



Universidad Nacional Autónoma De México

Centro de Física Aplicada y Tecnología Avanzada
Laboratorio Internacional de Investigación sobre el Genoma Humano

Híbridos e introgresiones en levaduras de agave

QUE PRESENTA:
Iván Eduardo Sedeño Jiménez

QUE PARA OBTENER EL TÍTULO DE:
Licenciado en Tecnología

T E S I S

TUTORAS:
Lucía Guadalupe Morales Reyes
Maritrini Colón González
Laboratorio Internacional de Investigación sobre el Genoma Humano



Santiago de Querétaro, Querétaro, México. Marzo 2023



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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.

Agradecimientos

Agradezco al Programa UNAM-PAPIIT IN209021 por la beca otorgada durante la realización de este trabajo de tesis.

Agradezco a las técnicas Alejandra Castillo Carbajal y Carina Díaz Uribe del LIIGH por su apoyo técnico en el laboratorio. De igual forma agradezco a Luis Alberto Aguilar Bautista y Jair Santiago García Sotelo del LAVIS-UNAM por su apoyo técnico en tecnologías de la información.

Esta investigación fue realizada gracias al Programa UNAM-PAPIIT IN209021 y a los donativos CONACYT CB-284992 y CF-103000.

También quiero agradecer a mi tutora, la Dra. Lucía Morales, por su paciencia y dedicación, sin su guía no hubiese podido lograr llegar hasta esta instancia. Gracias por darme la oportunidad de formar parte del Genome Evolution Laboratory del LIIGH y por compartir conmigo sus conocimientos.

De igual forma quiero agradecer a mi co-asesora de tesis, la Dra. Maritrini Colón, por sus aportaciones y correcciones precisas al momento de redactar la tesis.

A los sinodales, muchas gracias por dedicarme de su tiempo y tomarse la molestia de leer esta tesis y brindarme sus observaciones.

Agradecimientos a título personal

En primer lugar, quiero agradecer a mis padres que siempre me han apoyado a lograr todos mis objetivos académicos y personales. Me han impulsado a siempre a perseguir mis metas y nunca abandonarlas pese a las adversidades. Durante todos estos años me han brindado apoyo económico para que mi única preocupación sea concentrarme en mis estudios.

Quiero agradecer también a todos los miembros del GEL Lab, son un grupo de personas increíbles que siempre me hicieron sentir bienvenido y dispuestas a apoyarse entre sí. Gracias por sus invaluable opiniones y aportaciones al momento de desarrollar este trabajo.

También quiero darles las gracias a los miembros del consorcio Yeast Genomes MX, quienes durante varios años fueron parte de mi desarrollo académico y me vieron crecer en el área de la bioinformática, brindándome consejo y guía cuando lo necesite. Muchas gracias a los doctores Eugenio Mancera y Alexander de Luna cuya retroalimentación al realizar los análisis fue vital. Gracias al Dr. Luis Fernando con quien tuve la oportunidad de trabajar de la mano en el diseño y ejecución del pipeline bioinformático en lo relativo al proyecto de levaduras mezcaleras.

Es necesario mencionar también a mis amigos, quienes me han acompañado durante todos estos años y me han brindado su apoyo cuando más lo necesite, me han ayudado a distraerme cuando estaba agobiado y ocupado con todas mis actividades. Gracias por nunca dejarme solo y ser parte de mi vida.

Por último y en caso de que olvide a alguien, quiero agradecer a todas las personas que han sido parte de mi vida, especialmente a las personas que en este último año han aparecido en mi vida y me han ayudado a navegar este difícil capítulo de mi vida.

Índice

Lista de Figuras	5
Lista de Tablas	6
Lista de Material Suplementario.....	7
Resumen.....	8
1. Introducción.....	9
1.1 Las levaduras.....	9
1.2 Las levaduras y los humanos	9
1.3 Las levaduras como organismo modelo	10
1.4 Las levaduras: una taxonomía complicada	11
1.5 Hibridaciones e introgresiones	13
2. Antecedentes	14
3. Justificación	15
4. Objetivos.....	16
Principal.....	16
Específicos	16
5. Metodología	17
5.1 Muestreo y aislamiento de cepas de <i>Saccharomyces</i> de fermentaciones de agave.....	17
5.2 Selección de muestras usadas en este estudio	18
5.3 Secuenciación de genomas.....	18
5.4 Alineamiento a referencias concatenadas	19
5.5 Asignación a especie de los genomas secuenciados	20
5.6 Cálculo de la profundidad en la cobertura.....	20
5.7 Determinación de introgresiones.....	20
5.8 Asignación funcional de genes introgresados	22
6. Resultados	24
6.1 Colinealidad y divergencia entre los genomas de <i>S. paradoxus</i> y <i>S. cerevisiae</i>	24
6.2 Descripción de los aislados secuenciados.....	26
6.3 Asignación a especie mediante alineamiento de lecturas	27
6.4 Análisis de cobertura	29
6.4.1 Análisis de cobertura en aislados <i>S. cerevisiae</i>	29
6.4.2 Análisis de genomas híbridos	31

6.5 Introgresiones de <i>S. paradoxus</i>	34
6.6 Análisis de ontología de genes.....	35
6.6.1 Cantidad y distribución	35
6.6.2 Análisis de ontología de genes	38
7. <i>Discusión</i>	40
8. <i>Conclusiones</i>	44
9. <i>Perspectivas</i>	44
10. <i>Referencias bibliográficas</i>	46
11. <i>Bibliografía</i>	49
12. <i>Material Suplementario</i>	50

Lista de Figuras

Figura 1.1. Diferencia en la tasa de sustitución en distintos clados. En esta imagen se muestra el nivel de divergencia que existe entre proteínas homólogas en algunas especies de levaduras, animales y plantas. Destaca que en levaduras los niveles de divergencias son altos si se comparan con los animales y plantas.

Figura 5.1. Mapa de México con la distribución geográfica de zonas productoras de agaves. En esta imagen se muestran las distintas zonas de la república destinadas al cultivo de agaves y la producción de mezcales.

Figura 6.1. Alineamiento producido por NUCmer. En esta imagen se muestra el alineamiento que arrojó NUCmer al recibir como entrada las referencias de los genomas de *S. cerevisiae* S288C y *S. paradoxus* YPS138. La presencia de una línea azul continua indica el alto grado de colinealidad abarcando de forma uniforme a todos los cromosomas.

Figura 6.2. Bloques sinténicos producidos por SynChro. En esta imagen se muestra los bloques sinténicos calculados por SynChro al recibir como entrada las referencias de los genomas de *S. cerevisiae* S288C y *S. paradoxus* YPS138. Aunque la línea morada muestra pequeñas interrupciones abarca a todos los cromosomas casi en su totalidad.

Figura 6.3. Cobertura normalizada hacia *S. cerevisiae* y *S. paradoxus* del aislado híbrido XA082c2. En esta figura se presentan las gráficas de cobertura normalizada del aislado XA082c2 hacia *S. cerevisiae* y *S. paradoxus*, normalizadas dividiendo entre la media de la cobertura de las lecturas alineadas hacia *S. cerevisiae*.

Figura 6.4. Distribución de lecturas alineadas a *S. paradoxus*. Esta figura nos presenta el porcentaje de lecturas de cada aislado que se alinearon hacia *S. paradoxus*, los aislados se encuentran separados por origen de la secuencia y clado.

Figura 6.5. Proporción de las lecturas alineadas a cada genoma de referencia para los aislados híbridos. En esta figura se presenta la proporción de lecturas alineadas hacia cada especie contenida en la referencia concatenada (ver **Tabla 5.1**). Hay aislados con aproximadamente el 30% de sus lecturas alineadas hacia *S. paradoxus*, otros con el 50% y finalmente otros con el 60%. También hay una proporción reducida de lecturas alineadas hacia *Saccharomyces jurei* y *Saccharomyces kudriavzevi*.

Figura 6.6. Cobertura normalizada hacia *S. cerevisiae* y *S. paradoxus* de los aislados híbridos XA020c1, XA082c2 y XB003c8. En esta figura se presentan las gráficas de cobertura normalizada de los aislados XA020c1, XA082c2 y XB003c8 hacia *S. cerevisiae* y *S. paradoxus*, normalizadas dividiendo entre la media de la cobertura de las lecturas alineadas hacia *S. cerevisiae*.

Figura 6.7. Distribución de las introgresiones de *S. paradoxus*. En esta imagen se muestra la distribución del número de introgresiones de *S. paradoxus* separadas por grupos de muestras según su origen.

Figura 6.8. Histograma del número de introgresiones compartidas entre aislados. a) En esta imagen se muestra **a)** el número de introgresiones compartidas entre cierto porcentaje de aislados para todos los aislados. Véase que los híbridos comparten la gran mayoría de sus introgresiones. **b)** el número de introgresiones compartidas entre cierto porcentaje de aislados para los aislados de *S. cerevisiae* secuenciados para este trabajo. Contrario al caso anterior, los aislados de *S. cerevisiae* tienen pocas introgresiones comunes.

Figura 6.9. Ruta metabólica para la síntesis *de novo* de PLP y TDP. En esta imagen, modificada de Perli *et al.* se presenta un diagrama de la biosíntesis de vitamina B, donde se muestra la participación de los genes enriquecidos según el análisis de ontología genética.

Figura 6.10. Homólogos en *S. cerevisiae* de los genes enriquecidos de *S. paradoxus*. En esta imagen se aprecia la localización de los genes homólogos en *S. cerevisiae* de los genes de *S. paradoxus* enriquecidos según el análisis de categorías GO, los genes de interés están marcados con una estrella negra y se encuentran dentro del área amarilla; solamente para los cromosomas VI y XIV.

Lista de Tablas

Tabla 5.1. Genomas en la referencia concatenada. Esta tabla enlista la relación de cepas que se usaron para crear la referencia concatenada.

Tabla 5.2. ORFs introgresados según los distintos filtros. En esta tabla se presenta el número de ORFs que están presentes en distintos números de aislados para cada filtro utilizado.

Tabla 6.1. Estadísticas de sintenia obtenidas por SynChro. En esta tabla se encuentran el porcentaje de nucleótidos y genes que se encuentran en bloques sinténicos, así como su porcentaje de identidad a nivel de bases.

Tabla 6.2. Número de genomas analizados. En esta tabla se muestra la cantidad de genomas que secuenciamos para este trabajo separados por tecnología de secuenciación, aunado al número de genomas derivados de otros trabajos que usamos para complementar nuestro análisis.

Tabla 6.3. Asignación de especies a aislados secuenciados para este trabajo. En esta tabla se enlista el número de aislados asignado a cada especie.

Tabla 6.4. Asignación de especies a aislados secuenciados anteriormente. En esta tabla se enlista el número de aislados (provenientes de Peter y colaboradores (2018; Gallone y colaboradores (2019)) asignados a cada especie.

Tabla 6.5. Asignación de especies a aislados secuenciados. En esta tabla se enlistan las asignaciones a especie de los aislados secuenciados según su tecnología de secuenciación.

Tabla 6.6. Genes introgresados asociados a las categorías GO sobrerrepresentadas agrupados por función y ubicación. En esta tabla se agrupan los genes de *S. paradoxus* asociados a categorías GO sobrerrepresentadas agrupados por su función y ubicación dentro de la cepa de referencia, se muestra además el porcentaje de cepas mexicanas y del mundo que comparten la introgresión, y la anotación del gen, en *S. cerevisiae*, obtenida de Saccharomyces Genome Database (SGD, <http://www.yeastgenome.org>).

Lista de Material Suplementario

Tabla Suplementaria 1. Lista de genomas usados. Esta tabla enlista a las 316 secuencias que se usaron en este trabajo, la especie a la que pertenecen, su origen y plataforma de secuenciación.

Tabla Suplementaria 2. Lista de lecturas alineadas a cada especie de la referencia concatenada. En esta tabla se enlistan los 316 aislado secuenciados con él número total de secuencias aisladas y el porcentaje de lecturas que se alinearon a cada genoma de referencia.

