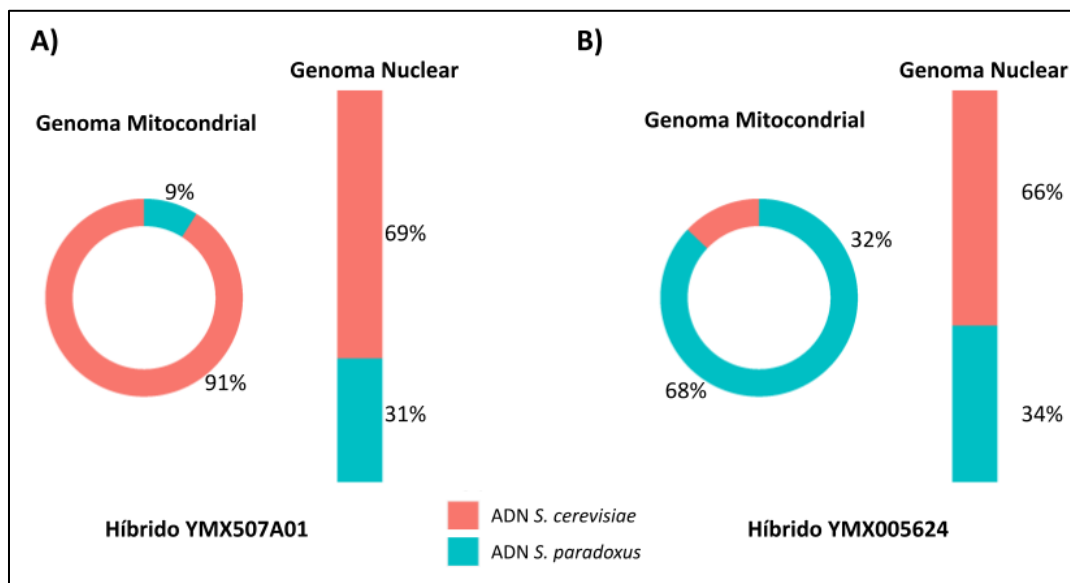


Figura 9. El origen del genoma mitocondrial prevalente no siempre coincide con el origen del genoma nuclear prevalente.

En esta figura se muestran las proporciones de ADN de *S. cerevisiae* y *S. paradoxus* en el genoma mitocondrial y nuclear de un híbrido representativo para cada grupo. A) El híbrido YMX507A01 presenta genomas nuclear y mitocondrial mayoritariamente de *S. cerevisiae*. B) El híbrido YMX005624 presenta genoma nuclear *S. cerevisiae* y mitocondrial *S. paradoxus*.

Tabla 1. Combinaciones encontradas de subgenomas en el núcleo y mitocondria.

	Número de aislados
<i>S. cerevisiae</i> predominante en núcleo y mitocondria	47
<i>S. paradoxus</i> predominante en núcleo y mitocondria	9
Subgenoma predominante distinto en núcleo y mitocondria	8

4.4 Los aislados parentales más cercanos a los híbridos fueron buscados por medio de filogenias

Para identificar a los aislados de *S. cerevisiae* y *S. paradoxus* más cercanos a los híbridos (o posibles parentales), se construyeron filogenias de máxima verosimilitud a partir de SNPs bialélicos comparando los aislados de las parentales de los clados cercanos (199 cepas *S. cerevisiae* y 84 de *S. paradoxus*) con cada uno de los híbridos. Como ejemplo, en la Figura 10 se muestra la filogenia del híbrido XA020c1 con las cepas de *S. cerevisiae* (panel A) y con las cepas de *S. paradoxus* (panel B). A partir del par de árboles de cada híbrido se calcularon las distancias filogenéticas entre todos los aislados y se identificaron a aquellas cepas más cercanas filogenéticamente con cada híbrido.

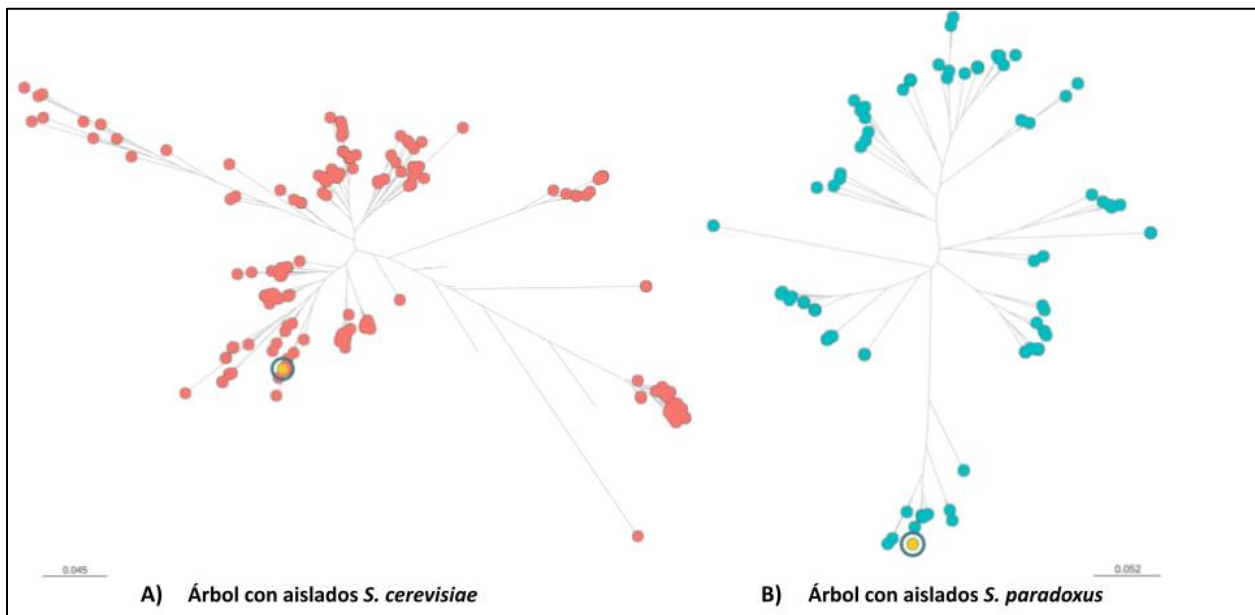


Figura 10. Árboles filogenéticos del híbrido XA020c1.

Para cada uno de los híbridos se generaron dos filogenias de máxima verosimilitud: Una incluyendo las 199 cepas *S. cerevisiae* (panel A) y otra con las 84 cepas *S. paradoxus* (panel B). Para el caso del híbrido XA020c1, el número de sitios utilizados para hacer la filogenia *S. cerevisiae* es de 359,053 y la escala en el árbol es de 0.045 sustituciones por sitios. Para la filogenia *S. paradoxus* se utilizaron 190,490 sitios y la escala en el árbol es de 0.052 sustituciones por sitio. En ambos árboles se resalta la posición del híbrido XA020c1 con un nodo de color amarillo.

Los datos se muestran en la Tabla Suplementaria 4, donde cada renglón corresponde a un híbrido y las columnas a las cepas aisladas más cercanas a los parentales junto con la distancia filogenética entre ellos, los SNPs empleados en la filogenia con la cual se calculó la distancia y el número de SNPs que comparte cada híbrido con sus putativos parentales. En total, contamos con 17 pares de cepas parentales diferentes.

Se graficaron curvas de densidad de las distancias filogenéticas de los híbridos a las cepas puras *S. cerevisiae* y *S. paradoxus* con el objetivo de observar su distribución (Figura 11). Para ambas especies, el rango de distancias abarca desde 0 a 0.25 sustituciones por sitio y la mediana de sus distribuciones cae en 0.03 sust. p/sitio para *S. cerevisiae* y en 0.04 sust. p/sitio para *S. paradoxus*; sin embargo, la varianza en las distancias a *S. cerevisiae* es ligeramente mayor que las de *S. paradoxus*. Cabe mencionar que las distancias filogenéticas no son directamente comparables con otros estimados de tasas de sustitución fuera de este análisis, ya que fueron calculadas a partir de SNPs bialélicos concatenados en un alineamiento múltiple con un conjunto de cepas específico. Sin embargo, para los propósitos de este estudio, las distancias filogenéticas nos aportan una medición de cercanía entre cepas. Un número menor de sustituciones por sitio implica que el subgenoma del híbrido y aquel de la cepa pura parental más cercana son más similares entre ellos.

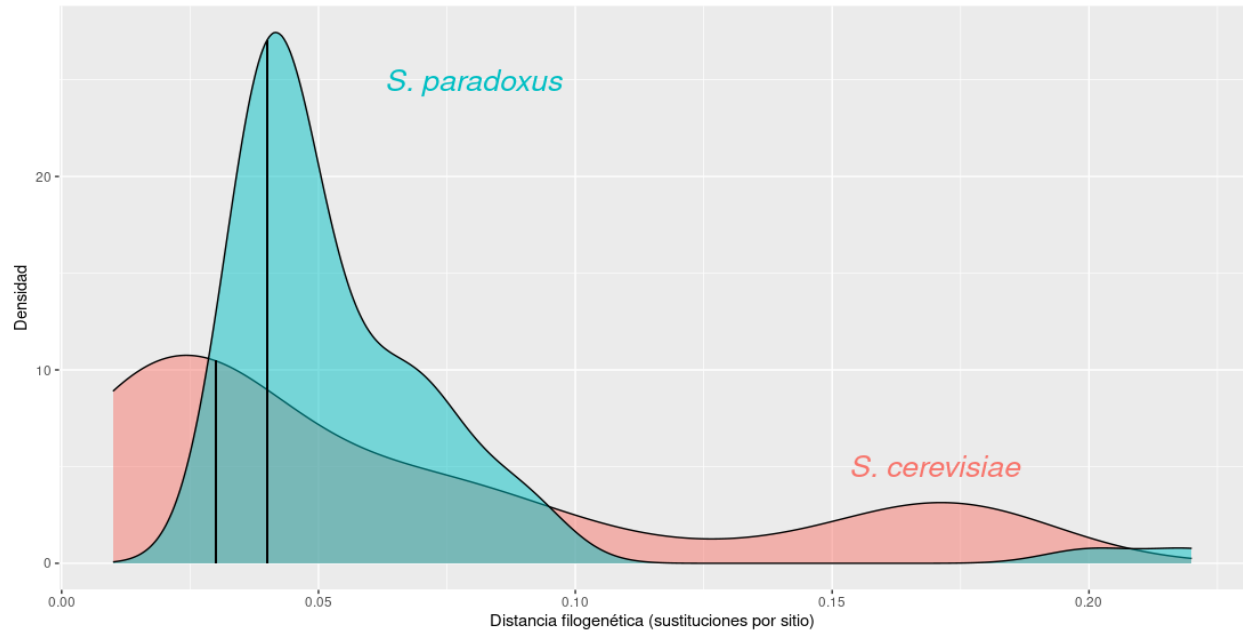


Figura 11. Densidad de las distancias filogenéticas de cada uno de los híbridos a las cepas más cercanas a los parentales *S. cerevisiae* y *S. paradoxus*.

En rojo se muestra la distribución de *S. cerevisiae* y en azul la de *S. paradoxus*, las medianas de las distribuciones se indican con líneas sólidas. Las distancias se encuentran en las columnas 3 y 7 de la Tabla Suplementaria 4.

Es importante prestar atención a las distancias filogenéticas de los híbridos a cada uno de sus padres putativos; un valor bajo en la distancia nos brinda confianza de realmente haber identificado una cepa cercana a los padres. Por ejemplo, si un híbrido A presenta una distancia de 0.02 sustituciones por sitio con su parental putativo *S. cerevisiae* y otro híbrido B presenta una de 0.17; la distancia del híbrido B, en comparación con el híbrido A, es 8.5 veces mayor y por lo tanto nos sugiere que el aislado secuenciado más cercano es muy similar al que fue el parental *S. cerevisiae* del híbrido A, pero no podríamos afirmar lo mismo para el híbrido B.

Para visualizar la información de las distancias filogenéticas calculadas a partir de las filogenias, se creó una gráfica de coordenadas paralelas en donde para cada aislado híbrido se relacionan las distancias hacia las cepas más cercanas a los parentales (Figura 12). Ésta nos permite visualizar e identificar con mayor facilidad aquellos híbridos para los cuales no contamos con la cepa más cercana a los parentales dentro de la colección empleada en este análisis. Por ejemplo, la distancia de los híbridos BR017c1

y BR018c1, coloreados en naranja, con las cepas más cercanas a los parentales *S. cerevisiae* es baja (0.04 sustituciones por sitio), mientras que con aquellas más cercanas a los parentales *S. paradoxus* es alta (0.22 sustituciones por sitio), con lo que podemos inferir que tenemos mayor seguridad del parental putativo de *S. cerevisiae* pero no del parental *S. paradoxus*.

Tomando en consideración los pares de cepas más cercanas a los híbridos en conjunto con los diferentes niveles de información previamente empleados para la clasificación de los híbridos (par de cepas secuenciadas identificadas como más cercanas, ploidía, proporción de lecturas nucleares y mitocondriales mapeadas, y lugar de aislamiento), se redefinieron los grupos y se clasificaron a los híbridos que no tenían un grupo asignado, resultando en **18 nuevas familias o grupos de híbridos** (Tabla 2).

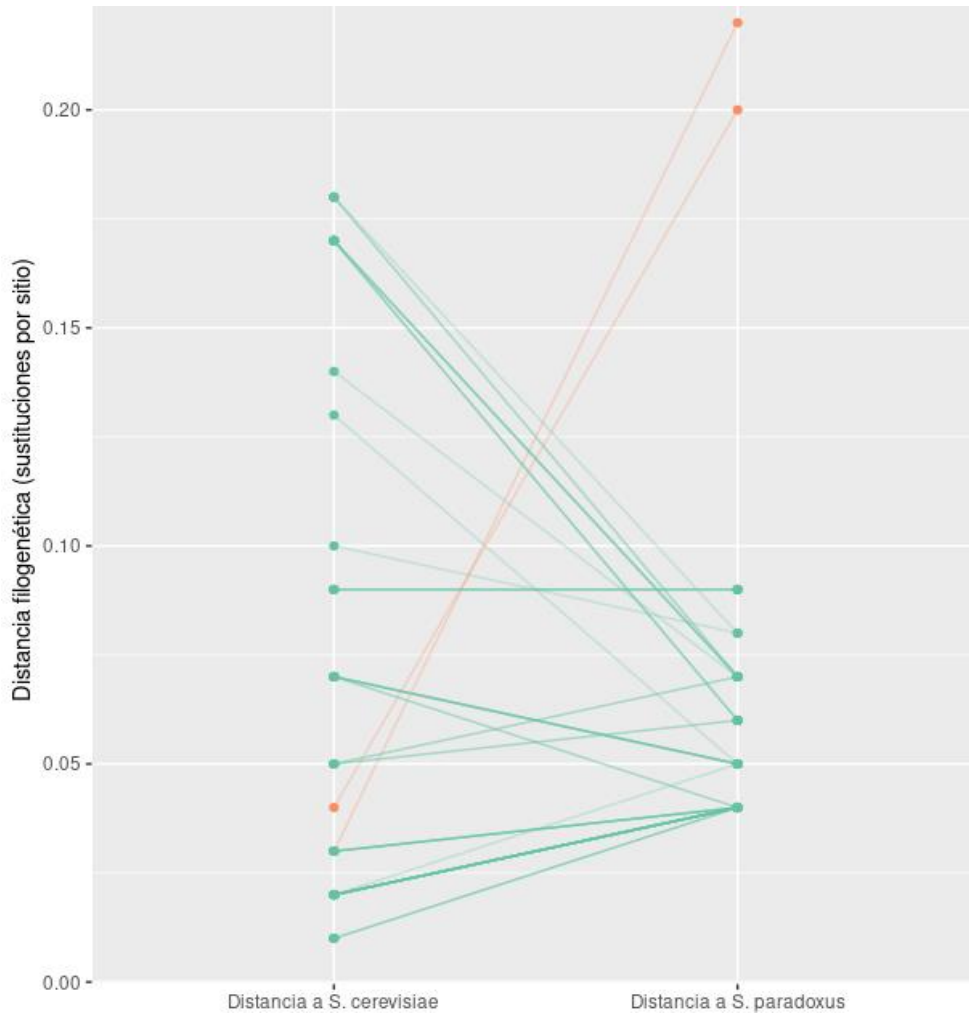


Figura 12. Gráfica de coordenadas paralelas de las distancias filogenéticas de los híbridos a las cepas más cercanas a los parentales *S. cerevisiae* y *S. paradoxus*.

La primera barra vertical representa la medida de distancia filogenética a la cepa más cercana a los parentales *S. cerevisiae*, la segunda indica la distancia a *S. paradoxus*. Cada recta relaciona las distancias de un híbrido a cada una de las especies parentales; varios híbridos presentan las mismas distancias, razón por la cual se aprecian menos puntos que híbridos.

Tabla 2. Resumen de los 18 grupos híbridos clasificados en este trabajo

Integrando la información obtenida hasta ahora (pares de cepas más cercanas, ploidía, proporción de lecturas nucleares y mitocondriales mapeadas, y estado de aislamiento) se definieron 18 familias de híbridos. Las proporciones de lecturas se muestran en rojo para *S. cerevisiae* y en azul para *S. paradoxus*. Esta tabla se divide en varias páginas por necesidad de presentación.

Grupo	Integrantes (ID)	Ploidía	Proporción SACE: SAPA nuclear	Proporción SACE : SAPA mitocondrial	Lugar de Aislamiento	Parental <i>S. cerevisiae</i>	Parental <i>S. paradoxus</i>
1	XA082c2 XB008c4 YMX004605 YMX004626 YMX005562	2n	53 : 44	90 : 10	Guanajuato	YMX507B05 (Jalisco)	YMX005537 (Guanajuato)
2a	YMX506F06	3n	78 : 22	91 : 9	Nuevo León	BR023c1 (Brazil)	YMX005537 (Guanajuato)
2b	YMX506G01 YMX507A08	3n	78 : 21	91 : 9	Nuevo León	XA121c18 (Tamaulipas)	YMX005537 (Guanajuato)
3	DK003c30 YMX004702	3n	34 : 66	33 : 67	San Luis Potosí	SRR800827 (Desconocido)	YMX005607 (Sonora)
4a	YMX005605	2n	47 : 52	67 : 33	Sonora	SRR800827 (Desconocido)	YMX004555 (Michoacán)
4b	YMX005623 YMX005624	3n	34 : 66	68 : 32	Sonora	SRR800827 (Desconocido)	YMX005607 (Sonora)
5	YMX005580	3n	67 : 39	92 : 8	Estado de México	XA017c1 (Estado de México)	YMX005613 (Guanajuato)
6a	XA192c4 YMX004655 YMX005579 YMX506C05 YMX506C06 YMX506C08 YMX506D05 YMX506D06	2n	49 : 51	64 : 35	Durango	YMX005597 (Durango)	YMX005613 (Guanajuato)

6b	<p>XA182c2 XA184c6 YMX004664 YMX004689 YMX005589 YMX506C11 YMX506D08 YMX506D09 YMX506E03 YMX506E04 YMX506H11 XA022c2 XA183c1 YMX004692 YMX005598 YMX506C09 YMX506D03 YMX506E02 YMX506H12 YMX507A03</p>	3n	64 : 34	64 : 35	Durango	XA191c2 (Durango)	YMX005613 (Guanajuato)
7	<p>YMX005578 YMX506C10 YMX506H10 YMX507A01</p>	3n	69 : 30	91 : 9	Durango	XA211c1 (Zacatecas)	YMX005613 (Guanajuato)
8	<p>YMX506G07</p>	3n	78 : 21	92 : 8	Nuevo León	XB001c4 (Oaxaca)	YMX005537 (Guanajuato)
9	<p>XA020c1 YMX004663 YMX005581</p>	3n	70 : 27	92 : 8	Estado de México	YMX005592 (Estado de México)	YMX005613 (Guanajuato)

10	XA256c15 XB003c8 XB255c12 YMX004544 YMX004587	3n	60 : 39	16 : 84	Michoacán	YMX005594 (Jalisco)	YMX005613 (Guanajuato)
11a	XB003s8 YMX004548 YMX004593 YMX005560	3n	37 : 61	16 : 84	San Luis Potosí	YMX005594 (Jalisco)	YMX004559 (Guanajuato)
11b	YMX005536	3n	30 : 67	16 : 84	San Luis Potosí	YMX005594 (Jalisco)	YMX005537 (Guanajuato)
12	DK002c39	3n	33 : 67	36 : 64	Oaxaca	YMX506F12 (Jalisco)	YMX005607 (Sonora)
13	YMX506F09	3n	68 : 31	91 : 9	Nuevo León	YMX506F12 (Jalisco)	YMX005537 (Guanajuato)
14	BR017c1 BR018c1	3n	76 : 23	30 : 70	Brasil	BR023c1 (Brazil)	YMX506E06 (Durango)

Figura 13. Resumen de las ploidías y genomas mitocondriales presentes en las 18 familias de híbridos.

En las filas se agrupan las familias por ploidía y proporciones de subgenomas parentales (*S. cerevisiae*, SACE; *S. paradoxus*, SAPA), por columnas se agrupan con base en el estado de la mitocondria y por el subgenoma parental predominante. Todas las familias diploides presentan subgenoma *S. cerevisiae* predominante en la mitocondria. Interesantemente, todas las familias triploides de proporción 1 SACE: 2 SAPA con mitocondrias recombinantes tienen como cepa más cercana al aislado YMX005607.

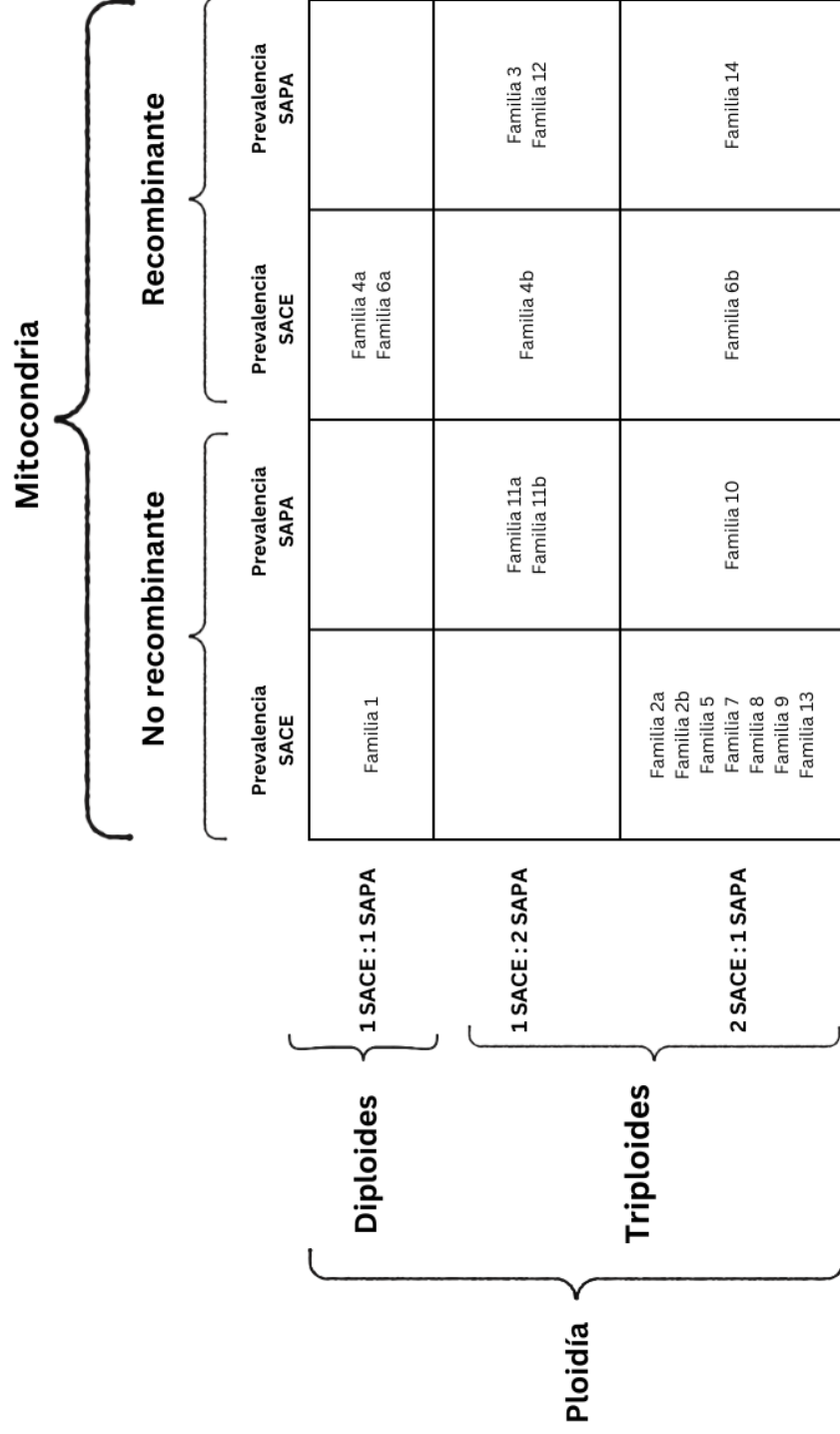


Tabla 3. Combinaciones encontradas de subgenomas en el núcleo y mitocondria en familias de híbridos.

	Número de aislados	Número de familias
<i>S. cerevisiae</i> predominante en núcleo y mitocondria	47	11
<i>S. paradoxus</i> predominante en núcleo y mitocondria	9	4
Subgenoma predominante distinto en núcleo y mitocondria	8	3

Posterior a la nueva clasificación por familias, se generaron filogenias de los sets *S. cerevisiae* y *S. paradoxus* puros con la finalidad de visualizar ambos árboles simultáneamente y representar a los híbridos uniendo mediante líneas los nodos de aquellas cepas identificadas como posibles parentales (Figura 14). Adicionalmente, se modificó la figura 14 para ilustrar las familias de híbridos definidas (Figura 15).

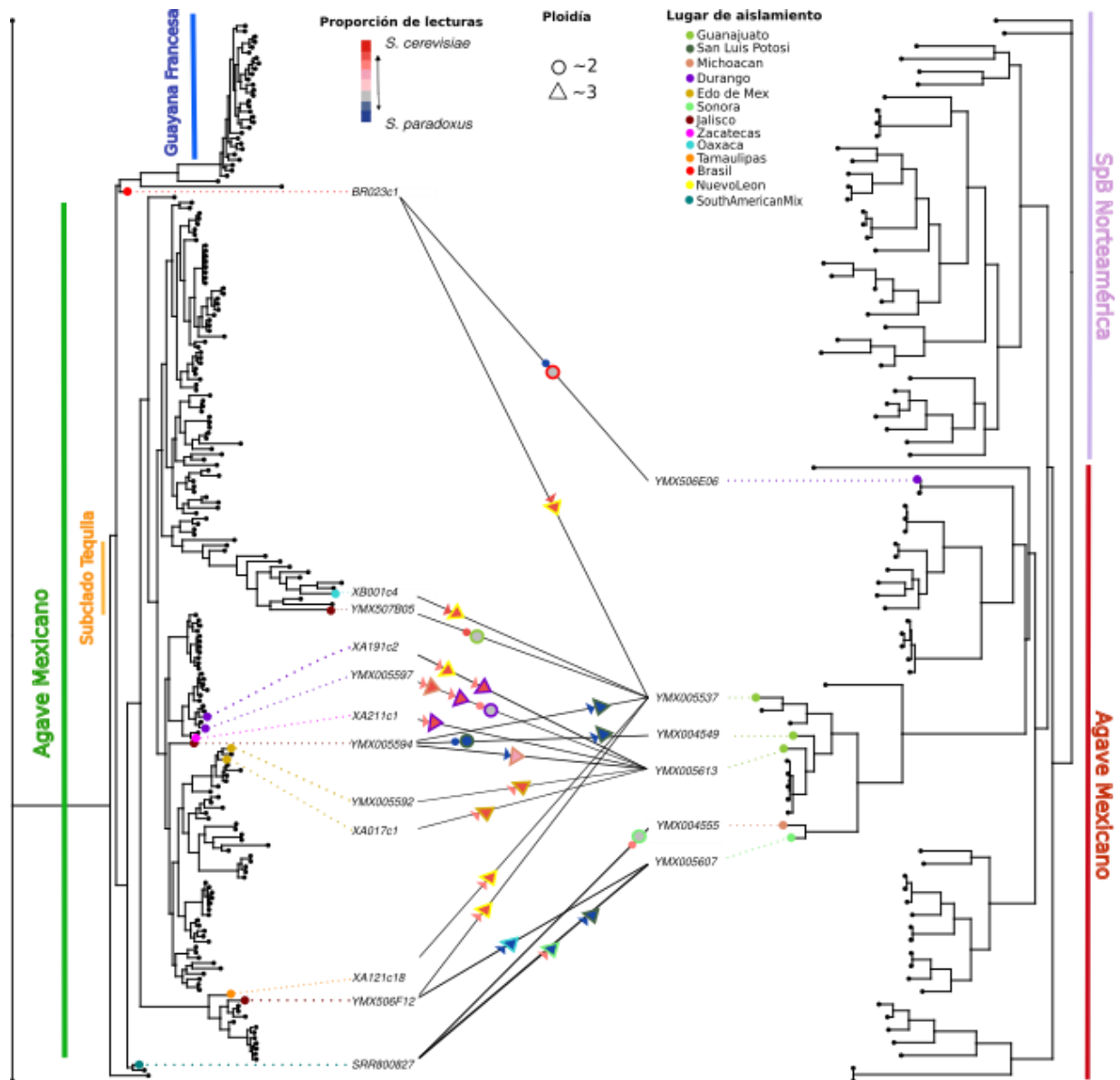


Figura 14. Co-filogenia de aislados parentales *S. cerevisiae* y *S. paradoxus*.