Resumen

En el mundo existe una gran variedad de alimentos y bebidas que en algún punto pasan por un proceso de fermentación, desde bebidas de popularidad internacional como la cerveza hasta alimentos endémicos como el kimchi. En muchos casos la fermentación es llevada a cabo por las levaduras del género *Saccharomyces*, en especial por la especie *Saccharomyces cerevisiae*. Esta especie alberga mucha variación genética lo que le permite adaptarse a los distintos ambientes en donde se lleva a cabo la fermentación. Cabe destacar que los miembros de esta familia tienen la característica de formar híbridos inter-especie, quebrantando así una de las reglas básicas para la definición biológica de especies.

En esta tesis presentamos el análisis de los genomas de 142 cepas de *Saccharomyces* aisladas de fermentación de agave de distintas partes del territorio mexicano. Se realizaron análisis de cobertura para identificar a que especie pertenecía cada cepa y explorar la existencia de flujo génico entre miembros del género. Logramos identificar la coexistencia *S. cerevisiae*, de *S. paradoxus* y de híbridos entre éstas dos especies en la población *Saccharomyces* asociadas a la fermentación de agave. Finalmente, describimos la cantidad y naturaleza de fragmentos de ADN provenientes de *S. paradoxus* en la población de *S. cerevisiae* que fermenta azúcares de agaves.

1. Introducción

1.1 Las levaduras

El término *levadura* comúnmente se utiliza para referirse a los organismos microscópicos usados durante la elaboración de bebidas o alimentos fermentados. Incluso en muchos idiomas este término hace referencia a “espumar” o “levar”, acciones que derivan del efecto de la producción de CO₂ al realizarse la fermentación. Desde un punto de vista evolutivo, las levaduras engloban a más de 1600 especies diferentes de hongos unicelulares que pertenecen a distintos grupos taxonómicos. Estas células miden en promedio entre 2 y 4 µm, y aunque se reproducen predominantemente de manera asexual por medio de gemación, algunas especies también llevan a cabo reproducción sexual facultativa (Kurtzman, *et al.*, 2011). Un rasgo característico de las levaduras es que pueden poseer dos nombres dependiendo de su estado sexual. Su nombre principal se basa en su estado sexual o teleomorfa, mientras que su nombre secundario se basa en su estado asexual o anamorfa (Kurtzman, *et al.*, 2011).

Kurtzman y colaboradores (2011) reconocen más de 1600 especies distintas de levaduras, las cuales tienen metabolismos heterotróficos verdaderamente diversos. Estas variaciones en el metabolismo les han permitido adaptarse a una amplia gama de hábitats (terrestres o marinos) y participar en un gran número de procesos ecológicos. Clásicamente a los hongos se les ubica como descomponedores primarios dentro de la cadena alimenticia. Típicamente las levaduras son de los primeros colonizadores en un sustrato rico en nutrientes, seguidos de múltiples microorganismos cuyo fin es degradar la materia en descomposición. Sin embargo, su rol no es el de simplemente descomponer materia sino de transformarla (Kurtzman, *et al.*, 2011).

Las levaduras han acompañado al ser humano desde hace milenios y tienen un gran número de aplicaciones biotecnológicas. Quizá su aplicación más obvia y conocida es la de su uso en la fermentación de bebidas como la cerveza y el vino, o la de su empleo en la elaboración de alimentos como pan, pero están lejos de ser sus únicas aplicaciones. Las levaduras también son usadas en la elaboración de bioetanol, así como la producción de enzimas para las industrias química y farmacéutica (Kurtzman, *et al.*, 2011). En el contexto actual de pandemia, las levaduras también se han usado para la elaboración de una nueva vacuna contra el virus SARS-CoV-2 (Dalvie, *et al.*, 2021). Por último, no hay que olvidar que ciertas especies son patógenas para el humano por lo que también son organismos de importancia médica (Kurtzman, *et al.*, 2011). En este trabajo analizaremos el genoma de un grupo de levaduras asociadas a las fermentaciones de azúcares de agaves cocidos en México.

1.2 Las levaduras y los humanos

Las levaduras no son exclusivas al nicho de agave ni de uso reciente. Estos microorganismos han acompañado a los seres humanos a lo largo de su historia y han ayudado a sustentar civilizaciones antiguas a través de la producción de alimentos fermentados. Por ejemplo, en la villa neolítica de Jiahu en China, se ha encontrado evidencia arqueológica de la producción de

bebidas fermentadas que datan del 7000 a.C. (Sicard & Legras, 2011). En otros lugares de Asia, se ha encontrado evidencia antigua de bebidas fermentadas producidas principalmente a partir de arroz, aunada a más evidencia en sitios como Egipto y Mesopotamia donde se creaban bebidas a partir de frutas (vino), miel (hidromiel) y cereales (cerveza). También hay evidencia de fabricación de pan levado en el antiguo Egipto (Sicard & Legras, 2011).

En un inicio, la preparación de alimentos y bebidas fermentadas era poco sofisticada, en el caso del vino del 5000 a.C., este se producía a partir de uvas recolectadas a mano y prensadas, el jugo resultante se colocaba en vasijas abiertas dejadas al aire libre para su fermentación. Del mismo modo hay evidencia que sugiere la existencia de pan levado en esta época, cuya elaboración implicaba el uso de trigo molido gruesamente. La datación de cuándo la fermentación pasó de dejarse a la naturaleza, y comenzó a volverse un proceso consciente y selectivo, sigue siendo un misterio (Sicard & Legras, 2011).

En la actualidad las levaduras siguen siendo microorganismos domesticados ligados estrechamente al desarrollo de las civilizaciones humanas en diferentes partes del planeta. Las usamos comúnmente para la producción de una gran variedad de alimentos y bebidas, algunos ejemplos serían: la cerveza, el vino, tequila, mezcal, sake, pan, cacao, chicha, kombucha, kéfir, soju, entre muchos otros (Sicard & Legras, 2011). En décadas recientes, su uso se ha extendido a las industrias farmacéuticas y energéticas. Los ejemplos anteriores evidencian la estrecha relación que hemos mantenido humanos y levaduras a largo de la historia (Kurtzman, *et al.*, 2011).

1.3 Las levaduras como organismo modelo

Como se ha mencionado previamente, las levaduras han acompañado al ser humano desde hace miles de años. Sin embargo, no fue hasta el año de 1789 cuando Antoine Lavoisier comenzó a estudiar la fermentación para entender como es que esto sucedía en la producción de vinos. Y no fue sino hasta varias décadas después, en 1860, que se descubrió que las levaduras eran capaces de convertir azúcares en etanol y dióxido de carbono, es decir, fermentar (Sicard & Legras, 2011). En los años siguientes los esfuerzos de la comunidad científica para comprender en profundidad el proceso de fermentación se redoblaron y hasta la fecha, las levaduras siguen siendo organismos ampliamente estudiados.

Fue hasta la primera mitad del siglo pasado que se realizaron los primeros experimentos genéticos en levaduras. En 1935, Ojvind Winge y Otto Lausten disecaron ascosporas de la especie de levadura *Saccharomyces cerevisiae* demostrando la segregación mendeliana. Desde estos primeros experimentos se hizo evidente que usar a *S. cerevisiae* como organismo modelo traía ventajas importantes como: un tiempo de generación corto, fácil manejo y una ploidía que puede estabilizarse. Además, de ser un organismo unicelular (característica compartida con organismos procariotas) y tener una organización celular compartimentalizada (característica compartida con los demás organismos eucariotas, incluyendo al resto de las levaduras).

Debido a su creciente relevancia científica y al número creciente de investigaciones llevadas a cabo en *S. cerevisiae*, en la segunda mitad del siglo pasado se comenzó a trabajar en la secuenciación de su genoma. Inicialmente, se publicaron únicamente una serie de mapas genéticos, cada vez más completos y extensos, hasta que finalmente en 1989 se acordó un proyecto cuya finalidad era secuenciar el genoma de *S. cerevisiae*. Finalmente, después de 6 años de trabajo y la colaboración de más de 600 científicos en distintos laboratorios del mundo, se publicó en 1996 la secuencia del genoma completo de *S. cerevisiae*, conteniendo 6340 genes y convirtiéndose en el primer microorganismo eucariota en ser secuenciado (Goffeau, *et al.*, 1996). Actualmente, contamos con más de 2000 aislados de *S. cerevisiae* secuenciados, y en meses recientes gracias a la secuencia de lecturas largas hemos alcanzando tal resolución que ya es posible discernir entre haplotipos de aislados diploides (Heasley & Argueso, 2022).

A la fecha existe más de un genoma secuenciado para cada una de las ocho especies del género *Saccharomyces*: *S. cerevisiae*, *S. paradoxus*, *S. jurei*, *S. arboricola*, *S. uvarum*, *S. eubayanus*, *S. kudriavzevii* y *S. mikatae*. Para las más estudiadas como *S. cerevisiae* existen más de 2000 mil cepas con genomas secuenciados, no obstante, todas las especies del género cuentan ya con varias cepas secuenciadas a excelente resolución (Peris, *et al.*, 2022). *S. cerevisiae* ha hecho de su hábitat los ambientes fermentativos asociados al humano, habiendo incluso quienes sugieren que se trata de una especie nómada sin hábitat natural específico (Steensels, *et al.*, 2019). En años recientes ha comenzado a criticarse su uso como organismo modelo para aspectos ecológicos, puesto que su alto nivel de domesticación no permite estudiar cómo operan algunos mecanismos en poblaciones naturales. Con el propósito de evitar estos problemas, recientemente se ha propuesto utilizar a *S. paradoxus*, la especie hermana no domesticada de *S. cerevisiae* (Zhang, *et al.*, 2010).

S. paradoxus, pese a ser capaz de fermentar, raramente se encuentra en ambientes fermentativos. Se cree que su hábitat son los bosques de encinos y se encuentra de manera abundante en Norteamérica, Asia y Europa, donde convive con otras especies del género *Saccharomyces*, incluyendo a *S. cerevisiae* (Zhang *et al.*, 2010).

1.4 Las levaduras: una taxonomía complicada

Las levaduras son hongos unicelulares de origen no monofilético, lo que quiere decir que están en diferentes ramas del reino Fungi intercaladas con otros hongos pluricelulares. Actualmente se han descrito alrededor de 1600 especies de levaduras, las cuales se clasifican en dos *Phyla* del subreino Dikarya: *Ascomycota* (con 1015 especies) y *Basidiomycota* (con 561 especies). Dentro de *Ascomycota* se pueden distinguir 3 *subphyla* distintos: *Saccharomycotina* (aquí es donde podemos encontrar al género *Saccharomyces*), *Pezizomycotina* y *Taphrinomycotina*. Por otro lado, el *phylum Basidiomycota* alberga entre otras especies a las del género *Cryptococcus*, conocidas por su patogenicidad (Kurtzman, *et al.*, 2011).

La categoría fundamental de organización biológica es la especie, cuya definición más aceptada fue dada por Ernst Mayr durante la primera mitad del siglo pasado. Mayr define a las especies como un grupo de poblaciones naturales capaces de entrecruzarse y que están aisladas

reproductivamente de otros grupos, a esta definición usualmente se le conoce como “especie biológica”. Y aunque normalmente esta definición es suficiente, en algunos escenarios o grupos de organismos, esta definición no es válida o suficiente, razón por la que han surgido otro tipo de definiciones que se puedan aplicar en distintos contextos (de Queiroz, 2005). Por ejemplo, en el caso de las levaduras, éstas en muchos casos son asexuales y no producen meiosporas (esporas haploides); mientras que otras son homotáticas (capaces de reproducirse por si mismas) y pueden producir meiosporas, por lo que realmente esta definición no aplicaría de manera precisa a todas las levaduras (Kurtzman, *et al.*, 2011).