En la izquierda se encuentra la filogenia de máxima verosimilitud de 199 cepas *S. cerevisiae* y en la derecha la filogenia de 84 cepas *S. paradoxus*. Cada línea que une dos nodos representa la cruce de una cepa *S. cerevisiae* con una *S. paradoxus*. Sobre la línea hay una figura (círculo o triángulo dependiendo de la ploidía) que representa un híbrido (o varios con la misma arquitectura genómica), el color externo del triángulo indica el lugar de aislamiento, mientras que el color interno corresponde a la proporción de lecturas nucleares mapeadas a las referencias *S. cerevisiae* y *S. paradoxus*. La figura pequeña contigua a cada híbrido representa su mitocondria y su color indica la proporción de lecturas mitocondriales que mapean a cada una de las referencias parentales.

En la figura 14 se puede observar que todas las cepas (de ambas especies) identificadas como más cercanas a los híbridos pertenecen a aislados obtenidos de destilerías de agave, a excepción de la cepa más cercana al parental *S. cerevisiae* de los híbridos brasileños (BR023c1) la cual también proviene de Brasil. Interesantemente, la cepa más cercana al parental *S. paradoxus* de los híbridos brasileños fue una mexicana aislada de ambientes naturales (YMX506E06); sin embargo, la distancia entre dicho padre y los híbridos brasileños es de 0.22 sustituciones por sitio, mientras que la distancia con su parental *S. cerevisiae* brasileño es de 0.03 sustituciones por sitio, sugiriendo que no contamos con la cepa parental *S. paradoxus* real de estos híbridos en el grupo de cepas empleado para estos análisis.

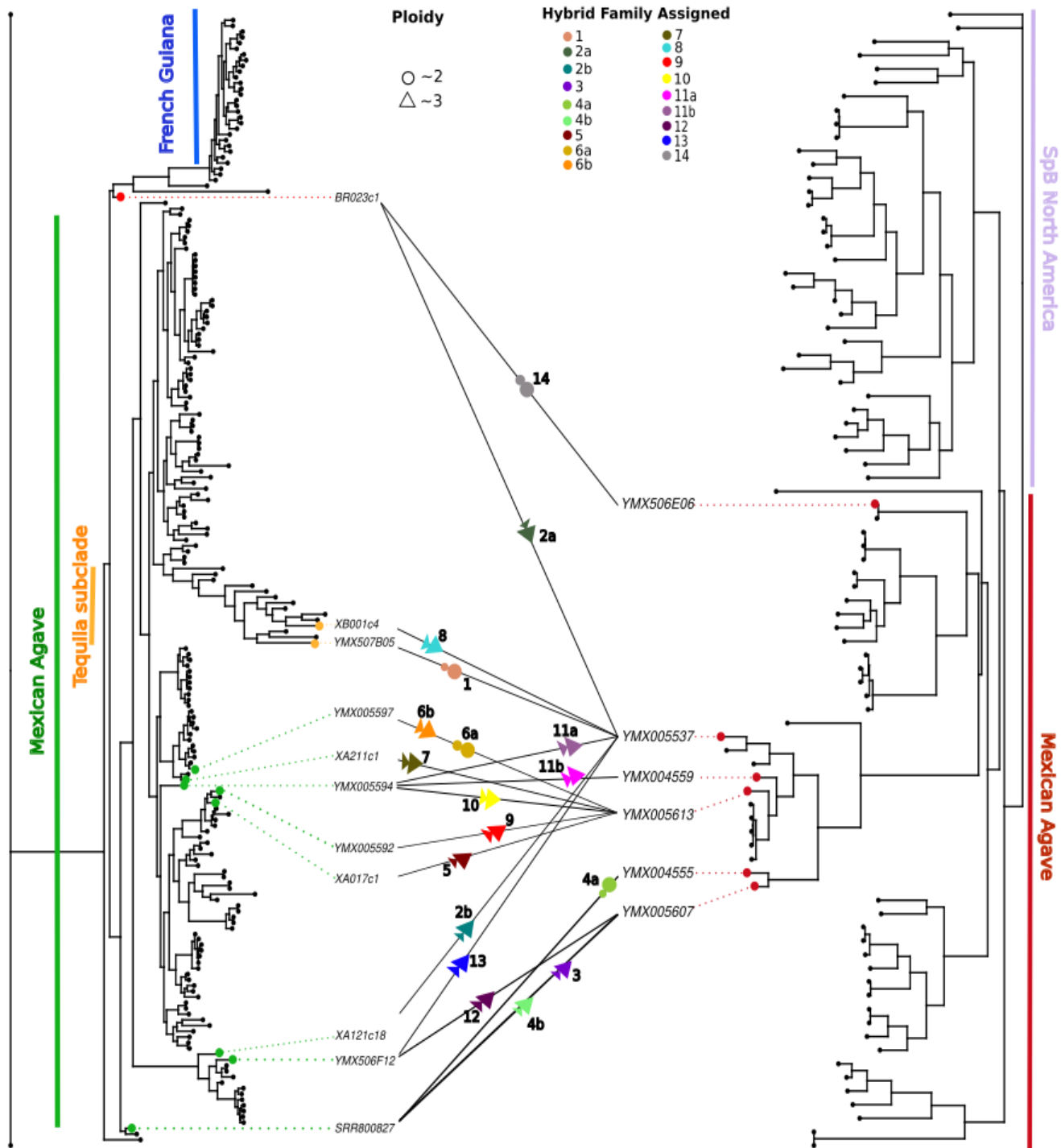


Figura 15. Co-filogenia de aislados parentales *S. cerevisiae* y *S. paradoxus* clasificados en familias.

Al igual que la figura anterior, cada línea que une dos nodos representa la cruce de una cepa *S. cerevisiae* con una *S. paradoxus*. Sobre la línea hay un círculo o triángulo dependiendo de la ploidía, coloreado dependiendo del grupo.

4.5 La proporción del contenido mitocondrial es similar dentro de los individuos de cada grupo

Para evaluar la similitud entre el contenido de ADN nuclear o mitocondrial procedente de uno de los parentales en los híbridos de un grupo, se construyó una gráfica de dispersión. Para facilitar la comparación, en esta gráfica se contrasta exclusivamente el contenido de ADN de *S. cerevisiae* en cada híbrido (Figura 16). Se puede observar que los integrantes de un mismo grupo comparten proporciones similares de contenido mitocondrial y nuclear.

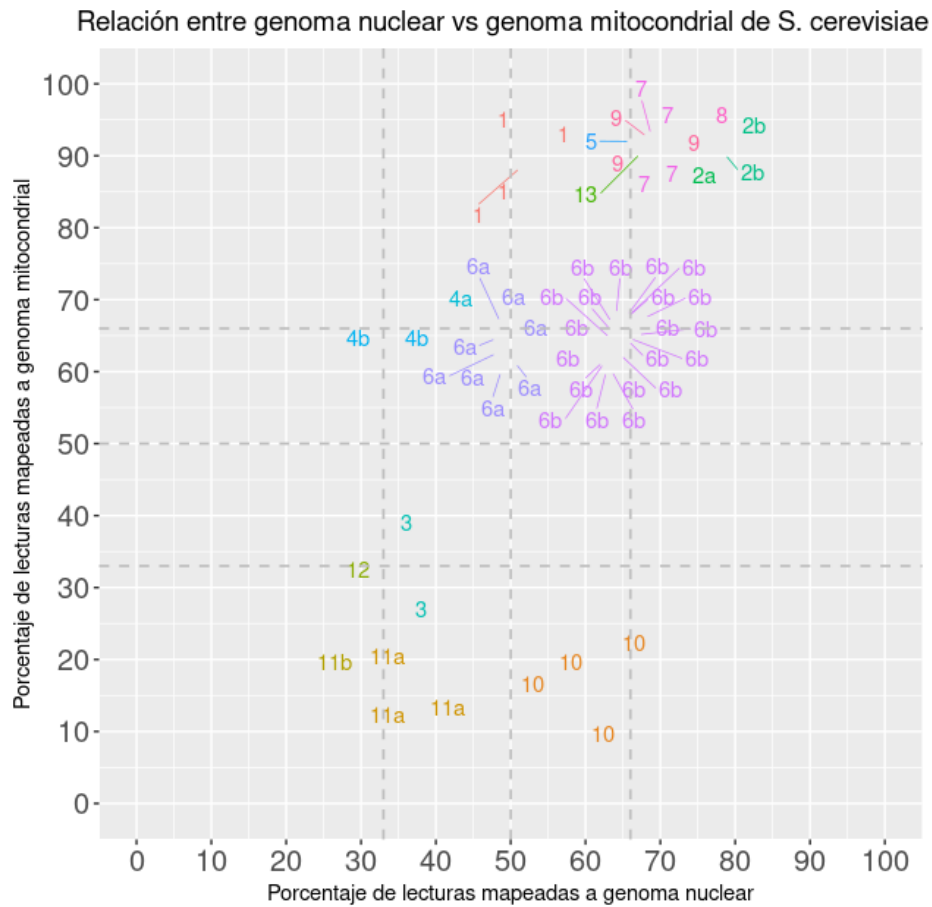


Figura 16. Gráfica de dispersión de porcentajes nucleares y mitocondriales.

El eje X indica el porcentaje de lecturas mapeadas al genoma de referencia nuclear *S. cerevisiae*, el eje Y el porcentaje de lecturas mitocondriales mapeadas al genoma mitocondrial de *S. cerevisiae* para cada híbrido. En ambos ejes se resaltan los porcentajes 30, 50 y 66.

Finalmente, se realizó una correlación de Spearman para tener una medida cuantitativa de este fenómeno (Figura 17). Se seleccionó aleatoriamente a un representante de cada grupo de híbridos y se utilizaron los porcentajes de contenido mitocondrial y nuclear para calcular la correlación con la paquetería estándar de R (versión 4.2.1).

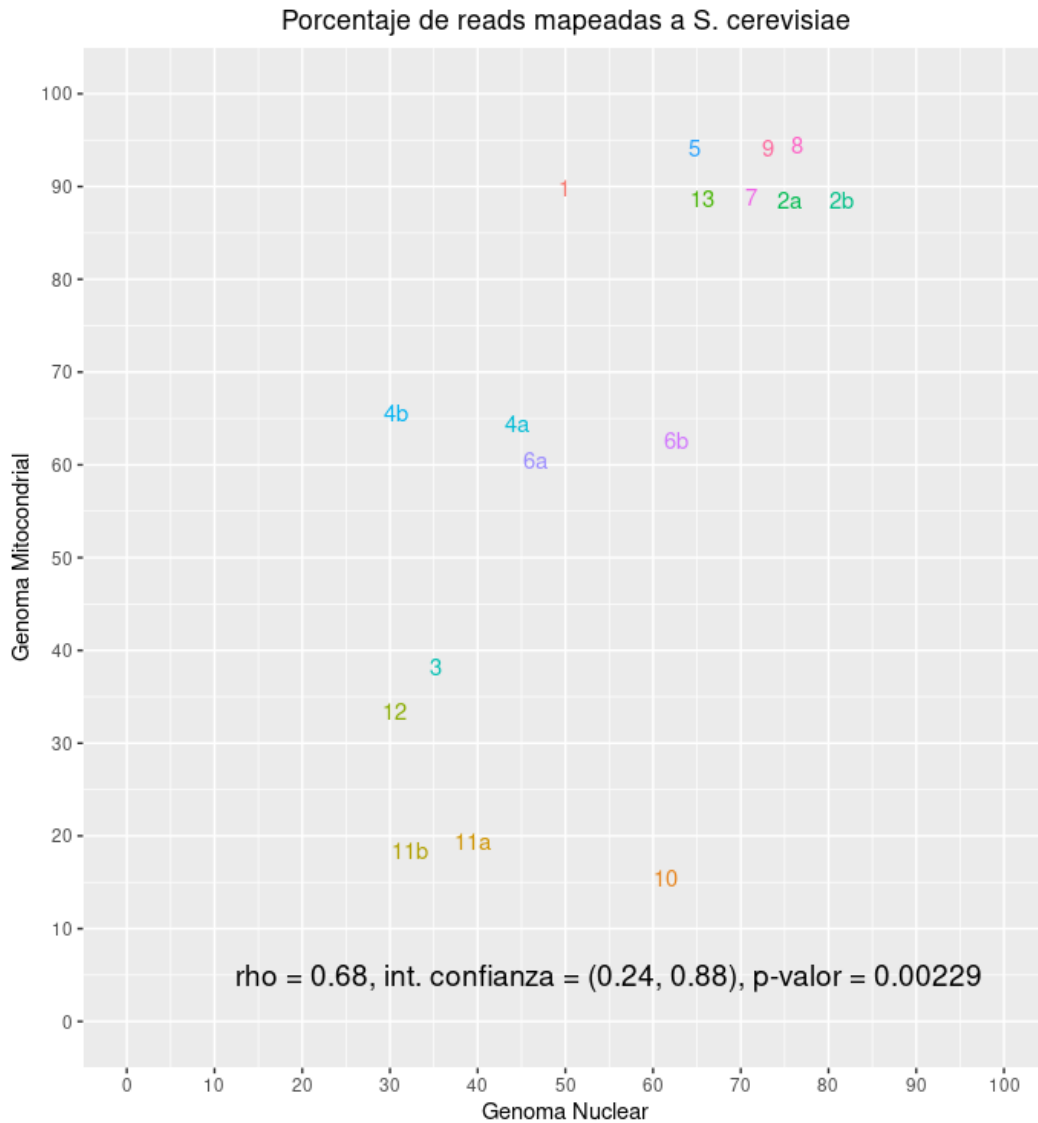


FIGURA 17. Correlación Spearman entre los porcentajes de contenido mitocondrial y nuclear

El eje X indica el porcentaje de lecturas mapeadas al genoma de referencia nuclear *S. cerevisiae*, el eje Y el porcentaje de lecturas mitocondriales mapeadas al genoma mitocondrial de *S. cerevisiae*.