Otro de los motivos por los cuales la definición de especie en levaduras puede ser complicada yace en que es común encontrar divergencias de hasta el 2% (aproximadamente) a nivel nucleotídico entre miembros de una misma especie. Las distancias evolutivas entre las levaduras *Saccharomycotina* son tan grandes que incluso llegan a superar a la que hay en el Filo completo de los Cordados (*Chordata*). Bajo este contexto, la divergencia nucleotídica que usualmente sería suficiente para diferenciar entre humanos y chimpancés, apenas es suficiente para distinguir entre cepas de una misma especie de levaduras (**Figura 1.1**, (Shen, *et al.*, 2018)).

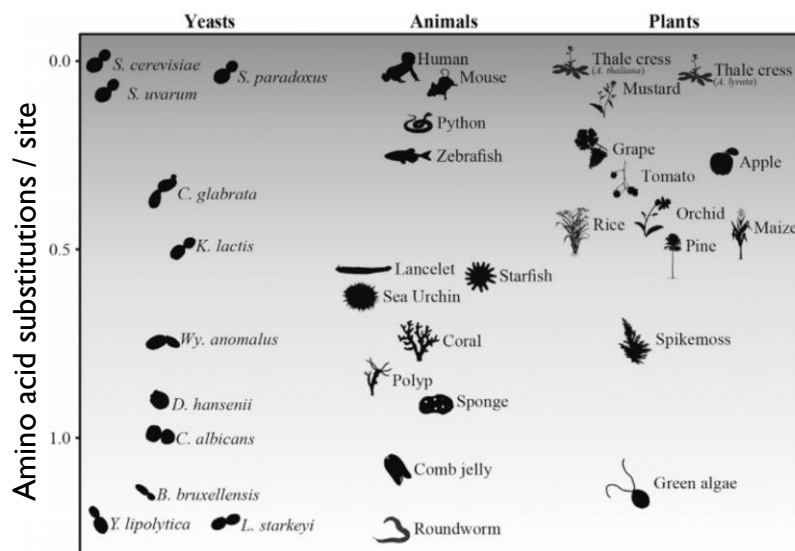


Figura 1.1: Diferencia en la tasa de sustitución en distintos clados. Nivel de divergencia entre proteínas homólogas en términos de sustituciones de aminoácidos por sitio en algunas especies de levaduras, animales y plantas. Se obtuvo la distancia filogenética entre especies destacadas de levaduras (*S. cerevisiae*), animales (*Homo sapiens*) y plantas (*Arabidopsis thaliana*) y otras especies representativas de cada linaje. Figura tomada de Shen y colaboradores (2018).

Además de lo previamente mencionado, el concepto de especie biológica no es suficiente para distinguir entre especies de levaduras. Algunas especies de levaduras son capaces de aparearse sexualmente con otras especies, formando híbridos que inicialmente se propagan por reproducción asexual o mitosis (Morales & Dujon, 2012). Aunque también se pueden encontrar híbridos en otro tipo de organismos como los mamíferos, tal es el caso de las mulas (cruza entre una yegua y un burro) o los ligres (cruza entre un león y una tigresa), en estos casos la progenie es infértil y por lo tanto no puede propagarse. En contraste, en el caso de las levaduras, aunque

sus híbridos tienen una fertilidad casi nula, lo cual les impide reproducirse sexualmente, aún se pueden reproducir clonalmente dando origen a una población híbrida (D'Angiolo, *et al.*, 2020; Morales & Dujon, 2012). Dicha población híbrida puede retrocruzarse con alguna de las especies parentales dejando fragmentos de la otra especie conocidos como introgresiones y descritos en la sección 1.5 de este trabajo.

Debido a su importancia industrial y su alta divergencia evolutiva, el género *Saccharomyces* ha sido ampliamente estudiado a nivel genómico, lo que ha llevado al descubrimiento de un gran número de híbridos poliploides, aneuploides y diploides en ambientes domesticados y naturales. Ahora es sabido que la hibridación interespecie ha participado activamente en la evolución de levaduras domesticadas. Probablemente, el ejemplo de hibridación interespecie más conocido lo conforman las levaduras involucradas en el proceso de fabricación de cervezas lager, entre ellas *S. pastorianus*. Esta especie contiene segmentos de ADN provenientes de 3 especies distintas y se cree que se formó por distintas hibridaciones antiguas que le ayudaron a adaptarse al frío (Sicard & Legras, 2011).

1.5 Hibridaciones e introgresiones

Es sabido que la hibridación interespecie es un proceso por el cual dos especies distintas juntan su material genético en una sola célula: el híbrido. Al igual que las levaduras parentales, este híbrido tiene la posibilidad de reproducirse asexualmente, generando una población que pese a ser infértil puede dar origen a muchos individuos. La hibridación interespecie es un evento recurrente en la evolución de *Saccharomyces*. Por ejemplo, evidencia reciente ha señalado que el clado producido por duplicación de genoma completo (whole genome duplication o WGD por sus siglas en inglés) es en realidad producto de una hibridación ancestral entre dos especies diploides que dieron lugar a una levadura alotetraploide que mediante pérdida masiva de genes originó un clado completo con el doble de cromosomas (Wolfe, 2015).

En la industria es muy recurrente la existencia de híbridos de *Saccharomyces* (Krogerus, *et al.*, 2018; Barbosa, *et al.*, 2016) y se ha propuesto que este mecanismo permite una adaptación rápida a ambientes estresantes como son los medios fermentativos.

La hibridación interespecie es un mecanismo que permite la formación de introgresiones, y particularmente en las especies del género *Saccharomyces* esta ocurre de manera recurrente en la naturaleza al tener nula o poca presencia de barreras precigóticas (Morales & Dujon, 2012). Una vez formado el híbrido este pierde casi por completo la capacidad para reproducirse sexualmente debido a que sus esporas se vuelven prácticamente inviables (<1% de viabilidad). El híbrido podría recuperar su fertilidad si llega a retrocruzarse con alguna especie parental volviéndose cada vez más fértil. Además, en cada generación, el híbrido resuelve las incompatibilidades a través de la reducción del material genético proveniente de uno de los parentales. Eventualmente, solo permanecen fragmentos pequeños del ADN de ese padre conocidos como introgresiones (D'Angiolo, *et al.*, 2020).

2. Antecedentes

En México, hay más de 200 especies de *Agave*, varias de ellas se usan para la elaboración de bebidas alcohólicas. A diferencia de la producción de tequila en la que se emplea sólo una especie de agave (*Agave tequilana*), para la producción de mezcal se utilizan más de 50 especies diferentes (Casas et al., 2017). De acuerdo a la Norma Oficial Mexicana, el mezcal se obtiene por medio de la destilación y rectificación del mosto preparado con los azúcares extraídos de las cabezas de los agaves, las cuales han sido previamente cocidas y fermentadas (Financiera Rural, 2011). Se distingue del proceso de fabricación de otros destilados de agave, como el tequila, debido a que su proceso de elaboración hasta la fecha es muy artesanal, y la mayoría de los productores no utiliza levaduras de uso industrial para inocular. Por esta razón, la fermentación del jugo de agave ocurre gracias a los microorganismos que están presentes en los tanques de fermentación y en los alrededores de las destilerías (Kirchmayr, et al., 2017).

A diferencia de otras bebidas fermentadas como el vino, se han hecho pocas investigaciones acerca de los organismos encargados de la fermentación del mosto de agave para la producción del mezcal. Estudios recientes han descrito a *S. cerevisiae* como la especie prevalente en los tanques de fermentación (Kirchmayr, et al., 2017; Lappe-Oliveras, et al., 2008; Garibay Marcelo, 2019).

En el 2018, Peter y colaboradores (Peter, et al., 2018) publicaron el estudio poblacional de levaduras *S. cerevisiae* más grande hasta la fecha. En dicho trabajo se analizaron los genomas de 1011 aislados de *S. cerevisiae* obtenidos de distintas partes del mundo. Se incluyeron aislados de diversos orígenes, tanto naturales como antropogénicos, entre ellos 9 aislados de mezcal. Derivado de este estudio, se encontraron cuatro grupos con un alto contenido de introgresiones provenientes de *S. paradoxus*, uno de estos el grupo de aislados de agave.

México es un país con una gran variedad de ecosistemas y climas, debido a esto, en cada región del país la elaboración del mezcal se lleva a cabo en condiciones diferentes, utilizando diversas especies de agave. Teniendo en cuenta que la producción de mezcal se realiza en distintos estados de la república (Sonora, Durango, Tamaulipas, Jalisco, Nayarit, Guanajuato, San Luis Potosí, Zacatecas, Estado de México, Guerrero, Michoacán, Oaxaca y Puebla), el estudio realizado por Peter y colaboradores (2018) no logra hacer justicia a la diversidad mexicana al utilizar únicamente nueve aislados de origen mexicano provenientes de Tamaulipas (n=7), Jalisco (n=1) y uno de origen desconocido. Por esta razón consideramos de gran importancia ecológica, evolutiva y cultural darnos a la tarea de realizar un muestreo sistemático de cepas aisladas de fermentación de agave en el que se incluyeron fermentos de todo el país con el objetivo de caracterizar sus introgresiones.

3. Justificación

Es sabido que en las levaduras asociadas a la fermentación de agave en México predomina la especie *S. cerevisiae*, y que en aislados de Tamaulipas y Jalisco son frecuentes las introgresiones de *S. paradoxus*. No obstante, aún no comprendemos cómo fue que estas introgresiones llegaron y se distribuyeron en esta población ni cual es su extensión y función.

En el desarrollo de esta tesis se busca ampliar el estado del arte de las introgresiones e hibridaciones de las levaduras mezcaleras mexicanas, comenzando por identificar cuales son estas introgresiones, y cuál es su prevalencia entre los aislados provenientes de distintas zonas de la República Mexicana.

4. Objetivos

Principal

Identificar los fragmentos de ADN introgresados de *S. paradoxus* en cepas de fermentaciones de agave de *S. cerevisiae*.

Específicos

1. Identificar mediante el mapeo a referencias de *Saccharomyces* las especies a las que pertenecen los aislados de agave.
2. Identificar los genes introgresados en cepas de *S. cerevisiae* analizando las regiones cubiertas en el genoma de *S. paradoxus* mediante el alineamiento competitivo.
3. Clasificar los genes introgresados según su prevalencia en los aislados de agave.
4. Explorar la existencia de categorías metabólicas funcionales enriquecidas en el grupo de genes introgresados.

5. Metodología

5.1 Muestreo y aislamiento de cepas de *Saccharomyces* de fermentaciones de agave

El muestreo que se presenta en este trabajo fue realizado en el contexto del consorcio Yeast Genomes MX fruto de una colaboración de la Dra. Lucía Morales, tutora de este trabajo, con los grupos de investigación dirigidos por los doctores Alexander de Luna Fors y Eugenio Mancera Ramos adscritos al LANGEBIO y al CINVESTAV Irapuato respectivamente. Con esta colaboración se buscó ampliar la diversidad de los genomas provenientes de muestras de fermentaciones de agave, ya que anteriormente sólo se contaba con representantes secuenciados de Tamaulipas y Jalisco. Las fermentaciones en México se llevan a cabo en una amplia gama de ecosistemas y climas en 13 estados de la república que abarcan 7 regiones distintas del país (**Figura 5.1**). Los resultados presentados en este trabajo incluyen a muestras de las 7 regiones ilustradas en la Figura 5.1.

1. **Noroeste Lejano:** Sonora.
2. **Noroeste:** Durango.
3. **Noreste:** Tamaulipas.
4. **Oeste:** Jalisco y Nayarit.
5. **Altiplano Potosino-Zacatecano:** Guanajuato, San Luis Potosí y Zacatecas.
6. **Cuenca del Balsas:** Estado de México, Guerrero y Michoacán.
7. **Centro-Sur:** Oaxaca y Puebla.

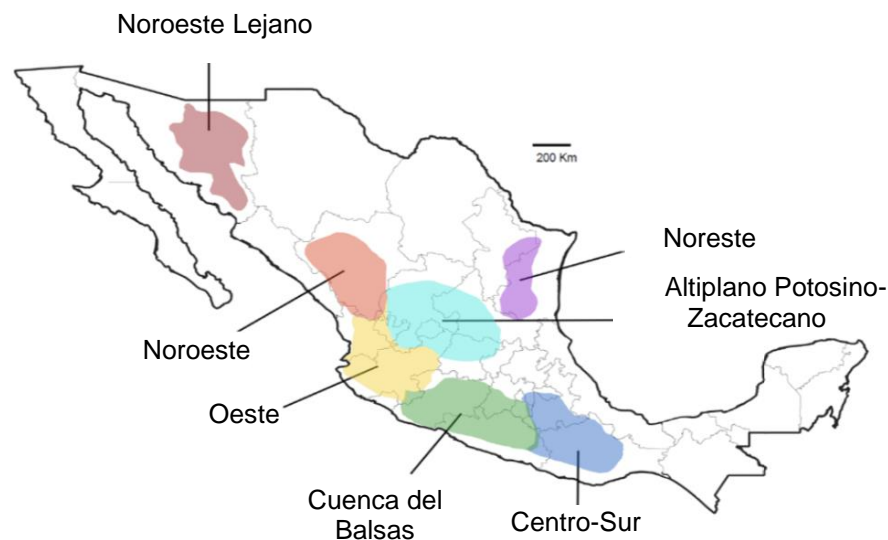


Figura 5.1: Mapa de México con la distribución geográfica de zonas productoras de agave. En este mapa se muestran las distintas zonas geográficas productoras de agave, de cada una de estas zonas se tomaron muestras para analizar. La imagen fue creada por Porfirio Gallegos, miembro del consorcio Yeast Genomes MX

Posterior al muestreo, en los laboratorios de nuestros colaboradores, se realizó un enriquecimiento en el que se buscaba aislar cepas del género *Saccharomyces* siguiendo el protocolo de Liti y colaboradores (2017), que consiste en añadir cloranfenicol para prevenir la propagación de bacterias y etanol para favorecer el crecimiento de *Saccharomyces*. Finalmente, se utilizó MALDI-TOF (desorción/ionización láser asistida por matriz) para identificar las cepas aisladas mediante espectrometría de masas del perfil proteico. El objetivo de esta caracterización fue el de identificar aquellas cepas pertenecientes al género *Saccharomyces*. En el contexto de este trabajo, es importante mencionar que debido a la resolución de la técnica y a la base de datos con la que se comparaban los perfiles de masas no fue posible distinguir con certeza entre *S. cerevisiae* y *S. paradoxus*; para este efecto y teniendo que dentro de los objetivos de este proyecto colaborativo está el análisis poblacional, se decidió recurrir a la secuenciación del genoma completo para un subconjunto de cepas clasificado como *Saccharomyces* por MALDI-TOF.

5.2 Selección de muestras usadas en este estudio

El consorcio Yeast Genomes MX ha aislado más de 1000 cepas de levadura provenientes de fermentaciones de agave. Sin embargo, en este estudio se analizaron únicamente 142 aislados intentando tener al menos un aislado de *Saccharomyces* de cada uno de los puntos de colecta visitados, logrando así representación de las diferentes regiones geográficas de México. La asignación de estos aislados a la familia *Saccharomyces* se realizó por medio de espectrometría de masas y posteriormente sus genomas se obtuvieron utilizando distintas tecnologías de secuenciación masiva en paralelo.

Adicionalmente se usaron para el análisis 170 genomas de *S. cerevisiae* previamente secuenciados y analizados por Peter y colaboradores (2018), estos aislados representan clados de todo el mundo y se utilizaron con el fin de poner en un contexto global a los aislados mexicanos.

Finalmente, se introdujeron otros cuatro aislados híbridos de Gallone y colaboradores (2019) con el fin de entender como se ven los análisis de aislados híbridos bajo la metodología aquí empleada.

5.3 Secuenciación de genomas

Para la extracción de ADN de los aislados identificados como *Saccharomyces* por MALDI-TOF se utilizó el kit MasterPure Yeast DNA. Posteriormente se mandó a secuenciar el ADN genómico en dos lotes distintos utilizando tecnologías de secuenciación distintas, el primer lote utilizó las plataformas MiSeq y NextSeq de Illumina, y el segundo utilizó la plataforma DNBSeg del Beijing Genome Institute (BGI). La secuenciación en BGI produce resultados similares a los de Illumina (Korostin, *et al.*, 2020) pero el costo es menor, razón por la cual se decidió secuenciar a la mayoría de las muestras utilizando esta tecnología.

Cabe mencionar que al momento de mandar a secuenciar se introdujeron aproximadamente 10% de cepas clasificadas por MALDI-TOF como *Kluyveromyces marxianus* para poder identificar la posible existencia de contaminación cruzada.

5.4 Alineamiento a referencias concatenadas

Como primer paso se elaboró un archivo con una referencia de genomas concatenados que incluía a las 7 especies del género *Saccharomyces* más otras dos referencias de levaduras más lejanas de las cuales se secuenciaron algunos genomas en el mismo lote de secuenciación, *Kluyveromyces marxianus* y *Pichia kudriavzevii*. Estas otras dos levaduras fueron usadas en otros proyectos y no son objeto de esta tesis. Para elaborar la referencia concatenada se descargaron los genomas de referencia de cada especie a utilizar (**Tabla 5.1**) de bases de datos públicas (yeastgenome.org, https://yjsx1217.github.io/Yeast_PacBio_2016/ y www.ncbi.nlm.nih.gov), posteriormente se dio formato a cada genoma de referencia de modo que los encabezados indicaran la especie y contig. Finalmente se concatenaron las referencias en un solo archivo de texto utilizando el comando **cat** como quedó plasmado en el archivo README de la carpeta ymez/data/ref/CONC_Mez1_v1.

Para todos los aislados descritos en la sección anterior (ver sección **5.2 Selección de muestras usadas en este estudio**), se descargaron lecturas crudas provenientes de la secuenciación del ADN genómico. Para los aislados pertenecientes a Peter y colaboradores (2018), los archivos fastq fueron descargados de Sequence Read Archive (SRA). Mientras que las lecturas obtenidas mediante Illumina fueron descargadas del servicio de secuenciación del Instituto Nacional de Medicina Genómica (INMEGEN). Por último, las lecturas secuenciadas por BGI fueron proporcionadas y descargadas mediante Amazon Web Services (AWS). Una vez que los archivos de las lecturas se descargaron en el servidor del LAVIS del campus Juriquilla de la UNAM, se utilizó la herramienta fastp v0.20.0 (Chen, *et al.*, 2018) para filtrar y eliminar las lecturas de baja calidad de cada uno de los aislados.

Tabla 5.1: Genomas en la referencia concatenada. Referencias genómicas utilizadas para elaborar la referencia concatenada, se especifica la cepa y el número de contigs de la referencia. La abreviación será utilizada en adelante.

Especie	Cepa	Abreviación	Contigs
<i>Saccharomyces cerevisiae</i>	S288C	SACE	17
<i>Saccharomyces paradoxus</i>	YPS138	SAPA	17
<i>Saccharomyces mikatae</i>	IFO 1815	SAMI	18
<i>Saccharomyces jurei</i>	M1	SAJU	17
<i>Saccharomyces arboricola</i>	H6	SAAR	17
<i>Saccharomyces uvarum</i>	CBS7001	SAUV	16
<i>Saccharomyces eubayanus</i>	FM318	SAEU	17
<i>Kluyveromyces marxianus</i>	DMKU3-1042	KLMA	9
<i>Pichia kudriavzevii</i>	CBS573	PIKU	6

Teniendo las lecturas limpias, estas se alinearon a la referencia concatenada, previamente descrita, utilizando `bwa-mem v0.7.8` (Li, 2014), posteriormente se marcaron las lecturas duplicadas con `picard v2.6.0` (<https://broadinstitute.github.io/picard/>) para obtener los archivos tipo bam con los alineamientos finales. Los scripts que realizan estos procesos se encuentran el cluster del LAVIS UNAM Campus Juriquilla en la carpeta *Public* (`/mnt/Timina/Imorales/Public`) del grupo de Imorales en la siguiente ruta: `ymez/bin/SGE/03_mapping/[!DdeMuestra]_vs_CONC.sge`.

5.5 Asignación a especie de los genomas secuenciados

Se utilizó una combinación de scripts de R y de Python v3.6 (`make_depth_plot.R` y `reference_percentage.py`) para determinar a que especie de la referencia concatenada pertenecía cada aislado. Como primer paso, se calculó la profundidad en la cobertura de cada base presente en la referencia, utilizando la herramienta `depth` de `samtools v1.10` (Li & Durbin, 2009). Después, se calculó la media de la cobertura en ventanas no sobrelapantes de 1 kb filtrando las lecturas por calidad de mapeo (MAPQ) de 0, 20 y 30. La profundidad de la cobertura por ventanas para cada filtro de calidad, se graficó utilizando R (`make_depth_plot.R`), lo que permitió identificar visualmente a que referencia se alineaban las lecturas del aislado, y asignarle una especie, o dos, en el caso de los híbridos. De igual forma se obtuvo el porcentaje de lecturas que se alinearon a cada referencia utilizando R (`reference_percentage.py`) y se determinó que la especie a la que pertenecía cada aislado en función al genoma con mayor porcentaje de lecturas alineadas. En el caso de los híbridos la señal se repartía en los dos subgenomas parentales.

5.6 Cálculo de la profundidad en la cobertura

Posteriormente, a partir de los archivos con el promedio de la cobertura calculado en ventanas de 1 Kb y aplicando el filtro de MAPQ=20, se calcularon por *contig* y por referencia las siguientes estadísticas: el total de lecturas alineadas, la media, la mediana, el mínimo, el máximo y los cuantiles 25 y 75, de la cobertura; además del porcentaje del genoma de referencia cubierto a más de 10 X (`make_depth_plot.R`).

La cobertura se normalizó dividiendo la media de cada ventana de 1 kb entre la media de cobertura del genoma completo de la referencia más cubierta, y nuevamente se obtuvieron el total de lecturas alineadas, la media, la mediana, el mínimo, el máximo y los cuantiles 25 y 75, de la cobertura por *contig* (`above5_coverage.R`).

Se elaboró un archivo de resumen utilizando Python v3.6 (Rossum & Drake, 2010) (`reference_percentage.py`) con el porcentaje de lecturas alineadas hacia cada referencia de cada aislado y una gráfica utilizando R (R Core Team, 2021) (`percentage_reference.R`) con estos porcentajes.

5.7 Determinación de introgresiones

Para cada aislado se determinó que genes estaban cubiertos en cada una de las referencias a partir del alineamiento competitivo de las lecturas. Lo anterior se obtuvo con scripts de Python

v3.6 (`overlapping_genes.py`) que relacionaban las bases cubiertas según el archivo tipo *bam* del alineamiento con las coordenadas y la función asociada de los genes de cada referencia según su archivo de anotación. Este paso es dependiente de la calidad de la anotación, siendo los genomas mejor estudiados aquellos que cuentan con archivos de anotación más completos (genes codificantes, centrómeros, telómeros, regiones subteloméricas, tRNAs, snoRNAs, etc.), mientras que otros organismos que han sido menos estudiados contienen menos información.