De la gráfica se puede observar un coeficiente de correlación moderado ($\rho=0.65$), indicando una asociación entre el subgenoma nuclear prevalente con el subgenoma mitocondrial prevalente en las muestras que analizamos; sin embargo, con un tamaño de muestra pequeño, el cálculo de la magnitud de la correlación se vuelve inestable y es importante contar con una estimación de los posibles valores de su valor real. Para ello se calculó un intervalo de confianza de 95%, el cual abarca un rango amplio de valores (0.24 , 0.88), indicando la alta incertidumbre con la que contamos acerca del valor real del coeficiente de correlación a pesar de haber obtenido un valor significativo.

La Tabla 4 integra toda la información obtenida en este trabajo, en ella se despliegan las cepas identificadas como más cercanas a los híbridos, el número de integrantes, ploidía, contenidos nucleares y mitocondriales, y distancias filogenéticas a los parentales putativos de cada familia. Observaciones relevantes incluyen la cantidad de grupos triploides (14 de 18); el parental putativo *S. cerevisiae* YMX005594 es identificado como más cercano para tres grupos de híbridos y en ellos las proporciones mitocondriales son las mismas, a pesar de que cada uno de los tres grupos tiene como parental putativo a una cepa *S. paradoxus* diferente; la cepa *S. paradoxus* YMX005613 es la cepa parental más cercana a un tercio de los grupos híbridos, posiblemente debido al número relativamente bajo de *S. paradoxus* usadas en este análisis ($n = 84$).

Todos los grupos a excepción del grupo 14, que contiene a los híbridos brasileños, tienen distancias filogenéticas menores a 0.09 con las cepas *S. paradoxus*. Esto es interesante ya que el set de *S. paradoxus* usadas en este análisis ($n = 84$) es menos de la mitad del tamaño del set *S. cerevisiae* y aun así las distancias filogenéticas con las cepas *S. paradoxus* es generalmente menor.

Tabla 4. Resumen de los aislados parentales más cercanos a los híbridos.

En rojo se muestran las cepas más cercanas a los parentales *S. cerevisiae* y en azul las cepas más cercanas a los parentales *S. paradoxus*. En cada celda se indica el nombre del grupo de híbridos con esos parentales (negritas), el número de híbridos en ese grupo, la ploidía, la proporción del sub genoma nuclear (Nuc.) o mitocondrial (Mic.) *S. cerevisiae* (rojo) o *S. paradoxus* (azul) y la distancia filogenética del híbrido a cada parental putativo (Dist.). Esta tabla se divide en varias páginas por necesidad de presentación.

	YMX005607	YMX005613	YMX005537	YMX004559	YMX004555	YMX506E06
YMX005597		Grupo 6a Durango n=8 Ploidía - 2n Nuc. (49 : 51) Mit. (64 : 35) Dist.(0.03, 0.04)				
XA191c2		Grupo 6b Durango n=20 Ploidía - 3n Nuc. (64 : 34) Mit. (64 : 35) Dist.(0.02, 0.04)				
YMX005594		Grupo 10 Michoacán n=5 Ploidía - 3n Nuc. (60 : 39) Mit. (16 : 84) Dist.(0.07, 0.05)	Grupo 11b San Luis Potosí n=1 Ploidía - 3n Nuc. (30 : 67) Mit. (16 : 84) Dist. (0.10, 0.08)	Grupo 11a San Luis Potosí n=4 Ploidía - 3n Nuc. (37 : 61) Mit. (16 : 84) Dist. (0.09, 0.09)		
SRR800827	Grupo 3 San Luis Potosí n=2 Ploidía - 3n Nuc. (34 : 66) Mit. (33 : 67)				Grupo 4a Sonora n=1 Ploidía - 2n Nuc. (47 : 52) Mit. (67 : 33)	

	Dist. (0.18, 0.07)				Dist. (0.07, 0.04)	
	Grupo 4b Sonora n=2 Ploidía - 3n Nuc. (34 : 66) Mit. (68 : 32) Dist. (0.07, 0.04)					
YMX507B05			Grupo 1a Guanajuato n=4 Ploidía - 2n Nuc. (53 : 44) Mit. (90 : 10) Distancias (0.17, 0.07)			
XA211c1		Grupo 7 Durango n=4 Ploidía - 3n Nuc. (69 : 30) Mit. (91 : 9) Dist. (0.05, 0.07)				
YMX005592		Grupo 9 Edo. Mex. n=3 Ploidía - 3n Nuc. (70 : 27) Mit. (92 : 8) Dist. (0.02, 0.04)				
BR023c1			Grupo 2a Nuevo León n=1 Ploidía - 3n Nuc. (78 : 22) Mit. (91 : 9) Distancias (0.17, 0.06)			Grupo 14 Brasil n=2 Ploidía - 2n Nuc. (54 : 43) Mit. (30 : 70) Dist. (0.04, 0.22)

YMX506F12	<p>Grupo 12 Oaxaca n=1 Ploidía - 3n Nuc. (33 : 67) Mit. (36 : 64) Dist. (0.18, 0.08)</p>		<p>Grupo 13 Nuevo León n=1 Ploidía - 3n Nuc. (68 : 31) Mit. (91 : 9) Distancias (0.13, 0.05)</p>			
XA121c18			<p>Grupo 2b Nuevo León n=2 Ploidía - 3n Nuc. (78 : 21) Mit. (91 : 9) Distancias (0.17, 0.06)</p>			
XA017c1		<p>Grupo 5 Edo. Mex. n=1 Ploidía - 3n Nuc. (67 : 29) Mit. (92 : 8) Dist. (0.07, 0.04)</p>				
XB001c4			<p>Grupo 8 Nuevo León n=1 Ploidía - 3n Nuc. (78 : 21) Mit. (92 : 8) Distancias (0.14, 0.07)</p>			

5. Discusión

En este trabajo se desarrolló un recurso de consulta en línea de los híbridos aislados por el consorcio YeastGenomesMX que permite el acceso a sus metadatos y a gráficas de cobertura. Se identificaron a los aislados parentales más cercanos a cada uno de los híbridos por medio de filogenias de máxima verosimilitud y se analizó la cobertura de las mitocondrias de dichos híbridos. Con base en las cepas parentales identificadas, en conjunto con diferentes niveles de información de los híbridos (estado de aislamiento, ploidía, proporción de lecturas nucleares mapeadas, etc.), se clasificaron en 18 grupos y encontramos que sus mitocondrias pueden ser categorizadas en tres tipos: mayoritariamente *S. cerevisiae*, mayoritariamente *S. paradoxus* y recombinantes.

La presencia de híbridos inter especie en ambientes antropogénicos de fermentación es un fenómeno previamente reportado (D'angiolo, et al., 2019; Gibson y Liti, 2014), sin embargo, la diversidad de arquitecturas genómicas de los híbridos aislados en México no tiene precedentes reportados en la literatura, probablemente debido a la naturaleza del proceso de fermentación en la elaboración de mezcal y las estructuras poblacionales de *S. cerevisiae* y *S. paradoxus* nativas de México.

Para algunos híbridos, la distancia filogenética con las cepas identificadas como parentales es lo suficientemente alta para considerar que no contamos con los padres reales dentro de nuestra colección, como es el caso de las familias 1, 2a, 3, 11b, 12, y 13; las cuales presentan una distancia filogenética mayor a 0.09 sustituciones por sitio con la cepa *S. cerevisiae* más cercana. La única familia con un valor alto en la distancia filogenética con *S. paradoxus* es la familia 14, la que agrupa a los híbridos brasileños, el resto tiene distancias valores menores a 0.09.

En futuras iteraciones de este trabajo será relevante aumentar el conjunto de posibles cepas parentales empleado con el fin de identificar con mayor precisión a las cepas más cercanas a los padres de los híbridos. También será relevante contar con algún control para determinar la distancia filogenética esperada entre padres e hijos, por ejemplo, un

par de cepas cuya relación de parentesco haya sido previamente determinada. Dichas cepas pueden pertenecer a nuestra colección o al conjunto de cepas empleadas por D'angiolo et al. (2019).

Este trabajo sienta las bases para el análisis de híbridos interespecie encontrados en México y aporta evidencias que sugieren un escenario de múltiples eventos de hibridación: 1) las múltiples arquitecturas genómicas y diferentes proporciones de lecturas parentales se podrían explicar por medio de resoluciones diferentes a la inestabilidad genómica post-hibridación, en el caso de un solo evento de hibridación, esperaríamos arquitecturas genómicas similares entre los aislados híbridos; 2) la identificación de diferentes pares de cepas más cercanas a los parentales (Figuras 14 y 15) sugiere diversidad en el origen de los subgenomas presentes en los híbridos; 3) la heterogeneidad en los genomas mitocondriales y la presencia de genomas mitocondriales recombinantes es un escenario probablemente debido a múltiples eventos de hibridación ya que la resolución de la heteroplasma implica la fijación de solo un haplotipo, esto implica que el surgimiento de diversos haplotipos recombinantes SACE x SAPA es más plausible si ocurren más apareamientos entre las dos especies.

Para corroborar el escenario de hibridaciones múltiples sería útil comparar los patrones de pérdida de heterocigosidad de los integrantes de cada grupo, determinar el tamaño y direccionalidad de los tractos de pérdida de heterocigosidad para determinar si las familias se generaron por eventos de hibridación distintos.

Con respecto a los genomas mitocondriales, sería de interés analizarlos a nivel de secuencia y, ahora que sabemos quién es el parental más cercano de cada híbrido, determinar la aportación específica de cada padre. Además, identificar elementos que promuevan o inhiban la compatibilidad entre el genoma nuclear y mitocondrial.

Estos híbridos nos presentan una gran cantidad de preguntas por contestar no solo al respecto de su composición genómica, sino también acerca de su historia evolutiva, y dinámica poblacional con sus progenitores. ¿Resultan de un sólo evento de hibridación

en el tanque y se han reproducido asexualmente o dentro de los tanques ocurren hibridaciones recurrentes? En caso de ser múltiples eventos, ¿con qué frecuencia ocurren? ¿Los híbridos encontrados son fértiles? ¿Qué ventajas presentan los híbridos que les permiten coexistir con sus progenitores? ¿Cuáles son los elementos genómicos responsables? Ahora que contamos con una descripción preliminar de los híbridos aislados, se abren las puertas para poder interrogarlos de forma más específica.

6. Conclusiones

En este trabajo se definió un set de 62 híbridos *S. cerevisiae* - *S. paradoxus* aislados por el consorcio YeastGenomesMX (más 2 híbridos de Brasil) 1.- para los cuales se creó una base de datos relacional que permite contar con un recurso en línea fácilmente manipulable y accesible de la información de aislados híbridos. Esta base de datos es actualizable, así que fácilmente se pueden agregar más entradas a medida que la colección de aislados mexicanos de *Saccharomyces* aumente.

2.- Se llevó a cabo un análisis filogenético que permitió la identificación de algunas de las cepas más cercanas a los parentales de los híbridos aislados. En conjunto con la alta diversidad en composiciones genómicas de los híbridos, los resultados del análisis filogenético aportan evidencia de una historia de hibridación compleja con, posiblemente, múltiples eventos de hibridación y posterior diversificación.

3.- Finalmente, se realizó un análisis de la cobertura de los genomas mitocondriales híbridos y se clasificaron en tres categorías dependiendo de la proporción de lecturas que mapean a una referencia parental o a la otra. Se observó que los integrantes de un mismo grupo híbrido tienden a compartir una proporción muy similar de lecturas mapeadas a las referencias parentales y observamos evidencia de una correlación positiva entre el genoma nuclear prevalente con el genoma mitocondrial prevalente.

Este trabajo proporciona un recurso de consulta de aislados híbridos provenientes de tinajas de fermentación de agave, resalta la diversidad genómica de dichos aislados y provee evidencia de una historia compleja de hibridación.

7. Material Suplementario

Las tablas suplementarias 1 a 4 están divididas en varias páginas por necesidad de presentación.

Tabla Suplementaria 1. Códigos utilizados y/o desarrollados en este trabajo

Esta tabla muestra los programas empleados en este trabajo, cada renglón corresponde a un *script*. Los códigos que modifiqué o desarrollé específicamente para este trabajo se indican con una celda de título de *script* color verde.