Una vez que se identificaron los genes de *S. cerevisiae* y *S. paradoxus* cubiertos en los aislados de fermentaciones de mezcal (este trabajo) y de Peter y colaboradores (2018), se utilizaron los scripts de Python v3.6 (`overlapping_genes.py`) para determinar las zonas con cobertura dentro de *S. paradoxus* que sugieren introgresiones. Para determinar cuáles fueron los genes cubiertos en *S. paradoxus* se aplicaron los siguientes filtros:

- Filtro 1. Debe tener un ortólogo 1:1 en *S. cerevisiae* y la divergencia entre estos debe ser igual o mayor al 5%.
- Filtro 2. Debe tener un ortólogo 1:1 en *S. cerevisiae* y ser el ORF con mayor porcentaje de lecturas alineadas entre los dos ortólogos o en su defecto debe estar cubierto por al menos el 25% del número de lecturas que cubren a su ortólogo en *S. cerevisiae*.
- Filtro 3. Debe tener un ortólogo 1:1 en *S. cerevisiae* y al menos la mitad del ortólogo en *S. paradoxus* debe de estar cubierto con por lo menos el 25% de la media de la profundidad en la cobertura del genoma completo.
- Filtro 4. Los filtros 1, 2 y 3.
- Filtro 5. Si tiene un ortólogo 1:1 en *S. cerevisiae*, entonces la divergencia entre estos debe ser de igual o mayor al 5%.
- Filtro 6. Si tiene un ortólogo 1:1 en *S. cerevisiae*, entonces ser el ORF con mayor porcentaje de lecturas alineadas entre los dos ortólogos o en su defecto debe estar cubierto por al menos el 25% de número de las lecturas que cubren a su ortólogo en *S. cerevisiae*.
- Filtro 7. Si tiene un ortólogo 1:1 en *S. cerevisiae*, entonces al menos la mitad del ortólogo en *S. paradoxus* debe de estar cubierto con lo menos el 25% de la media de la profundidad en la cobertura del genoma completo.
- Filtro 8. Los filtros 5, 6 y 7 en conjunto.

En general los filtros se pueden dividir en dos, la primera mitad contempla que de manera obligatoria las introgresiones deben de tener ortólogos para que al momento de buscarse se registren como una introgresión. Mientras que la segunda mitad considera opcional las relaciones de ortología.

Los filtros y el número de ORFs en distintos casos se encuentran plasmados en la **Tabla 5.2**.

Tabla 5.2: ORFs introgresados según los distintos filtros. Número de ORFs introgresados utilizando distintos filtros para definir las introgresiones.

	Filtro									
	Sin filtros	Tiene ortólogo en SACE	Filtro 1	Filtro 2	Filtro 3	Filtro 4	Filtro 5	Filtro 6	Filtro 7	Filtro 8
ORFs introgresados	790	558	528	522	437	743	730	599	423	577
ORFs presentes en un solo aislado	171	139	134	123	103	159	140	116	101	113
ORFs presentes en todos los aislados	1	1	1	0	1	1	0	1	0	0
ORFs compartidos por el 90% de los aislados	89	26	23	24	21	85	78	54	17	49
ORFs compartidos por el 85% de los aislados	112	31	31	31	27	105	100	69	24	66
ORFs compartidos por el 75% de los aislados	138	47	46	47	40	129	125	85	36	80

Se determinó que lo más acertado para identificar las introgresiones y eliminar los posibles alineamientos espurios por similitud entre secuencias era aplicar el filtro 8 (**Tabla 5.2**). No es necesario que los genes introgresados tengan un ortólogo en *S. cerevisiae* pero de tenerlo, este debe de tener al menos 5% de divergencia nucleotídica, al menos el 25% de las lecturas que cubren al ortólogo con mayor cobertura deben de cubrir al de *S. paradoxus*, y al menos la mitad del ortólogo en *S. paradoxus* debe de estar cubierto con por lo menos el 25% de la media de la profundidad de la cobertura del genoma completo.

5.8 Asignación funcional de genes introgresados

Una vez que se determinó qué regiones con cobertura se contarían como introgresiones, se realizó un análisis de ontología génica sobre los genes introgresados (n=790) para determinar si es que existe enriquecimiento de alguna categoría funcional específica. Debido a que los genes de *S. paradoxus* no tienen categorías GO, se tomaron los homólogos en *S. cerevisiae* de los genes introgresados y sus categorías GO asociadas reduciendo el universo de los genes analizados a n=558.

La ontología genética provee una metodología para nombrar a los genes, a la vez que describe los atributos de dicho gen. Las categorías ontológicas se dividen en tres grupos generales: función molecular, proceso biológico y localización por componente celular.

Para el análisis de ontología génica se utilizó la herramienta web GOrilla (Eden, *et. al.*, 2009), con el método de dos listas sin ordenar, los valores por defecto y un p-valor de 10^{-3} . Se tomaron los genes introgresados compartidos por el 75% de las cepas, estos se utilizaron como “*target*” y la lista completa de genes de *S. paradoxus* como “*background*”.

6. Resultados

6.1 Colinealidad y divergencia entre los genomas de *S. paradoxus* y *S. cerevisiae*

El objetivo principal de este trabajo es identificar la cantidad y naturaleza de las regiones introgresadas en los genomas de *S. cerevisiae* aislados de fermentaciones de agave. Existen varias aproximaciones para identificar a los genes introgresados. Algunas de ellas requieren identificar marcadores que nos permitan diferenciar si un fragmento de ADN proviene de una especie o de otra, en nuestro caso diferenciar entre las especies *S. cerevisiae* y *S. paradoxus*. En el marco de mi tesis empecé a explorar esta vía usando dos herramientas para determinar cuales son las coordenadas genómicas equivalentes y la magnitud de conservación en la sintenia. La primera herramienta utilizada fue NUCmer (Kurtz, *et al.*, 2004) dándole como input la referencia de *S. cerevisiae* S288C y de *S. paradoxus* la referencia de la cepa YPS138. Se usó esta referencia de *S. paradoxus* porque es la que se ha descrito como presente en el continente americano y además por ser la más cercana al genoma parental de *S. paradoxus* de los híbridos encontrados como parte de este proyecto colaborativo. NUCmer alinea ambas secuencias de ADN para encontrar regiones idénticas. Dicho alineamiento hizo evidente la presencia de colinealidad entre estos genomas, mostrando más de 6.5 Mb alineadas equivalentes a 60% de ambos genomas aproximadamente, dichos alineamientos tienen una longitud promedio de 6 Kb y están presentes a lo largo de todos los cromosomas siendo separados en promedio por 4.5 Kb, abarcando la totalidad de ambos genomas como lo muestra la línea azul de la **Figura 6.1**.

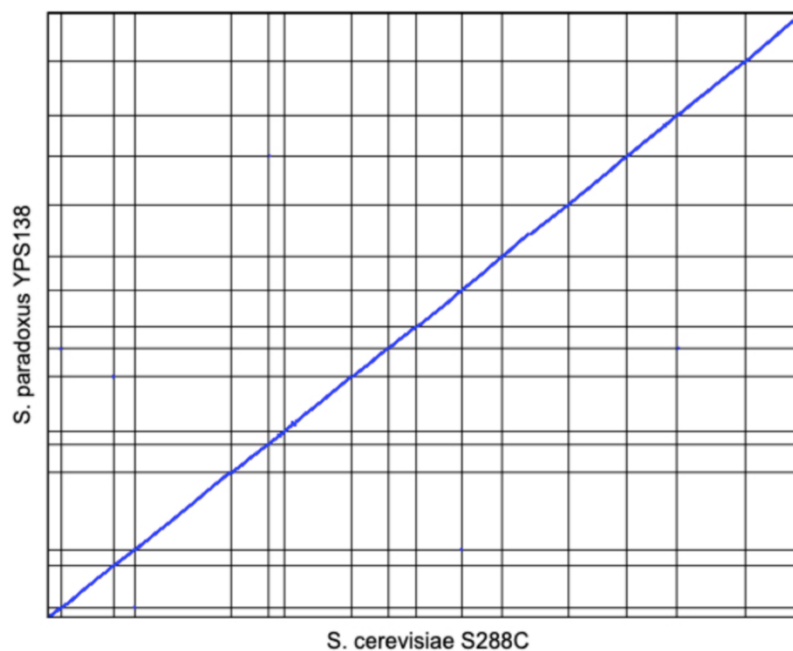


Figura 6.1: Alineamiento producido por NUCmer. Visualización del alineamiento entre los genomas de *S. cerevisiae* y *S. paradoxus*. Se aprecia que la línea azul es continua a lo largo de los 16 cromosomas de ambos genomas, mostrando el alto grado de colinealidad entre ambas especies.

Para la segunda estrategia se empleó la herramienta SynChro (Drillon, *et al.*, 2014), la cual utiliza las secuencias de aminoácidos para localizar los genes y definir los bloques de genes sinténicos conservados. SynChro reveló la presencia de bloques sinténicos de gran tamaño, abarcando prácticamente la totalidad de los genomas de ambas especies dentro de bloques sinténicos, tanto a nivel nucleotídico como a nivel de genes (Tabla 6.1, Figura 6.2) y compartiendo una identidad promedio del 92.8 % a nivel de pares de bases.

Tabla 6.1: Estadísticas de sintenia obtenidas por SynChro. Porcentaje de los genomas de *S. cerevisiae* y *S. paradoxus* conservados en bloques sinténicos, así como su identidad promedio.

	Bloques sinténicos		
	%Nucleótidos	%Genes	%Identidad
SACE	97.11	99.02	92.80
SAPA	97.66	97.30	

Por un lado, NUCmer nos revela la presencia de regiones homólogas sinténicas en el 60% del genoma de las dos especies, y por el otro, SynChro nos indica que son prácticamente idénticos, la diferencia radica en que NUCmer busca alineamientos idénticos en el ADN de todo el genoma distribuidos en la secuencia de los cromosomas en regiones inter y intra génicas, mientras que SynChro busca similitudes sólo entre los genes codificantes y su localización, aunque no sean completamente iguales.

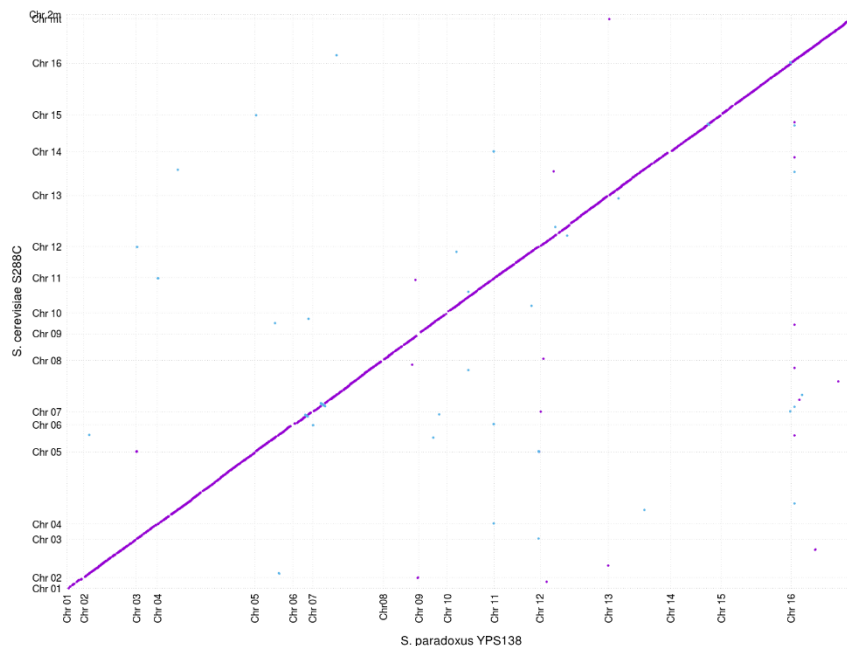


Figura 6.2: Bloques sinténicos producidos por SynChro. Visualización de los bloques sinténicos compartidos entre *S. cerevisiae* y *S. paradoxus*. La línea morada indica que ambos genomas tienen una fuerte sintenia conservada.

Ambos análisis confirman que pese a la divergencia de las secuencias de 7.2% en promedio, la sintenia está altamente conservada (**Figura 6.1** y **Figura 6.2**). Esta conservación de sintenia es relevante al estudiar el posible origen de las introgresiones ya que esta colinealidad es la que permite la hibridación y el intercambio genético mediado por micro-homología entre los cromosomas de ambas especies. Los datos generados en estos análisis además de que serán usados en trabajos posteriores para evaluar un segundo método de detección de introgresiones, nos permitieron considerar al alineamiento competitivo hacia una referencia concatenada como herramienta sencilla para detectar introgresiones.