Nombre del <i>script</i>	Acción	Entrada	Salida
<i>clean_fastq.py</i>	Control de calidad de archivos fastq	- Archivos fastq crudos: {ID}_R1.fastq.gz, {ID}_R2.fastq.gz	- Archivos fastq filtrados y procesados por control de calidad: {ID}_R1_clean.fastq.gz {ID}_R2_clean.fastq.gz
<i>mapping2ref.py</i>	Mapeo con bwa a una referencia especificada	- Lecturas procesadas por fastp: {ID}_R1_clean.fastq.gz {ID}_R2_clean.fastq.gz	- Archivo BAM ordenado por coordenadas: {ID}_CONC.rmdup.addgp.bam
<i>coverageplot.py</i>	Grafica la cobertura de un alineamiento a una referencia especificada	- Archivo BAM ordenado por coordenadas y fasta de un genoma de referencia: {ID}_CONC.rmdup.addgp.bam {Referencia}_allChr.fasta	- Imagen en archivo .png con la gráfica de cobertura: {ID}_{Referencia}_Q{###}.png

<p><i>vcalling_genotyping_isolates_GATK.py</i></p>	<p>Realiza llamado de variantes y genotipación por muestra con respecto a una referencia especificada</p>	<ul style="list-style-type: none"> - Archivo BAM ordenado por coordenadas y fasta de un genoma de referencia: {ID}_CONC.rmdup.addgp.bam {Referencia}_allChr.fasta 	<ul style="list-style-type: none"> - Archivo VCF y VCF comprimido {ID}_{Referencia}.g.vcf {ID}_{Referencia}.g.vcf.gz
<p><i>SNPs_onlyChr.py</i></p>	<p>Filtra aquellas variantes del subgenoma especificado y las almacena en un VCF</p>	<ul style="list-style-type: none"> - Archivo VCF con todas las variantes: {ID}_{Referencia}.g.vcf 	<ul style="list-style-type: none"> - Archivo VCF filtrado por SNPs provenientes sólo del subgenoma especificado: {ID}_{Referencia}.SNP_onlychr_{Subgenoma}.g.vcf
<p><i>genotype_and_runPhyloTree_ForOneHybrid.py</i></p>	<p>Une los archivos VCF de las muestras especificadas, hace genotipado conjunto y genera un árbol filogenético de máxima verosimilitud con el híbrido especificado</p>	<ul style="list-style-type: none"> - Archivos VCF y fasta de un genoma de referencia {ID}_{Referencia}.g.vcf {Referencia}_allChr.fasta 	<ul style="list-style-type: none"> - Archivo VCF conjunto filtrado por SNPs, alineamiento múltiple phylip y árboles de máxima verosimilitud: Matrix_SNPs_{Subgenoma}_from_{Referencia}_gt_onlySNPs_filtered_mising_10_plus_{Híbrido}.recode.vcf Matrix_SNPs_{Subgenoma}_from_{Referencia}_gt_onlySNPs_filtered_mising_10_plus_{Híbrido}.recode.min4.phy <p>RxML_bipartitionsBranchLabels.Ma</p>

				<p>trix_SNPs_{Subgenoma}_from_{Referencia}_gt_onlySNPs_filtered_missing_10_plus_{Híbrido}.recode_min4_AllHybSites.tree</p>
<i>vcf2phylipv28.py</i>	<p>Convierte un VCF a formato Phylip para utilizarse en la generación del árbol filogenético</p>	<p>- Archivo VCF conjunto filtrado por SNPs de un subgenoma: Matrix_SNPs_SAPA_from_CONC_gt_onlySNPs_filtered_missing_10_plus_YMX004544.recode.vcf</p>	<p>- Archivo phylip Matrix_SNPs_{Subgenoma}_from_{Referencia}_gt_onlySNPs_filtered_missing_10_plus_{Híbrido}.recode.min4.phy</p>	
<i>reduce_phylip.py</i>	<p>Remueve aquellas posiciones del alineamiento múltiple donde la muestra especificada no tiene información</p>	<p>- Archivo phylip: Matrix_SNPs_{Subgenoma}_from_{Referencia}_gt_onlySNPs_filtered_missing_10_plus_{Híbrido}.recode.min4.phy</p>	<p>- Archivo phylip filtrado Matrix_SNPs_{Subgenoma}_from_{Referencia}_gt_onlySNPs_filtered_missing_10_plus_{Híbrido}.recode.min4_AllHybSites.phy</p>	
<i>get_hyb_dists.py</i>	<p>Emplea la librería APE de R para calcular una matriz de distancias a partir de un archivo phylip</p>	<p>- Archivo phylip: Matrix_SNPs_{Subgenoma}_from_{Referencia}_gt_onlySNPs_filtered_missing_10_plus_{Híbrido}.recode.min4.phy</p>	<p>- Archivo CSV HybridParentals_{Referencia}.csv</p>	

<i>reference_percentagev4_onlyNuclearPercentages.py</i>	Obtiene el número de lecturas <u>nucleares</u> mapeadas a cada referencia, calcula sus proporciones y almacena esa información.	- Archivo CSV con los IDs de las muestras a procesar AllHybs_onlyIDs.csv	- Archivo CSV con los números de lecturas <u>mitocondriales</u> mapeadas, no mapeadas y porcentajes. RefPercentageSheet_63hybs.csv
<i>mit_reference_percentage3.py</i>	Obtiene el número de lecturas <u>mitocondriales</u> mapeadas a cada referencia, calcula sus proporciones, almacena y envía esa información al programa <i>mod_percentage_referencev2.R</i>	- Archivo CSV con los IDs de las muestras a procesar AllHybs_onlyIDs.csv	- Archivo CSV con los números de lecturas <u>mitocondriales</u> mapeadas, no mapeadas y porcentajes. MitRefPercentageSheet_63hybs.csv
<i>mit_percentage_reference2.R</i>	Utiliza la información de lecturas obtenida por el <i>script mod_percentage_referencev3.py</i> para hacer gráficas de barras representando las proporciones de lecturas mapeadas	Este programa emplea una tabla generada por el <i>script mod_percentage_referencev3.py</i> .	- Archivos PNG con gráficas de barras representando proporciones de lecturas mitocondriales mapeadas percentage_reference_mitDNA_Q20_###.png
<i>NuclearAndMitPlots.R</i>	Creación de gráficas lado a lado: una gráfica de barras con las proporciones de lecturas nucleares mapeadas y una gráfica de donas con las proporciones de lecturas mitocondriales	- Archivo CSV con los números de lecturas mapeadas, no mapeadas y porcentajes. RefPercentageSheet_63hybs.csv	- Archivos PNG con gráficas de barras y donas para cada híbrido. NuclearAndMitDNA_{ID}.png

<p><i>ScatterPlotMitrNuclear.R</i></p>	<p>mapeadas</p> <p>Realiza pruebas de normalidad en los datos de proporciones nucleares y mitocondriales, crea una gráfica de dispersión de los híbridos en función a sus proporciones de lecturas nucleares y mitocondriales de <i>S. cerevisiae</i> y corre un test de correlación de Spearman</p>	<p>- Archivo CSV con los números de lecturas mapeadas, no mapeadas y porcentajes.</p> <p>RefPercentageSheet_63hybs.csv</p>	<p>- Archivo PNG con gráfica de dispersión</p> <p>NuclearMitCorr.png</p>
--	--	--	--

Tabla Suplementaria 2. Aislados empleados en la filogenia *S. paradoxus*

84 cepas pertenecientes a los clados de agave mexicano y norteamericanos (SpB y SpC)

Nombre	Especie	Localización	Estado	Municipio	País
XA096c3	SAPA	Hacienda Santa Rosa de Ochoa	Guanajuato	San Luis de la Paz	México
XB077c1	SAPA				México
XB007c9	SAPA	Hacienda Vergel de Guadalupe	Guanajuato	San Luis de la Paz	México
YMX004555	SAPA	Leopoldo Guerrero	Michoacán	Los Lobos	México
YMX004559	SAPA	Hacienda Vergel de Guadalupe	Guanajuato	San Luis de la Paz	México
YMX004625	SAPA	El caramelo	Oaxaca	San Juan Bautista Jayacatlán	México
YMX005537	SAPA	Hacienda Vergel de Guadalupe	Guanajuato	San Luis de la Paz	México
YMX005582	SAPA	El diablo	Durango	Temohaya Mezquital	México
YMX005607	SAPA	Los Sausés	Sonora	Moctezuma	México
YMX005613	SAPA	Hacienda Vergel de Guadalupe	Guanajuato	San Luis de la Paz	México
YMX507A09	SAPA	LaCanoa02	NuevoLeón	Aramberri	México
YMX507A10	SAPA	LaCanoa02	NuevoLeón	Aramberri	México
YMX507A11	SAPA	LaCanoa	NuevoLeón	Aramberri	México
SRR7500266	SAPA	Pointe Platon, Québec, Canadá,	Québec, Canadá	Québec	Canadá
SRR7500208	SAPA	Pointe Platon, Québec, Canadá,	Québec, Canadá	Québec	Canadá

SRR7500202	SAPA	Pointe Platon, Québec, Canadá,	Québec, Canadá	Québec	Canadá
SRR7500200	SAPA	Pointe Platon, Québec, Canadá,	Québec, Canadá	Québec	Canadá
SRR7500201	SAPA	Pointe Platon, Québec, Canadá,	Québec, Canadá	Québec	Canadá
SRR7500205	SAPA	Riviere-Verte, New Brunswick, Canadá,	New Brunswick, Canadá	New Brunswick	Canadá
SRR7500252	SAPA	Lac Pink, Parc de la Gatineau, Québec, Canadá	Québec, Canadá	Québec	Canadá
SRR7500253	SAPA	Lac Pink, Parc de la Gatineau, Québec, Canadá	Québec, Canadá	Québec	Canadá
SRR7500249	SAPA	Lac Meech (Est), Parc de la Gatineau, Québec, Canadá	Québec, Canadá	Québec	Canadá
SRR7500248	SAPA	Lac Meech (Est), Parc de la Gatineau, Québec, Canadá	Québec, Canadá	Québec	Canadá
SRR7500247	SAPA	Lac Meech (Est), Parc de la Gatineau, Québec, Canadá	Québec, Canadá	Québec	Canadá
SRR7500246	SAPA	Lac Meech (Est), Parc de la Gatineau, Québec, Canadá	Québec, Canadá	Québec	Canadá
SRR7500256	SAPA	Lac Meech (Est), Parc de la Gatineau, Québec, Canadá	Québec, Canadá	Québec	Canadá

SRR7500257	SAPA	Lac Meech (Est), Parc de la Gatineau, Québec, Canadá	Québec, Canadá	Québec	Canadá
SRR7500259	SAPA	Lac Meech (Est), Parc de la Gatineau, Québec, Canadá	Québec, Canadá	Québec	Canadá
SRR7500260	SAPA	Lac Meech (Est), Parc de la Gatineau, Québec, Canadá	Québec, Canadá	Québec	Canadá
SRR7500263	SAPA	Sucrerie, Parc de la Gatineau, Québec, Canadá	Québec, Canadá	Québec	Canadá
SRR7500258	SAPA	Sucrerie, Parc de la Gatineau, Québec, Canadá	Québec, Canadá	Québec	Canadá
SRR7500228	SAPA	Sucrerie, Parc de la Gatineau, Québec, Canadá	Québec, Canadá	Québec	Canadá
SRR7500229	SAPA	University of Toronto Mississauga, Ontario, Canadá,	Ontario, Canadá	Ontario	Canadá
SRR7500232	SAPA	University of Toronto Mississauga, Ontario, Canadá,	Ontario, Canadá	Ontario	Canadá
SRR7500240	SAPA	Riverwood Park, Mississauga, Ontario, Canadá	Ontario, Canadá	Ontario	Canadá
SRR7500243	SAPA	Afton, Michigan, EE. UU.,	Michigan, EE. UU.	Michigan	EE. UU.
SRR7500244	SAPA	Afton, Michigan, EE. UU.,	Michigan, EE. UU.	Michigan	EE. UU.

SRR7500241	SAPA	Afton, Michigan, EE. UU.,	Michigan, EE. UU.	Michigan	EE. UU.
SRR7500242	SAPA	Afton, Michigan, EE. UU.,	Michigan, EE. UU.	Michigan	EE. UU.
SRR7500235	SAPA	Afton, Michigan, EE. UU.,	Michigan, EE. UU.	Michigan	EE. UU.
SRR7500236	SAPA	Afton, Michigan, EE. UU.,	Michigan, EE. UU.	Michigan	EE. UU.
SRR7500209	SAPA	Afton, Michigan, EE. UU.,	Michigan, EE. UU.	Michigan	EE. UU.
SRR7500206	SAPA	Afton, Michigan, EE. UU.,	Michigan, EE. UU.	Michigan	EE. UU.
SRR7500213	SAPA	Afton, Michigan, EE. UU.,	Michigan, EE. UU.	Michigan	EE. UU.
SRR7500218	SAPA	Cornell, New York, EE. UU.,	New York, EE. UU.	New York	EE. UU.
SRR7500222	SAPA	Cornell, New York, EE. UU.,	New York, EE. UU.	New York	EE. UU.
SRR1868751	SAPA	Mont St Hilaire, Québec, Canadá,	Québec, Canadá	Québec	Canadá
SRR1868666	SAPA	St Michel de Bellechasse, Québec, Canadá,	Québec, Canadá	Québec	Canadá
SRR1868728	SAPA	Oak Point Provincial Parc, New Brunswick, Canadá,	New Brunswick, Canadá	New Brunswick	Canadá
SRR1868780	SAPA	London, Ontario, Canadá,	Ontario, Canadá	Ontario	Canadá
SRR4074411	SAPA	Pennsylvania, EE. UU.	Pennsylvania, EE. UU.	Pennsylvania	EE. UU.
SRR4074412	SAPA	Tijuca forest, Rio de Janeiro, Brasil,	Rio de Janeiro, Brasil	Rio de Janeiro	Brasil

SRR1868775	SAPA	Canadá	Canadá		Canadá
SRR4453439	SAPA	University of Toronto Mississauga	Ontario, Canadá	Ontario	Canadá
SRR4453454	SAPA	University of Toronto Mississauga	Ontario, Canadá	Ontario	Canadá
SRR4453542	SAPA	University of Toronto Mississauga	Ontario, Canadá	Ontario	Canadá
SRR4453416	SAPA	University of Toronto Mississauga	Ontario, Canadá	Ontario	Canadá
YMX506B01	SAPA	NuevePuntas3	Oaxaca	Santiago Matatlan	México
YMX506A12	SAPA	NuevePuntas3	Oaxaca	Santiago Matatlan	México
YMX506B04	SAPA	NuevePuntas3	Oaxaca	Santiago Matatlan	México
YMX506B07	SAPA	NiDain03	Oaxaca	Santiago Matatlan	México
YMX506B12	SAPA	MatatlanEste03	Oaxaca	Santiago Matatlan	México
YMX506C01	SAPA	MatatlanEste03	Oaxaca	Santiago Matatlan	México
YMX506D01	SAPA	Vinata malpais	Durango	Nombre de Dios	México
YMX506D02	SAPA	Vinata malpais	Durango	Nombre de Dios	México
YMX506D04	SAPA	EncinaGorda01	Durango	La Michilia	México
YMX506D10	SAPA	EncinaGorda01	Durango	La Michilia	México
YMX506D12	SAPA	EncinaGorda_Camino	Durango	La Michilia	México
YMX506E01	SAPA	EncinaGorda_Camino	Durango	La Michilia	México
YMX506E05	SAPA	EIPicacho1	Durango	Nombre de Dios	México
YMX506E06	SAPA	EIPicacho1	Durango	Nombre de Dios	México
YMX506F01	SAPA	DeSoto02	NuevoLeon	Aramberri	México
YMX506E12	SAPA	DeSoto02	NuevoLeon	Aramberri	México

YMX506F02	SAPA	EIDiente03	Tamaulipas	San Carlos	México
YMX506F05	SAPA	EIDiente02	Tamaulipas	San Carlos	México
YMX506F08	SAPA	LaCanoa02	NuevoLeon	Aramberri	México
YMX506F07	SAPA	LaCanoa02	NuevoLeon	Aramberri	México
YMX506F11	SAPA	LaCanoa02	NuevoLeon	Aramberri	México
YMX506G04	SAPA	LaCanoa01	NuevoLeon	Aramberri	México
YMX506G03	SAPA	LaCanoa01	NuevoLeon	Aramberri	México
YMX506G12	SAPA	EIDiente01	Tamaulipas	San Carlos	México
YMX506H01	SAPA	EIDiente01	Tamaulipas	San Carlos	México
YMX506G09	SAPA	EIDiente03	Tamaulipas	San Carlos	México
YMX506G10	SAPA	EIDiente03	Tamaulipas	San Carlos	México
YMX506G11	SAPA	EIDiente01	Tamaulipas	San Carlos	México
YMX506G08	SAPA	EIDiente01	Tamaulipas	San Carlos	México
YMX506H04	SAPA	EIDiente04	Tamaulipas	San Carlos	México
YMX506H02	SAPA	EIDiente04	Tamaulipas	San Carlos	México

Tabla Suplementaria 3. Aislados empleados en la filogenia *S. cerevisiae*

Ciento noventa y nueve cepas de los clados de agave mexicano, sudamericano y Guayana francesa.

Nombre	Especie	Localización	País
DK001c27	SACE	Santiago Matatlán; Oaxaca; México	México
DK003c56	SACE	Laguna Seca; San Luis Potosí; México	México
DK016c5	SACE	Eduardo Neri; Guerrero; México	México
DK017c4	SACE	Chilapa; Guerrero; México	México
DK018c4	SACE	Huitzucó; Guerrero; México	México
DK019c1	SACE	Charo; Michoacán; México	México
XA017c1	SACE	Palmar de Guadalupe; Estado de México; México	México
XA023c1	SACE	Palmar de Guadalupe; Estado de México; México	México
XA030c1	SACE	San José Chalmita; Estado de México; México	México
XA031c2	SACE	San José Chalmita; Estado de México; México	México
XA057c4	SACE	San Isidro Gusihe; Oaxaca; México	México
XA070c1	SACE	San Juan Bautista Jayacatlán; Oaxaca; México	México
XA072c4	SACE	San Juan Bautista Jayacatlán; Oaxaca; México	México
XA074c2	SACE	Santa María Ixcatlán; Oaxaca; México	México
XA081c6	SACE	San Felipe; Guanajuato; México	México
XA083c1	SACE	San Felipe; Guanajuato; México	México
XA083c6	SACE	San Felipe; Guanajuato; México	México
XA097c15	SACE	San Luis de la Paz; Guanajuato; México	México
XA097c4	SACE	San Luis de la Paz; Guanajuato; México	México
XA098c2	SACE	San Felipe; Guanajuato; México	México
XA121c14	SACE	San Carlos; Tamaulipas; México	México
XA121c18	SACE	San Carlos; Tamaulipas; México	México
XA121c20	SACE	San Carlos; Tamaulipas; México	México

XA121c6	SACE	San Carlos; Tamaulipas; México	México
XA126c7	SACE	San Carlos; Tamaulipas; México	México
XA132c2	SACE	San Nicolas Huajuapán; Puebla; México	México
XA133c5	SACE	San Nicolas Huajuapán; Puebla; México	México
XA134c1	SACE	San Nicolas Huajuapán; Puebla; México	México
XA142c1	SACE	San Juan del Río; Oaxaca; México	México
XA143c7	SACE	San Luis Amatlan; Oaxaca; México	México
XA144c3	SACE	San Luis Amatlan; Oaxaca; México	México
XA145c3	SACE	San Luis Amatlan; Oaxaca; México	México
XA146c1	SACE	San Agustín Amatengo; Oaxaca; México	México
XA154c1	SACE	Mascota; Jalisco; México	México
XA186c3	SACE	Nombre de Dios; Durango; México	México
XA189c1	SACE	Nombre de Dios; Durango; México	México
XA190c1	SACE	Nombre de Dios; Durango; México	México
XA191c2	SACE	Nombre de Dios; Durango; México	México
XA196c1	SACE	Yonora Mezquitlan; Durango; México	México
XA197c1	SACE	Yonora Mezquitlan; Durango; México	México
XA200c2	SACE	Temohaya Mezquitlan; Durango; México	México
XA211c1	SACE	Huitzila; Zacatecas; México	México
XA251c2	SACE	Santiago Matatlan; Oaxaca; México	México
XA252c3	SACE	Santiago Matatlan; Oaxaca; México	México
XA254c23	SACE	Piedras de lumbre; Michoacán; México	México
XA255c1	SACE	Santiago Matatlan; Oaxaca; México	México
XA258c4	SACE	Santiago Matatlan; Oaxaca; México	México
XA261c23	SACE	Santiago Matatlan; Oaxaca; México	México
XB001c4	SACE	Santiago Matatlan; Oaxaca; México	México
XB011c1	SACE	Etucuaró; Michoacán; México	México
XB041c1	SACE	Santa Catarina de Albarradas; Oaxaca; México	México
XB044c4	SACE	Santiago Matatlan; Oaxaca; México	México

XB045c4	SACE	Santiago Matatlan; Oaxaca; México	México
XB047c2	SACE	Santiago Matatlan; Oaxaca; México	México
XB048c1	SACE	Santiago Matatlan; Oaxaca; México	México
XB050c1	SACE	Santiago Matatlan; Oaxaca; México	México
XB051c4	SACE	Santiago Matatlan; Oaxaca; México	México
XB052c1	SACE	Santiago Matatlan; Oaxaca; México	México
XB053c1	SACE	Santiago Matatlan; Oaxaca; México	México
XB054c5	SACE	Santiago Matatlan; Oaxaca; México	México
XB056c1	SACE	Miahuatlan; Oaxaca; México	México
XB064c2	SACE	Santa Catarina de Minas; Oaxaca; México	México
XB065c1	SACE	Santa Catarina de Minas; Oaxaca; México	México
XB076c1	SACE	Apetlanca; Guerrero; México	México
XB094c2	SACE	San Juan Espanatica; Jalisco; México	México
XB251c1	SACE	Morelia Pino Bonito; Michoacan; México	México
XB252c1	SACE	Morelia Pino Bonito; Michoacan; México	México
XB254c1	SACE	Piedras de lumbre; Michoacan; México	México
XB256c3	SACE	Santiago Matatlan; Oaxaca; México	México
XB258c4	SACE	Santiago Matatlan; Oaxaca; México	México
DK001c10	SACE	Santiago Matatlan; Oaxaca; México	México
XA024c2	SACE	Palmar de Guadalupe; Estado de México; México	México
XA137c1	SACE	San Juan del Rio; Oaxaca; México	México
XA139c1	SACE	San Juan del Rio; Oaxaca; México	México
XA140c2	SACE	San Juan del Rio; Oaxaca; México	México
XA195c1	SACE	Yonora Mezquitil; Durango; México	México
YMX004566	SACE	Etucuario; Michoacan; México	México
YMX004569	SACE	San Juan Bautista Jayacatlan; Oaxaca; México	México
YMX004648	SACE	San Agustin Amatengo; Oaxaca; México	México
YMX004659	SACE	La Estancia; Jalisco; México	México
YMX004668	SACE	La Estancia; Jalisco; México	México

YMX004685	SACE	Santa Maria Ixcatlan; Oaxaca; México	México
YMX004695	SACE	San Pedro Totolapan; Oaxaca; México	México
YMX005538	SACE	San Juan Bautista Jayacatlan; Oaxaca; México	México
YMX005539	SACE	Santa Catarina de Minas; Oaxaca; México	México
YMX005541	SACE	Santa Catarina de Albarradas; Oaxaca; México	México
YMX005542	SACE	Miahuatlan; Oaxaca; México	México
YMX005545	SACE	San Juan Espanatica; Jalisco; México	México
YMX005546	SACE	San Juan Espanatica; Jalisco; México	México
YMX005547	SACE	Santiago Matatlan; Oaxaca; México	México
YMX005548	SACE	Santiago Matatlan; Oaxaca; México	México
YMX005549	SACE	Santa Catarina de Minas; Oaxaca; México	México
YMX005550	SACE	Santiago Matatlan; Oaxaca; México	México
YMX005551	SACE	Piedras de lumbre; Michoacan; México	México
YMX005552	SACE	Piedras de lumbre; Michoacan; México	México
YMX005553	SACE	Santiago Matatlan; Oaxaca; México	México
YMX005554	SACE	Santa Catarina de Albarradas; Oaxaca; México	México
YMX005555	SACE	Morelia Pino Bonito; Michoacan; México	México
YMX005556	SACE	Morelia Pino Bonito; Michoacan; México	México
YMX005557	SACE	Miahuatlan; Oaxaca; México	México
YMX005558	SACE	Santiago Matatlan; Oaxaca; México	México
YMX005559	SACE	Etucuario; Michoacan; México	México
YMX005561	SACE	Apetlanca; Guerrero; México	México
YMX005563	SACE	Santa Maria Ixcatlan; Oaxaca; México	México
YMX005564	SACE	Santa Maria Ixcatlan; Oaxaca; México	México
YMX005565	SACE	San Carlos; Tamaulipas; México	México
YMX005566	SACE	San Carlos; Tamaulipas; México	México
YMX005569	SACE	San Agustin Amatengo; Oaxaca; México	México
YMX005570	SACE	San Agustin Amatengo; Oaxaca; México	México
YMX005571	SACE	San Juan del Rio; Oaxaca; México	México

YMX005572	SACE	San Luis Amatlan; Oaxaca; México	México
YMX005575	SACE	San Nicolas Huajuapán; Puebla; México	México
YMX005576	SACE	Temohaya Mezquital; Durango; México	México
YMX005577	SACE	Yonora Mezquital; Durango; México	México
YMX005583	SACE	Temohaya Mezquital; Durango; México	México
YMX005584	SACE	Palmar de Guadalupe; Estado de México; México	México
YMX005585	SACE	Mascota; Jalisco; México	México
YMX005586	SACE	La Estancia; Jalisco; México	México
YMX005587	SACE	San Carlos; Tamaulipas; México	México
YMX005588	SACE	San Carlos; Tamaulipas; México	México
YMX005590	SACE	Piedra Parada; Estado de México; México	México
YMX005591	SACE	Mascota; Jalisco; México	México
YMX005592	SACE	San José Chalmita; Estado de México; México	México
YMX005593	SACE	San José Chalmita; Estado de México; México	México
YMX005594	SACE	Las Guasimas Cabo Corrientes; Jalisco; México	México
YMX005595	SACE	San Juan del Rio; Oaxaca; México	México
YMX005596	SACE	San Juan del Rio; Oaxaca; México	México
YMX005597	SACE	Nombre de Dios; Durango; México	México
YMX005600	SACE	Bacanora; Sonora; México	México
YMX005601	SACE	Bacanora; Sonora; México	México
YMX005603	SACE	Bacanora; Sonora; México	México
YMX005609	SACE	Puerta del Sol; Sonora; México	México
YMX005616	SACE	Mazatlan; Guerrero; México	México
YMX005618	SACE	Mazatlan; Guerrero; México	México
YMX005619	SACE	San Luis de la Paz; Guanajuato; México	México
YMX005620	SACE	San Pedro Ures; Sonora; México	México
YMX005621	SACE	San Pedro Ures; Sonora; México	México
YMX005622	SACE	Bacanora; Sonora; México	México
YMX005626	SACE	Mazocahui; Sonora; México	México

YMX005628	SACE	Baviacora; Sonora; México	México
YMX005630	SACE	Matope Villa Pesqueira; Sonora; México	México
YMX005632	SACE	Pueblo de Alamos; Sonora; México	México
YMX005633	SACE	Moctezuma; Sonora; México	México
YMX005639	SACE	Tototlan; Jalisco; México	México
YMX005642	SACE	Tototlan; Jalisco; México	México
YMX005645	SACE	Ciénega; Jalisco; México	México
YMX506H07	SACE	Santa Catarina de Minas; Oaxaca; México	México
YMX506H08	SACE	Santiago Matatlán; Oaxaca; México	México
YMX506H09	SACE	Santiago Matatlán; Oaxaca; México	México
YMX507A04	SACE	San Carlos; Tamaulipas; México	México
YMX507A05	SACE	San Nicolás; Tamaulipas; México	México
YMX507A06	SACE	San Carlos; Tamaulipas; México	México
YMX507B01	SACE	Amatitan; Jalisco; México	México
YMX507B02	SACE	Amatitan; Jalisco; México	México
YMX507B03	SACE	Amatitan; Jalisco; México	México
YMX507B04	SACE	Amatitan; Jalisco; México	México
YMX507B05	SACE	Amatitan; Jalisco; México	México
YMX507B06	SACE	Amatitan; Jalisco; México	México
YMX507B07	SACE	Amatitan; Jalisco; México	México
YMX506F12	SACE	San Nicolás; Tamaulipas; México	México
JS109c1	SACE	México	México
JS497c1	SACE	Jalisco; México	México
JS326c1	SACE	Guayana Francesa	Guayana Francesa
JS210c1	SACE	Guayana Francesa	Guayana Francesa
JS818c1	SACE	Guayana Francesa	Guayana Francesa