6.2 Descripción de los aislados secuenciados

En el muestreo realizado por el consorcio Yeast Genomes MX se encontraron a más de 1000 aislados de levaduras asociados con la fermentación de agave y clasificados por espectrometría de masas como pertenecientes al género *Saccharomyces*. Con el objetivo de conocer cuáles son las características genómicas específicas de estas cepas, y en particular para este trabajo cuántos, y cuáles son los genes introgresados, seleccionamos entre esos mil aislados a 142 representantes de las diferentes regiones geográficas de México para secuenciar su genoma. De las cepas seleccionadas 120 fueron secuenciadas utilizando la tecnología de DNBSeg™ de BGI, y 22 utilizando las plataformas de secuenciación de Illumina (7 utilizando Illumina MiSeq y 15 utilizando Illumina NextSeq; **Tabla 6.2**).

Tabla 6.2. Número de genomas analizados. En esta tabla se muestra la cantidad de genomas que secuenciamos para este trabajo separados por tecnología de secuenciación, aunado al número de genomas derivados de otros trabajos que usamos para complementar nuestro análisis.

Yeast Genomes MX BGI	Yeast Genomes MX Illumina	Peter <i>et al.</i>	Gallone <i>et al.</i>	Total
120	22 (15 NextSeq + 7 Miseq)	170	4	316

A pesar de haber usado distintas tecnologías de secuencia como se muestra en la tabla anterior, ya ha sido publicado con anterioridad que al ser técnicas de secuenciación similares los resultados de los análisis no se ven afectados por esta variable (Korostin, *et al.*, 2020). En promedio se generaron 2.26 Gb de secuencia por genoma dando una cobertura media de $188X \pm 33.98$ por aislado. Las coberturas entre aislados son similares, exceptuando los 7 casos secuenciados con la plataforma Illumina MiSeq, en cuyo caso la cobertura resultó ser menor (67.06 ± 68.91) a la del resto de los aislados secuenciados mediante BGI u otras plataformas de Illumina.

Además, se utilizaron 170 secuencias de *S. cerevisiae* previamente descritas en el artículo de Peter y colaboradores (2018), estas cepas pertenecen a aislados de diferentes partes del planeta y representan a los principales clados de *S. cerevisiae* previamente descritos (Peter, *et al.*, 2018); hay que destacar que la totalidad de los aislados de *S. cerevisiae* pertenecientes a cepas mexicanas de agave (9 en total) fueron incluidas dentro de las 170 secuencias adicionales de

Peter y colaboradores (2018). Estas muestras están cubiertas en promedio a una profundidad de 155 X. Por último, se incluyeron las secuencias de 4 cepas pertenecientes a híbridos interespecie de *Saccharomycotina* previamente descritas por Gallone y colaboradores (2019), para contar con un total de 316 secuencias (**Tabla suplementaria 1**).

6.3 Asignación a especie mediante alineamiento de lecturas.

Cuando elegimos a los 142 aislados a secuenciar, sólo teníamos la certeza de que pertenecían al género *Saccharomyces*, pero ignorábamos qué especie eran o si eran híbridos. Con el objetivo de identificar la especie (o especies en el caso de los híbridos) a la que pertenecía cada aislado, al igual que encontrar las introgresiones, las lecturas obtenidas de la secuenciación de cada cepa se alinearon a un archivo con distintos genomas de referencia (**Tabla 5.1**) mediante un mapeo competitivo. Este proceso permitió que cada lectura se alineara al genoma con el que compartía mayor similitud, y así determinar la especie o especies a las que pertenecía. Gracias al análisis de divergencia realizado en la sección anterior de este trabajo sabíamos que para ciertas regiones era posible distinguir entre secuencias provenientes de *S. cerevisiae* o de *S. paradoxus*, si usábamos filtros en la calidad de mapeo. Los alineamientos competitivos fueron filtrados por calidad de mapeo (MAPQ), parámetro que disminuye cuando una lectura alinea con buena calidad a más de una secuencia en la referencia. Usamos el filtro de MAPQ > a 20, que nos garantiza que en promedio el 99% de los alineamientos sean reales, disminuyendo la cantidad de alineamientos ambiguos o espurios, ya que, en este trabajo, nos basamos en ellos para identificar las introgresiones. En este contexto, es importante mencionar que las especies del complejo *Saccharomyces*, presentes en el archivo de referencias concatenadas (**Tabla 5.1**), tienen sintenia conservada y regiones de alta similitud (ver sección Colinealidad entre los genomas de *S. paradoxus* y *S. cerevisiae*). Además, las especies de este complejo son capaces de aparearse e intercambiar material genético a través de procesos como la hibridación (Giannakou, *et al.*, 2020). Se ha observado que algunas cepas de *S. cerevisiae* contienen introgresiones genómicas provenientes de *S. paradoxus* (D'Angiolo, *et al.*, 2020; Barbosa, *et al.*, 2016), por lo que si realizábamos los mapeos de forma independiente a cada genoma de referencia, alineamientos de secuencias provenientes de un genoma forzadas a alinearse a otro provocarían lecturas pobremente alineadas.

Tabla 6.3: Asignación de especies a aislados secuenciados para este trabajo. Identificación de cepas secuenciadas para el proyecto de investigación del cual forma parte esta tesis utilizando el porcentaje de lecturas alineadas hacia cada genoma de referencia determinado, el asterisco (*) indica híbridos interespecie.

Proyecto Mezcal

Parental	No. cepas
SACE	124
SAPA	5
SACE-SAPA*	13

De los aislados mexicanos secuenciados identificamos a **124 cepas de *S. cerevisiae*** en donde más del 95% de las lecturas alinearon con alta calidad a esta referencia cubriéndola en su totalidad con una profundidad promedio de 188X; **5 cepas de *S. paradoxus*** en donde más del

95% de las lecturas de alinearon con alta calidad a esta referencia cubriéndola en su totalidad con una profundidad promedio de 133X, y finalmente **13 aislados híbridos *S. cerevisiae* x *S. paradoxus*** en donde entre el 32% y el 71% alinearon con alta calidad a genoma de *S. cerevisiae* mientras que entre el 68% y el 29% alineaba al de *S. paradoxus*. Las coberturas medias a cada genoma parental de estas cepas híbridas varían en los rangos de [50-50%], [66-33%] y [33-66%], calculando el porcentaje de lecturas alineadas a cada genoma de referencia del total de lecturas alineadas (**Tabla suplementaria 2**). Cabe recalcar que en este paso del análisis cinco aislados fueron eliminados del conjunto de datos debido a que fue imposible asignarles una especie. Estos aislados tenían lecturas alineadas hacia múltiples referencias en porcentajes altos lo cuál indicaría posibles contaminaciones, o que pertenecen a especies no incluidas en los genomas de referencia (**Tabla suplementaria 2**).

De las secuencias del resto del mundo de Peter y colaboradores (2018) se identificaron correctamente los 170 aislados pertenecientes a *S. cerevisiae*, además de identificarse correctamente los 4 híbridos de *Saccharomycotina* (Gallone, *et al.*, 2019) utilizando este método (**Tabla 6.4**).

Tabla 6.4: Asignación de especies a aislados secuenciados anteriormente. Identificación de cepas de Peter y colaboradores (2018) e híbridos de Gallone y colaboradores (2019) utilizando el porcentaje de lecturas alineadas hacia cada genoma de referencia determinado.

Cepas del mundo e híbridos

Parental	No. cepas
SACE	170
SACE-SAKU	1
SACE-SAEU	1
SAUV-SAEU	2

Tabla 6.5: Asignación de especies a aislados secuenciados. Lista de las asignaciones a especie de los aislados secuenciados agrupados por tecnología de secuenciación.

Tipo de aislado	Yeast Genomes MX BGI (n=120)	Yeast Genomes MX Illumina (n=22)	Peter <i>et al.</i> (n=170)	Gallone <i>et al.</i> (n=4)	Total (n=316)
SACE	105	19 (13 Nextseq + 6 Miseq)	170	0	294
SAPA	3	2 (1 Nextseq + 1 Miseq)	0	0	5
Híbridos	12	1 (Nextseq)	0	4	17
TOTAL	120	22 (15 NextSeq + 7 Miseq)	170	4	316

Este paso de asignación a especie resaltó dos observaciones importantes con respecto a las introgresiones: 1) Nos sugirió que entre los aislados de *S. cerevisiae* de agave existen diferentes niveles de introgresión de *S. paradoxus*; y 2) al menos 9.1% de los aislados son híbridos entre las dos especies, que de hacer retrocruzas con alguna de las especies parentales, generarían introgresiones. Con respecto al primer punto, la cepa que tuvo mayor número de lecturas alineadas con alta calidad al genoma de *S. paradoxus* fue la XA121c21 con 4.9 %, en contraste la que menos tuvo fue XB057c7 con 0.3 %. De las 124 cepas de *S. cerevisiae* mencionadas en el párrafo anterior 55 (40.3%) se encuentran en la categoría entre 3% y 5% de lecturas alineadas a *S. paradoxus*; 68 (54.8%) en la categoría entre el 2% y el 3%; finalmente 6 (4.8%) en la categoría con menos del 2 %. Si contrastamos estos porcentajes con los aislados del mundo usados en este análisis en donde la mayoría 117 (68.4%) caen en la categoría con menos del 1 % de lecturas alineadas al genoma de *S. paradoxus*, y con los aislados de alpechín, un subproducto de la producción de aceite de oliva, en donde la mayoría (85.7%) caen en la categoría entre 3% y 5%, se evidencia que en promedio los fragmentos introgresados en las cepas de agave son menos que en las de alpechín pero más que en el resto del mundo.

En contraste con lo expuesto en el párrafo anterior, los aislados que son mayoritariamente *S. paradoxus* tienen pocas lecturas que mapean al genoma de *S. cerevisiae* (0.49%, 0.52%, 1.24%, 1.39%, 1.58%).

En el caso de los aislados híbridos, 6 son mayoritariamente *S. cerevisiae* (de 62% a 71% de lecturas alineadas a *S. cerevisiae*), 4 tienen proporciones similares de ambos genomas (de 48% a 57% de lecturas alineadas a *S. cerevisiae*) y las tres restantes son mayoritariamente *S. paradoxus* (con 32% a 36% de lecturas alineadas a *S. cerevisiae*).

6.4 Análisis de cobertura

A partir del mapeo competitivo se graficó la media de la cobertura normalizada en ventanas de 1Kb, permitiendo observar qué regiones de cada genoma estaban cubiertas y a cuánta profundidad, resaltando así aneuploidías y duplicaciones del orden de kilobases. Para el caso de las introgresiones en las cepas de *S. cerevisiae*, este procedimiento permitió identificar de qué partes de la especie donadora, *S. paradoxus*, provenían éstas. Mientras que para los aislados híbridos este paso permitió explorar la estructura de los genomas destacando la proporción de las aportaciones de cada genoma parental de forma global y también las regiones que habían sufrido de pérdidas de heterocigosidad. A continuación, describiré los resultados en ambos tipos de cepas.

6.4.1 Análisis de cobertura en aislados *S. cerevisiae*

Las gráficas con las coberturas normalizadas, en los aislados de *S. cerevisiae* permitieron identificar aneuploidías y duplicaciones. Estas variantes se muestran como regiones que tienen una cobertura que se sale de la media. En este conjunto de datos se encontró una aneuploidía recurrente localizada en el brazo derecho del cromosoma IX de aproximadamente el 10% de los aislados secuenciados de *S. cerevisiae* (**Figura 6.3**).

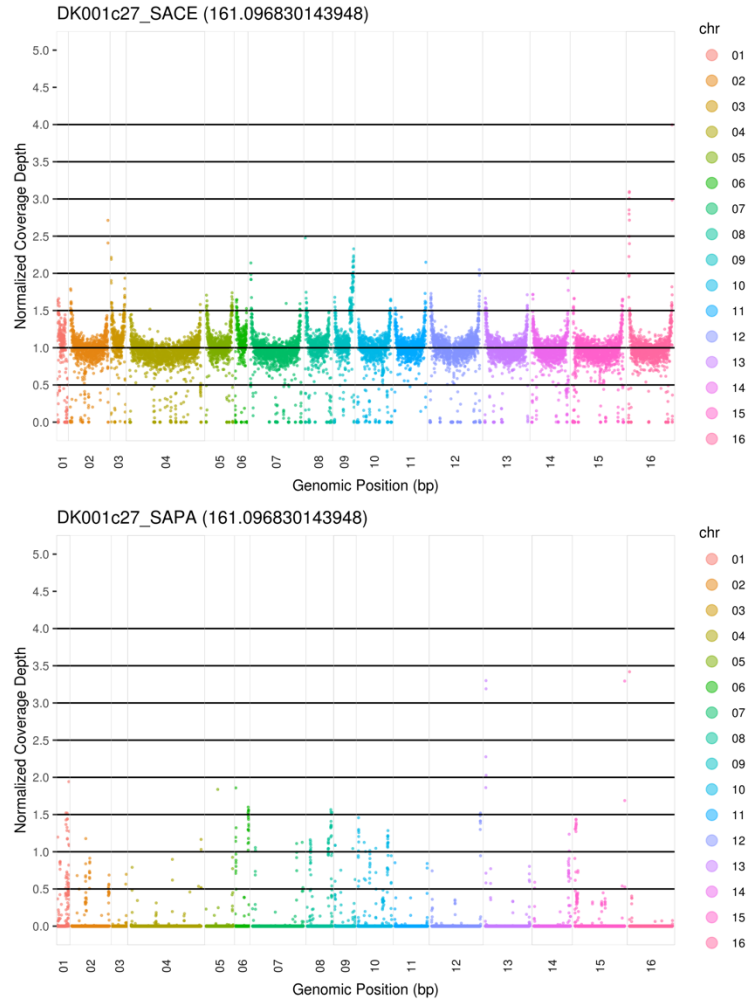


Figura 6.3 Cobertura normalizada hacia *S. cerevisiae* y *S. paradoxus* en el alineamiento competitivo para la cepa DK001c27. Se observan pocas regiones con cambios abruptos en la cobertura, siendo el cambio más notorio el brazo derecho del cromosoma IX. Se observa que la cobertura normalizada hacia la referencia a *S. cerevisiae* está alrededor de 1, mientras que a la referencia de *S. paradoxus* es mayoritariamente de 0.

En las cepas mexicanas de *S. cerevisiae* secuenciadas para este estudio se encontró que el 90% de los aislados cuentan con al menos 2% de sus lecturas alineadas hacia *S. paradoxus*. En el caso de las cepas agave mexicanas de Peter y colaboradores (2018) una tercera parte de las cepas cumplen este mismo criterio (**Figura 6.4**), en ambos casos el máximo de lecturas alineadas a *S. paradoxus* rozaba el 5%. Este hallazgo concuerda con lo reportado por Peter y colaboradores (2018).

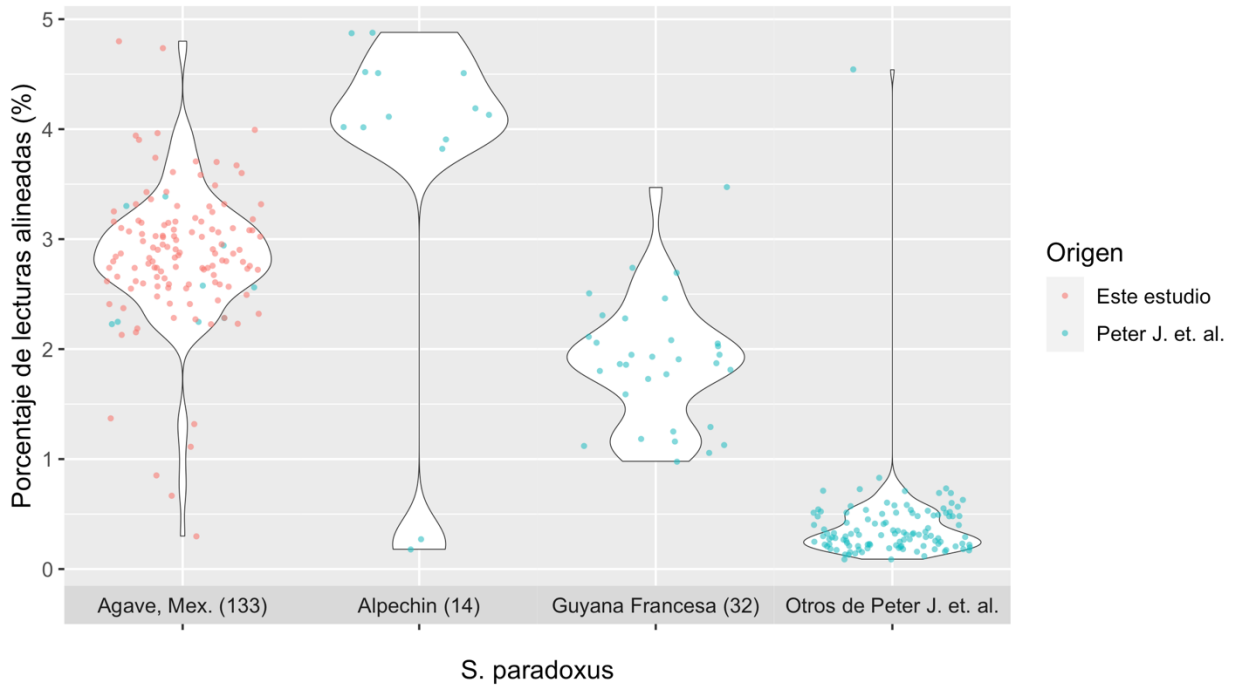


Figura 6.4: Distribución de lecturas alineadas a *S. paradoxus* de aislados de *S. cerevisiae*. Distribución del porcentaje de lecturas alineadas hacia *S. paradoxus*, en color salmón las cepas secuenciadas para este análisis, y en azul las cepas de Peter y colaboradores (2018). En general las cepas mexicanas de mezcal tienen un mayor porcentaje de lecturas mapeadas hacia *S. paradoxus*, que las cepas del resto del mundo, a excepción de los aislados de alpechin.

6.4.2 Análisis de genomas híbridos

En los aislados híbridos se identificaron regiones en las cuales la cobertura de uno de los parentales baja y en esa misma región la cobertura del otro parental aumenta, tal es el caso del asilado XA082c2 (cromosomas VIII y XII, **Figura 6.5**), lo cual podría ser un indicativo de pérdida de heterocigosis y podría ayudar a determinar la edad del híbrido.

En este estudio se identificaron 13 aislados híbridos *S. cerevisiae* x *S. paradoxus* los cuales poseen composiciones genómicas variadas, yendo desde mayoritariamente *S. cerevisiae* donde más de un tercio de las lecturas se alinearon hacia este parental, hasta mayoritariamente *S. paradoxus* donde ocurre el caso opuesto (**Figura 6.5**).

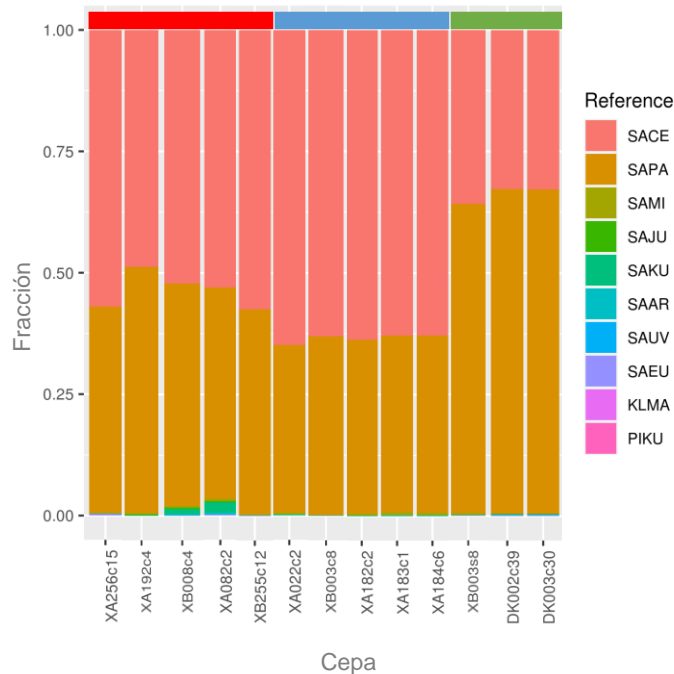


Figura 6.5: Proporción de las lecturas alineadas a cada genoma de referencia para los aislados híbridos. Proporción de lecturas alineadas hacia cada genoma de referencia para las cepas híbridas mexicanas. De los 13 aislados híbridos, únicamente 3 tienen claramente un mayor número de lecturas alineadas hacia *S. paradoxus*. Las muestras se dividen en tres grupos, en el rojo se encuentran aquellas muestras con proporción 1:1, en el azul aquellos mayoritariamente *S. cerevisiae* y en el verde aquellas que son principalmente *S. paradoxus*.

Se identificaron 13 aislados con genomas híbridos, en 5 aislados el ADN proveniente de *S. cerevisiae* está conservado en mayor medida, con una proporción de lecturas mapeadas hacia el parental de esta especie de 2:1 con respecto a *S. paradoxus* (panel azul). De igual forma se encontraron 5 aislados con proporción 1:1 (panel rojo) y solamente 3 de los aislados híbridos fueron mayoritariamente *S. paradoxus* con una proporción 1:2 (panel verde; **Figura 6.5**).

Los aislados híbridos mostraron una variedad estructural importante, habiendo aislados con coberturas bastante homogéneas en ambos parentales, como en el aislado XA082c2 (**Figura 6.6 a**) en el cual únicamente el brazo derecho de los cromosomas VIII y XIII de *S. cerevisiae* presenta el doble de cobertura mientras que su contraparte en *S. paradoxus* tiene una disminución en esta. Hay otras cepas como la XA020c1 (**Figura 6.6 b**) que poseen múltiples variaciones en su cobertura que además son de tamaño importante, esta cepa en particular posee un aumento en la cobertura del brazo derecho del cromosoma IV de *S. cerevisiae* (uno de los cromosomas más grandes) y otras variaciones en la cobertura de los cromosomas VIII, IX, XII y XV. También hay otros aislados que poseen variaciones constantes en la cobertura a lo largo de ambos genomas y cuando la cobertura aumenta en uno de los parentales disminuye en el otro parental reflejando pérdidas de heterocigosidad como ocurre en el aislado XB003c8 (**Figura 6.6 c**).