JS835c1	SACE	Guayana Francesa	Guayana Francesa
JS171c1	SACE	Guayana Francesa	Guayana Francesa
JS208c1	SACE	Guayana Francesa	Guayana Francesa
JS014c1	SACE	Guayana Francesa	Guayana Francesa
JS889c1	SACE	Guayana Francesa	Guayana Francesa
JS654c1	SACE	Guayana Francesa	Guayana Francesa
JS150c1	SACE	Guayana Francesa	Guayana Francesa
JS933c1	SACE	Guayana Francesa	Guayana Francesa
JS440c1	SACE	Guayana Francesa	Guayana Francesa
JS387c1	SACE	Guayana Francesa	Guayana Francesa
JS922c1	SACE	Guayana Francesa	Guayana Francesa
JS114c1	SACE	Guayana Francesa	Guayana Francesa
JS277c1	SACE	Guayana Francesa	Guayana Francesa
JS861c1	SACE	Guayana Francesa	Guayana Francesa
JS691c1	SACE	Guayana Francesa	Guayana Francesa
JS468c1	SACE	Guayana Francesa	Guayana Francesa

JS263c1	SACE	Guayana Francesa	Guayana Francesa
JS599c1	SACE	Guayana Francesa	Guayana Francesa
JS527c1	SACE	Guayana Francesa	Guayana Francesa
JS310c1	SACE	Guayana Francesa	Guayana Francesa
JS311c1	SACE	Guayana Francesa	Guayana Francesa
JS759c1	SACE	Guayana Francesa	Guayana Francesa
JS165c1	SACE	Guayana Francesa	Guayana Francesa
JS728c1	SACE	Guayana Francesa	Guayana Francesa
JS246c1	SACE	Guayana Francesa	Guayana Francesa
JS270c1	SACE	Guayana Francesa	Guayana Francesa
JS754c1	SACE	Guayana Francesa	Guayana Francesa
ERR1309514	SACE	Yasuni National Park; Orellana; Ecuador	Ecuador
ERR1309020	SACE	Aba; Abia State; Nigeria	Nigeria
BR021c1	SACE	Riparian forest of Córrego; Buritizal; Tocantins; Brasil	Brasil
BR023c1	SACE	Research Center Canguçu; T. da Praia; Tocantins; Brasil	Brasil
BR010c1	SACE	Taguatinga; Tocantins; Brasil	Brasil
BR011c1	SACE	Taguatinga; Tocantins; Brasil	Brasil
SRR800827	SACE	No reportado	No reportado

Tabla Suplementaria 4. Aislados parentales más cercanos a los híbridos

En esta tabla se despliegan, para cada híbrido, los IDs de las cepas identificadas como más cercanas a los parentales junto con la distancia filogenética a su híbrido más cercano, el número de SNPs usados en la filogenia y número de SNPs compartidos con su híbrido más cercano. La información de la cepa más cercana al parental *S. cerevisiae* se resalta en rojo y naranja, la de *S. paradoxus* en azul; la intensidad del color en la columna de distancias indica la magnitud del valor. Los híbridos con grupo NatHyb corresponden a dos híbridos aislados de ambientes naturales en México y todavía no se cuenta con la información suficiente para clasificarlos bajo los criterios utilizados en este trabajo.

Grupo híbrido	ID Híbrido	<i>S. cerevisiae</i> más cercano	Distancia filogenética (sust.p/sitio)	SNPs usados en la filogenia	SNPs compartidos HíbridoPadre	<i>S. paradoxus</i> más cercano	Distancia filogenética(s) ust.p/sitio)	SNPs usados en la filogenia	SNPs compartidos HíbridoPadre
1	XB008c4	YMX507B05	0,17	3,42E+05	3,13E+05	YMX005537	0,07	2,21E+05	2,08E+05
1	YMX004605	YMX507B05	0,17	3,52E+05	3,22E+05	YMX005537	0,07	2,23E+05	2,10E+05
1	YMX004626	YMX507B05	0,17	3,42E+05	3,13E+05	YMX005537	0,07	2,21E+05	2,08E+05
1	YMX005562	YMX507B05	0,17	3,42E+05	3,13E+05	YMX005537	0,07	2,21E+05	2,08E+05
3	YMX004702	SRR800827	0,18	2,99E+05	2,75E+05	YMX005607	0,07	2,18E+05	2,05E+05
3	DK003c30	SRR800827	0,18	2,99E+05	2,75E+05	YMX005607	0,07	2,18E+05	2,05E+05
5	YMX005580	XA017c1	0,02	3,57E+05	3,50E+05	YMX005613	0,04	1,98E+05	1,91E+05
7	YMX005578	XA211c1	0,05	3,40E+05	3,31E+05	YMX005613	0,07	1,58E+05	1,48E+05
7	YMX506C10	XA211c1	0,05	3,32E+05	3,24E+05	YMX005613	0,06	1,48E+05	1,40E+05
7	YMX506H10	XA211c1	0,05	3,39E+05	3,30E+05	YMX005613	0,07	1,58E+05	1,50E+05
7	YMX507A01	XA211c1	0,05	3,32E+05	3,23E+05	YMX005613	0,06	1,48E+05	1,40E+05
8	YMX506G07	XB001c4	0,14	3,55E+05	3,27E+05	YMX005537	0,07	9,71E+04	9,16E+04

9	YMX004663	YMX005592	0,02	3,59E+05	3,53E+05	YMX005613	0,04	1,91E+05	1,83E+05
9	YMX005581	YMX005592	0,02	3,53E+05	3,47E+05	YMX005613	0,04	2,15E+05	2,06E+05
9	XA020c1	YMX005592	0,02	3,59E+05	3,53E+05	YMX005613	0,04	1,90E+05	1,83E+05
10	YMX004544	YMX005594	0,07	3,42E+05	3,25E+05	YMX005613	0,05	1,77E+05	1,69E+05
10	YMX004587	YMX005594	0,07	3,33E+05	3,16E+05	YMX005613	0,05	1,75E+05	1,67E+05
10	XA256c15	YMX005594	0,07	3,34E+05	3,18E+05	YMX005613	0,05	1,74E+05	1,67E+05
10	XB003c8	YMX005594	0,07	3,45E+05	3,28E+05	YMX005613	0,05	1,77E+05	1,69E+05
10	XB255c12	YMX005594	0,07	3,42E+05	3,25E+05	YMX005613	0,05	1,76E+05	1,68E+05
12	DK002c39	YMX506F12	0,18	2,95E+05	2,70E+05	YMX005607	0,08	2,16E+05	2,03E+05
13	YMX506F09	YMX506F12	0,13	3,49E+05	3,24E+05	YMX005537	0,05	1,73E+05	1,64E+05
14	BR017c1	BR023c1	0,04	2,29E+05	2,25E+05	YMX506E06	0,20	1,33E+05	1,13E+05
14	BR018c1	BR023c1	0,03	3,42E+05	3,37E+05	YMX506E06	0,22	2,07E+05	1,75E+05
11a	XB003s8	YMX005594	0,09	2,69E+05	2,54E+05	YMX004559	0,09	2,06E+05	1,90E+05
11a	YMX004548	YMX005594	0,09	2,73E+05	2,58E+05	YMX004559	0,09	2,08E+05	1,91E+05
11a	YMX004593	YMX005594	0,09	2,70E+05	2,55E+05	YMX004559	0,09	2,08E+05	1,91E+05
11a	YMX005560	YMX005594	0,09	2,70E+05	2,55E+05	YMX004559	0,09	2,08E+05	1,91E+05
11b	YMX005536	YMX005594	0,10	2,49E+05	2,35E+05	YMX005537	0,08	2,16E+05	2,02E+05
2a	YMX506F06	BR023c1	0,17	3,55E+05	3,28E+05	YMX005537	0,06	9,92E+04	9,37E+04
2b	YMX506G01	XA121c18	0,17	3,56E+05	3,16E+05	YMX005537	0,06	9,78E+04	9,26E+04
2b	YMX507A08	XA121c18	0,17	3,56E+05	3,16E+05	YMX005537	0,06	9,56E+04	9,02E+04
4a	YMX005605	SRR800827	0,07	3,11E+05	2,98E+05	YMX004555	0,04	2,10E+05	2,02E+05
4b	YMX005623	SRR800827	0,07	3,22E+05	3,08E+05	YMX005607	0,05	2,27E+05	2,18E+05
4b	YMX005624	SRR800827	0,07	3,05E+05	2,92E+05	YMX005607	0,04	2,30E+05	2,20E+05
6a	XA192c4	YMX005597	0,03	3,47E+05	3,42E+05	YMX005613	0,04	2,19E+05	2,10E+05
6a	YMX004655	YMX005597	0,03	3,47E+05	3,42E+05	YMX005613	0,04	2,20E+05	2,10E+05

6a	YMX005579	YMX005597	0,03	3,44E+05	3,39E+05	YMX005613	0,04	2,19E+05	2,09E+05
6a	YMX506C05	YMX005597	0,03	3,42E+05	3,38E+05	YMX005613	0,04	2,16E+05	2,07E+05
6a	YMX506C06	YMX005597	0,02	3,46E+05	3,41E+05	YMX005613	0,05	2,19E+05	2,09E+05
6a	YMX506C08	YMX005597	0,02	3,45E+05	3,40E+05	YMX005613	0,04	2,20E+05	2,10E+05
6a	YMX506D05	YMX005597	0,03	3,45E+05	3,40E+05	YMX005613	0,04	2,17E+05	2,08E+05
6a	YMX506D06	YMX005597	0,03	3,43E+05	3,38E+05	YMX005613	0,04	2,19E+05	2,09E+05
6b	XA182c2	XA191c2	0,02	3,59E+05	3,53E+05	YMX005613	0,04	2,16E+05	2,07E+05
6b	XA184c6	XA191c2	0,01	3,58E+05	3,52E+05	YMX005613	0,04	2,17E+05	2,08E+05
6b	YMX004664	XA191c2	0,02	3,59E+05	3,53E+05	YMX005613	0,04	2,16E+05	2,07E+05
6b	YMX004689	XA191c2	0,01	3,58E+05	3,52E+05	YMX005613	0,04	2,17E+05	2,08E+05
6b	YMX005589	XA191c2	0,02	3,59E+05	3,53E+05	YMX005613	0,04	2,17E+05	2,07E+05
6b	YMX506C11	XA191c2	0,01	3,57E+05	3,52E+05	YMX005613	0,04	2,16E+05	2,07E+05
6b	YMX506D08	XA191c2	0,02	3,57E+05	3,52E+05	YMX005613	0,04	1,97E+05	1,89E+05
6b	YMX506D09	XA191c2	0,02	3,57E+05	3,52E+05	YMX005613	0,04	1,97E+05	1,89E+05
6b	YMX506E03	XA191c2	0,02	3,56E+05	3,50E+05	YMX005613	0,04	2,13E+05	2,04E+05
6b	YMX506E04	XA191c2	0,02	3,58E+05	3,53E+05	YMX005613	0,04	2,16E+05	2,06E+05
6b	YMX506H11	XA191c2	0,02	3,59E+05	3,53E+05	YMX005613	0,04	2,11E+05	2,01E+05
6b	XA022c2	YMX005597	0,02	3,58E+05	3,53E+05	YMX005613	0,04	2,14E+05	2,04E+05
6b	XA183c1	YMX005597	0,02	3,59E+05	3,54E+05	YMX005613	0,04	2,19E+05	2,09E+05
6b	YMX004692	YMX005597	0,02	3,57E+05	3,53E+05	YMX005613	0,04	2,19E+05	2,09E+05
6b	YMX005598	YMX005597	0,02	3,57E+05	3,52E+05	YMX005613	0,04	2,15E+05	2,06E+05
6b	YMX506C09	YMX005597	0,02	3,57E+05	3,52E+05	YMX005613	0,04	2,19E+05	2,10E+05
6b	YMX506D03	YMX005597	0,02	3,57E+05	3,52E+05	YMX005613	0,04	2,14E+05	2,05E+05
6b	YMX506E02	YMX005597	0,02	3,50E+05	3,46E+05	YMX005613	0,04	2,07E+05	1,99E+05
6b	YMX506H12	YMX005597	0,02	3,57E+05	3,53E+05	YMX005613	0,04	2,18E+05	2,08E+05

6b	YMX507A03	YMX005597	0,02	3,58E+05	3,53E+05	YMX005613	0,04	2,09E+05	2,00E+05
NatHyb	YMX508B11	YMX507B05	1,32	3,02E+03	2,23E+03	YMX506B12	9,80E-05	2,27E+05	2,27E+05
NatHyb	YMX508B12	XA211c1	0,05	3,30E+05	3,21E+05	YMX506G04	0,07	1,41E+05	1,35E+05

Tabla Suplementaria 5. Distancias filogenéticas entre los aislados parentales *S. cerevisiae*

Esta tabla muestra las distancias filogenéticas entre las 12 cepas *S. cerevisiae* identificadas como más cercanas

	YMX0055 97	XA191c2	YMX0055 94	SRR8008 27	YMX507B 05	XA211c1	YMX0055 92	BR023c1	YMX506F 12	XA121c1 8	XA017c1	XB001c4
YMX0055 97	0	0.017	0.08	0.11	0.29	0.04	0.13	0.13	0.22	0.20	0.13	0.30
XA191c2	0.02	0	0.083	0.12	0.29	0.04	0.13	0.13	0.23	0.21	0.13	0.30
YMX0055 94	0.08	0.08	0	0.10	0.28	0.07	0.11	0.11	0.21	0.19	0.10	0.28
SRR8008 27	0.11	0.12	0.1	0	0.29	0.1	0.15	0.04	0.17	0.15	0.14	0.29
YMX507B 05	0.29	0.29	0.28	0.29	0.00	0.278	0.33	0.3	0.40	0.38	0.32	0.20
XA211c1	0.04	0.04	0.07	0.1	0.28	0	0.12	0.11	0.21	0.19	0.11	0.29
YMX0055 92	0.13	0.13	0.11	0.15	0.33	0.12	0	0.16	0.26	0.24	0.02	0.33
BR023c1	0.13	0.13	0.11	0.04	0.30	0.11	0.16	0	0.19	0.17	0.16	0.31
YMX506F 12	0.22	0.23	0.21	0.17	0.40	0.21	0.26	0.18	0	0.07	0.25	0.40
XA121c1 8	0.2	0.21	0.19	0.15	0.38	0.19	0.24	0.16	0.07	0	0.23	0.38
XA017c1	0.13	0.13	0.1	0.14	0.32	0.11	0.02	0.16	0.25	0.23	0	0.33
XB001c4	0.3	0.3	0.28	0.29	0.20	0.29	0.33	0.31	0.4	0.38	0.33	0

Tabla Suplementaria 6. Distancias filogenéticas entre los aislados parentales *S. paradoxus*

Esta tabla muestra las distancias filogenéticas entre las 8 cepas *S. paradoxus* identificadas como más cercanas

	YMX005607	YMX005613	YMX005537	YMX004559	YMX004555	YMX506E06	YMX506G04	YMX506B12
YMX005607	0	0.1	0.12	0.09	0.02	0.25	0.1	0.35
YMX005613	0.1	0	0.09	0.03	0.11	0.26	0.02	0.35
YMX005537	0.12	0.09	0	0.08	0.13	0.28	0.09	0.38
YMX004559	0.09	0.03	0.08	0	0.1	0.25	0.03	0.35
YMX004555	0.02	0.11	0.13	0.1	0	0.26	0.1	0.35
YMX506E06	0.25	0.26	0.28	0.25	0.26	0	0.26	0.25
YMX506G04	0.1	0.02	0.09	0.03	0.1	0.26	0	0.35
YMX506B12	0.35	0.35	0.38	0.35	0.35	0.25	0.35	0

8. Glosario

Aislado: Muestra de un organismo obtenida para estudio o experimentación. En este contexto se usa para referir a las cepas de levadura analizadas.

Cobertura Genómica: Medida de la cantidad de veces que una base es secuenciada, puede dar evidencia sobre la composición genómica del organismo secuenciado.

Filogenia: Determinación de las relaciones evolutivas entre organismos, en este proyecto se utilizan para entender una parte de la historia evolutiva de los híbridos encontrados.

Híbrido: Organismo resultante de la cruce entre individuos de dos especies distintas, en el contexto de este trabajo, entre *Saccharomyces cerevisiae* y *Saccharomyces paradoxus*.

Introgresión: Incorporación de material genético de una especie en el genoma de otra, se pueden originar por hibridaciones o por otros mecanismos como la transferencia horizontal de material genético.

Lecturas Mapeadas: Secuencias de ADN que se han alineado a una referencia genómica, dependiendo de la referencia a la que hayan alineado se puede inferir el origen de dichas lecturas.

Mapeo Competitivo: Técnica bioinformática utilizada para alinear secuencias de ADN a múltiples referencias genómicas simultáneamente, con el objetivo de determinar la procedencia de las moléculas de ADN secuenciadas.

Secuenciación: Proceso de determinar la sucesión exacta de nucleótidos en una molécula de ADN.

Transferencia horizontal: Transferencia de material genético entre organismos de la misma o diferentes especies, esta forma de transmisión de información genética es distinta a la herencia de progenitores a hijos (vertical).

9. Referencias

- Alsammar, H. (2019). Environmental genomics of the *Saccharomyces sensu stricto* group. Research Explorer The University of Manchester. Retrieved March 7, 2024, from <https://research.manchester.ac.uk/en/studentTheses/environmental-genomics-of-the-saccharomyces-sensu-stricto-group>
- Alsammar, H., & Delneri, D. (2020). An update on the diversity, ecology and biogeography of the *Saccharomyces* genus. *FEMS Yeast Research*, 20(3), foaa013. <https://doi.org/10.1093/femsyr/foaa013>
- Barbosa, R., Almeida, P., Safar, S. V. B., Santos, R. O., Morais, P. B., Nielly-Thibault, L., Leducq, J.-B., Landry, C. R., Gonçalves, P., Rosa, C. A., & Sampaio, J. P. (2016). Evidence of Natural Hybridization in Brazilian Wild Lineages of *Saccharomyces cerevisiae*. *Genome Biology and Evolution*, 8(2), 317–329. <https://doi.org/10.1093/gbe/evv263>
- Bendixsen, D. P., Frazão, J. G., & Stelkens, R. (2022). *Saccharomyces* yeast hybrids on the rise. *Yeast*, 39(1–2), 40–54. <https://doi.org/10.1002/yea.3684>
- D'Angiolo, M., De Chiara, M., Yue, J.-X., Irizar, A., Stenberg, S., Persson, K., Llored, A., Barré, B., Schacherer, J., Marangoni, R., Gilson, E., Warringer, J., & Liti, G. (2020). A yeast living ancestor reveals the origin of genomic introgressions. *Nature*, 587(7834), Article 7834. <https://doi.org/10.1038/s41586-020-2889-1>
- De Chiara, M., Friedrich, A., Barré, B., Breitenbach, M., Schacherer, J., & Liti, G. (2020). Discordant evolution of mitochondrial and nuclear yeast genomes at population level. *BMC Biology*, 18(1), 49. <https://doi.org/10.1186/s12915-020-00786-4>
- Fischer, G., James, S. A., Roberts, I. N., Oliver, S. G., & Louis, E. J. (2000). Chromosomal evolution in *Saccharomyces*. *Nature*, 405(6785), Article 6785. <https://doi.org/10.1038/35013058>

- Fischer, G., Liti, G., & Llorente, B. (2021). The budding yeast life cycle: More complex than anticipated? *Yeast*, 38(1), 5–11. <https://doi.org/10.1002/yea.3533>
- Fritsch, E. S., Chabbert, C. D., Klaus, B., & Steinmetz, L. M. (2014). A genome-wide map of mitochondrial DNA recombination in yeast. *Genetics*, 198(2), 755–771. <https://doi.org/10.1534/genetics.114.166637>
- Futcher, A. B. (1988). The 2 μ m circle plasmid of *Saccharomyces cerevisiae*. *Yeast*, 4(1), 27–40. <https://doi.org/10.1002/yea.320040104>
- Gabaldón, T. (2020). Hybridization and the origin of new yeast lineages. *FEMS Yeast Research*, 20(5), foaa040. <https://doi.org/10.1093/femsyr/foaa040>
- Gallegos-Casillas, P., García-Ortega, L. F., Espinosa-Cantú, A., Torres-Lagunes, C. G., Avelar-Rivas, J. A., Cano-Ricardez, A., García-Acero, Á. M., Ruiz-Castro, S., Flores-Barraza, M., Castillo, A., González-Zozaya, F., Delgado-Lemus, A., Molina-Freaner, F., Jacques-Hernández, C., Hernández-López, A., Delaye, L., Aguirre-Dugua, X., Kirchmayr, M. R., Morales, L., ... DeLuna, A. (2023). Yeast diversity in open agave fermentations across México (p. 2023.07.02.547337). *bioRxiv*. <https://doi.org/10.1101/2023.07.02.547337>
- Gibson, B., & Liti, G. (2015). *Saccharomyces pastorianus*: Genomic insights inspiring innovation for industry. *Yeast*, 32(1), 17–27. <https://doi.org/10.1002/yea.3033>
- Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., & Oliver, S. G. (1996). Life with 6000 genes. *Science (New York, N.Y.)*, 274(5287), 546, 563–567. <https://doi.org/10.1126/science.274.5287.546>
- Hénault, Mathieu, Souhir Marsit, Guillaume Charron, and Christian R. Landry. “Hybridization Drives Mitochondrial DNA Degeneration and Metabolic Shift in a Species with Biparental Mitochondrial Inheritance.” *Genome Research* 32, no. 11–12 (2022): 2043–56. <https://doi.org/10.1101/gr.276885.122>.

- Knop, M. (2006). Evolution of the hemiascomycete yeasts: On life styles and the importance of inbreeding. *BioEssays*, 28(7), 696–708. <https://doi.org/10.1002/bies.20435>
- Kurtzman, C., Fell, J. W., & Boekhout, T. (2011). *The Yeasts: A Taxonomic Study*. Elsevier.
- Leducq, J.-B., Nielly-Thibault, L., Charron, G., Eberlein, C., Verta, J.-P., Samani, P., Sylvester, K., Hittinger, C. T., Bell, G., & Landry, C. R. (2016). Speciation driven by hybridization and chromosomal plasticity in a wild yeast. *Nature Microbiology*, 1(1). Scopus. <https://doi.org/10.1038/nmicrobiol.2015.3>
- Libkind, D., Hittinger, C. T., Valério, E., Gonçalves, C., Dover, J., Johnston, M., Gonçalves, P., & Sampaio, J. P. (2011). Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 108(35), 14539–14544. <https://doi.org/10.1073/pnas.1105430108>
- Majdak, A. (2020). Comparison of Wine Aroma Compounds Produced by *Saccharomyces paradoxus* and *Saccharomyces cerevisiae* Strains. *Journal Article*.
- Malina, C., Larsson, C., & Nielsen, J. (2018). Yeast mitochondria: An overview of mitochondrial biology and the potential of mitochondrial systems biology. *FEMS Yeast Research*, 18(5), foy040. <https://doi.org/10.1093/femsyr/foy040>
- Marcet-Houben, M., & Gabaldón, T. (2015). Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker's Yeast Lineage. *PLOS Biology*, 13(8), e1002220. <https://doi.org/10.1371/journal.pbio.1002220>
- Martin, S. H., & Jiggins, C. D. (2017). Interpreting the genomic landscape of introgression. *Current Opinion in Genetics & Development*, 47, 69–74. <https://doi.org/10.1016/j.gde.2017.08.007>
- Mercado, L. V. (2022, May 23). Mula: ¿Por qué es un animal híbrido y no puede tener hijos?. *El Tiempo*. <https://www.eltiempo.com/vida/ciencia/mula-por-que-no-puede-tener-hijos-646162#:~:text=Las%20mulas%20son%20conocidas%20por,similares%20a%20las%20del%20burro.>

- Merz, S., & Westermann, B. (2009). Genome-wide deletion mutant analysis reveals genes required for respiratory growth, mitochondrial genome maintenance and mitochondrial protein synthesis in *Saccharomyces cerevisiae*. *Genome Biology*, 10(9), R95. <https://doi.org/10.1186/gb-2009-10-9-r95>
- Mixão, V., & Gabaldón, T. (2018). Hybridization and emergence of virulence in opportunistic human yeast pathogens. *Yeast (Chichester, England)*, 35(1), 5–20. <https://doi.org/10.1002/yea.3242>
- Mixão, V., & Gabaldón, T. (2020). Genomic evidence for a hybrid origin of the yeast opportunistic pathogen *Candida albicans*. *BMC Biology*, 18, 48. <https://doi.org/10.1186/s12915-020-00776-6>
- Morales, L., & Dujon, B. (2012). Evolutionary Role of Interspecies Hybridization and Genetic Exchanges in Yeasts. *Microbiology and Molecular Biology Reviews*, 76(4), 721–739. <https://doi.org/10.1128/membr.00022-12>
- Nikulin, J., Vidgren, V., Krogerus, K., Magalhães, F., Valkeemäki, S., Kangas-Heiska, T., & Gibson, B. (2020). Brewing potential of the wild yeast species *Saccharomyces paradoxus*. *European Food Research and Technology*, 246(11), 2283–2297. <https://doi.org/10.1007/s00217-020-03572-2>
- Notion Labs Inc. (2016), Notion [Software en línea]. <https://www.notion.so/product>
- Peris, D., Arias, A., Orlić, S., Belloch, C., Pérez-Través, L., Querol, A., & Barrio, E. (2017). Mitochondrial introgression suggests extensive ancestral hybridization events among *Saccharomyces* species. *Molecular Phylogenetics and Evolution*, 108, 49–60. <https://doi.org/10.1016/j.ympev.2017.02.008>
- Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., ... Schacherer, J. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature*, 556(7701), Article 7701. <https://doi.org/10.1038/s41586-018-0030-5>

- Samarasinghe, H., & Xu, J. (2018). Hybrids and hybridization in the *Cryptococcus neoformans* and *Cryptococcus gattii* species complexes. *Infection, Genetics and Evolution*, 66, 245–255. <https://doi.org/10.1016/j.meegid.2018.10.011>
- Schröder, M. S., Vicente, K. M. de S., Prandini, T. H. R., Hammel, S., Higgins, D. G., Bagagli, E., Wolfe, K. H., & Butler, G. (2016). Multiple Origins of the Pathogenic Yeast *Candida orthopsilosis* by Separate Hybridizations between Two Parental Species. *PLOS Genetics*, 12(11), e1006404. <https://doi.org/10.1371/journal.pgen.1006404>
- Sedeño, Ivan. (2023). Híbridos e introgresiones en levaduras de agave [Tesis de Licenciatura], Universidad Nacional Autónoma de México.
- Solieri, L. (2010). Mitochondrial inheritance in budding yeasts: Towards an integrated understanding. *Trends in Microbiology*, 18(11), 521–530. <https://doi.org/10.1016/j.tim.2010.08.001>
- Stelkens, R., & Bendixsen, D. P. (2022). The evolutionary and ecological potential of yeast hybrids. *Current Opinion in Genetics & Development*, 76, 101958. <https://doi.org/10.1016/j.gde.2022.101958>
- Strope, P. K., Skelly, D. A., Kozmin, S. G., Mahadevan, G., Stone, E. A., Magwene, P. M., Dietrich, F. S., & McCusker, J. H. (2015). The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Research*, 25(5), 762–774. <https://doi.org/10.1101/gr.185538.114>
- Tsai, I. J., Bensasson, D., Burt, A., & Koufopanou, V. (2008). Population genomics of the wild yeast *Saccharomyces paradoxus*: Quantifying the life cycle. *Proceedings of the National Academy of Sciences*, 105(12), 4957–4962. <https://doi.org/10.1073/pnas.0707314105>
- Wolfe, K. H. (2015). Origin of the Yeast Whole-Genome Duplication. *PLoS Biology*, 13(8), e1002221. <https://doi.org/10.1371/journal.pbio.1002221>