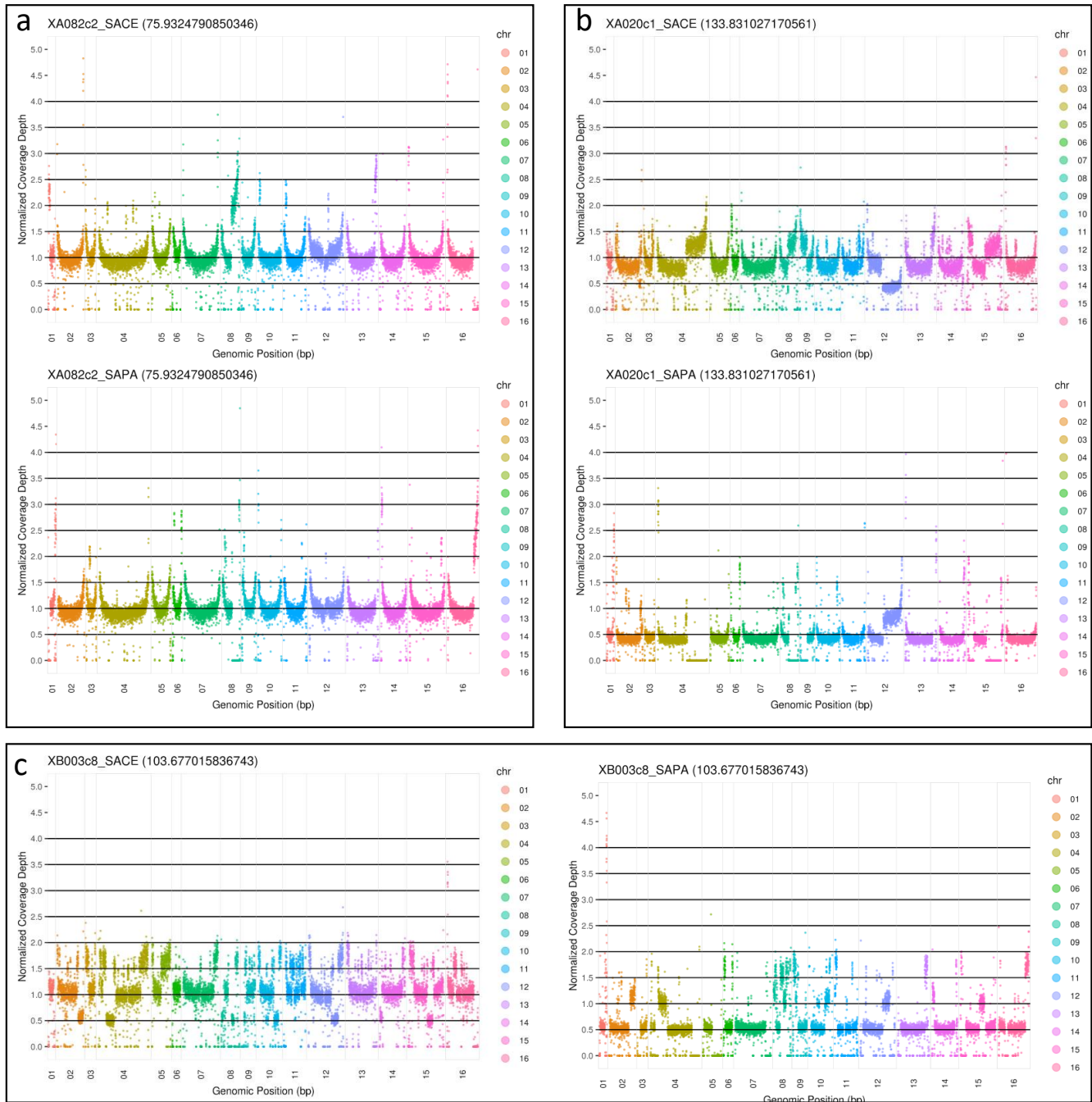


Figura 6.6: Cobertura normalizada hacia *S. cerevisiae* y *S. paradoxus* de tres distintos aislados híbridos. a) Cobertura de la cepa XA082c2 Se observan pocas regiones con cambios abruptos en la cobertura, siendo el cambio más notorio el brazo derecho del cromosoma VIII. Además, la cobertura hacia ambas referencias ronda alrededor del 1; **b)** cobertura de la cepa XA020c1. Se observan distintos cambios abruptos en la cobertura en los cromosomas IV, VIII, IX, XII y XV, además la cobertura hacia *S. cerevisiae* ronda alrededor de 1 mientras que hacia *S. paradoxus* es de alrededor de 0.5 y **c)** cobertura de la cepa XB003c8. Se observa que las coberturas hacia ambas referencias tienen un gran número de aumentos y disminuciones abruptos.

6.5 Introgresiones de *S. paradoxus*

Luego de asignar la especie a cada uno de los aislados, se hizo evidente la presencia de material genético proveniente de *S. paradoxus* dentro de los aislados identificados como *S. cerevisiae*, por lo que el siguiente paso fue explorar el origen de dicho material genético. Debido a la similitud entre los genomas de *S. cerevisiae* y *S. paradoxus* (ver sección 6.1 **Colinealidad entre los genomas de *S. paradoxus* y *S. cerevisiae***), es importante determinar cuáles zonas de *S. paradoxus* con cobertura realmente provienen de esta otra especie y no solamente lo aparentan debido a la similitud de sus genomas y a la conservación de genes homólogos.

Al aplicar distintos filtros (ver sección 5.7 **Determinación de introgresiones**), se observó que el filtro que marca una mayor diferencia es aquel que delimita los genes con cobertura dentro de *S. paradoxus* que tengan un ortólogo 1:1 dentro de *S. cerevisiae*. El pedir que tengan un ortólogo 1:1 elimina a 232 genes (**Tabla 5.2**). Luego de observar la cobertura de los genes introgresados bajo estos filtros, se decidió definir las introgresiones provenientes de *S. paradoxus* para este trabajo como: un gen que puede o no tener un ortólogo en *S. cerevisiae*. En caso de tener un ortólogo, este debe de tener al menos 5% de divergencia nucleotídica. Además, el ortólogo de *S. paradoxus* debe ser el de mayor cobertura o de lo contrario debe de estar cubierto por al menos el 25% de las lecturas que cubren su ortólogo de *S. cerevisiae*, y al menos la mitad del ortólogo en *S. paradoxus* debe de estar cubierto con por lo menos el 25% de la cobertura del genoma completo.

Para todos los aislados analizados se identificaron los genes de *S. paradoxus* que resultaron cubiertos por las lecturas de la secuenciación. Como era de esperarse, las cepas con mayor número de genes cubiertos son las que fueron identificadas como *S. paradoxus*, cada una con más de 5000 genes cubiertos. Les siguen de cerca los aislados híbridos, con entre 4200 y 5400 de genes cubiertos. Para el resto de las cepas, se puede observar que las pertenecientes al clado de alpechín de Peter y colaboradores (2018) tienen un número alto de introgresiones (203.3 en promedio), como se ha reportado con anterioridad (Peter, *et al.*, 2018; Pontes, *et al.*, 2019), seguidas de las cepas mexicanas de *S. cerevisiae*, que poseen en promedio 137.5 genes introgresados, ambos grupos son los que albergan a un mayor número de introgresiones (**Figura 6.7**).

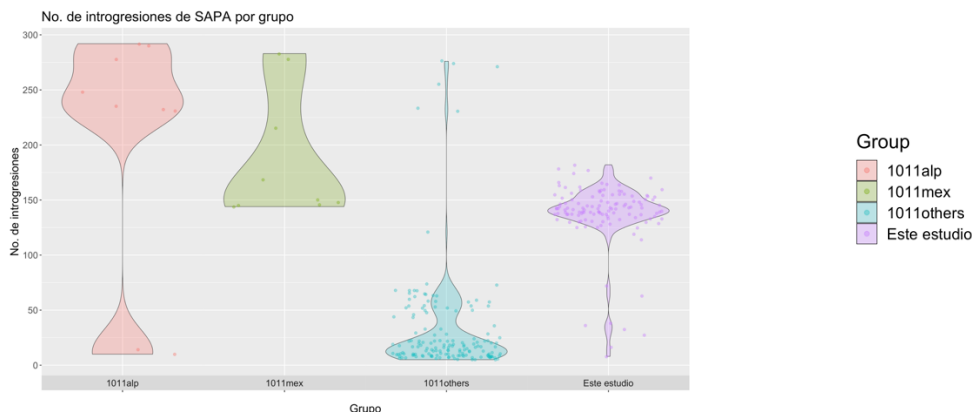


Figura 6.7: Distribución de las introgresiones de *S. paradoxus*. Distribución de genes de *S. paradoxus* introgresados en cada cepa, las cepas están separadas según su especie o secuenciación.

En los híbridos, aproximadamente el 50% de los genes están compartidos por al menos el 90% de las cepas, mientras que en el resto de los aislados ocurre lo contrario, teniendo un número reducido de genes (11%) compartidos por al menos el 90% de las cepas. Además, en los aislados híbridos el número de genes exclusivos de una cepa son muy pocos, ocurre lo contrario en el resto de los aislados donde el número de introgresiones compartidas por máximo 10% de las cepas es alto (**Figura 6.8**).

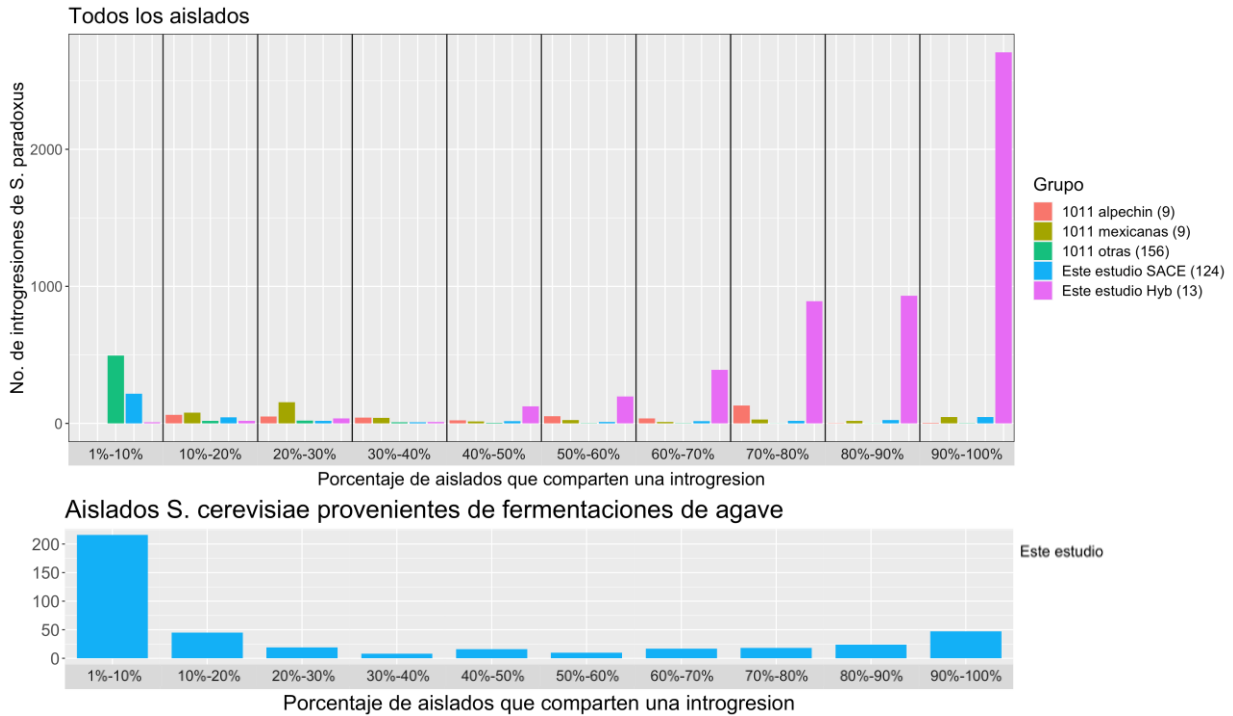


Figura 6.8: Histograma del número de introgresiones compartidas entre aislados. Histograma de introgresiones de *S. paradoxus* compartidas entre cepas, **a**: Todos los grupos de aislados y **b**: solo los aislados *S. cerevisiae* de agave. Los híbridos tienen en común la mayoría de sus introgresiones, mientras que en el resto de los aislados *S. cerevisiae* las introgresiones son compartidas entre pocos aislados.

6.6 Análisis de ontología de genes

6.6.1 Cantidad y distribución

La ontología génica realizada sobre el universo de ORFs introgresados, conformado por 577 genes, reveló el enriquecimiento de 29 categorías GO asociadas a procesos biológicos, 4 categorías GO asociadas a funciones moleculares y una categoría asociada a componentes celulares; todas las categorías GO poseen un p-valor menor a 10^{-4} (**Tabla 6.5**). A las categorías GO sobrerrepresentadas se les asociaron 17 genes (**Tabla 6.5**), entre los que se encuentran genes de las familias *THI* (*THI5*, *THI11*, *THI12*, *THI13*), *SNO* (*SNO2*, *SNO3*) y *SNZ* (*SNZ2*, *SNZ3*) que participan de manera activa en la biosíntesis de tiamina (Perli, *et al.*, 2020) (**Figura 6.9**).

Los genes sobrerrepresentados se encuentran distribuidos a lo largo de las regiones subtroméricas de los cromosomas IV, VI, VIII, X y XIV de *S. paradoxus* (Perli, *et al.*, 2020) (**Figura**

6.10), que son conocidas por su plasticidad genética y por contener genes necesarios para la interacción entre el organismo y el ambiente (Winzeler, *et al.*, 2003).

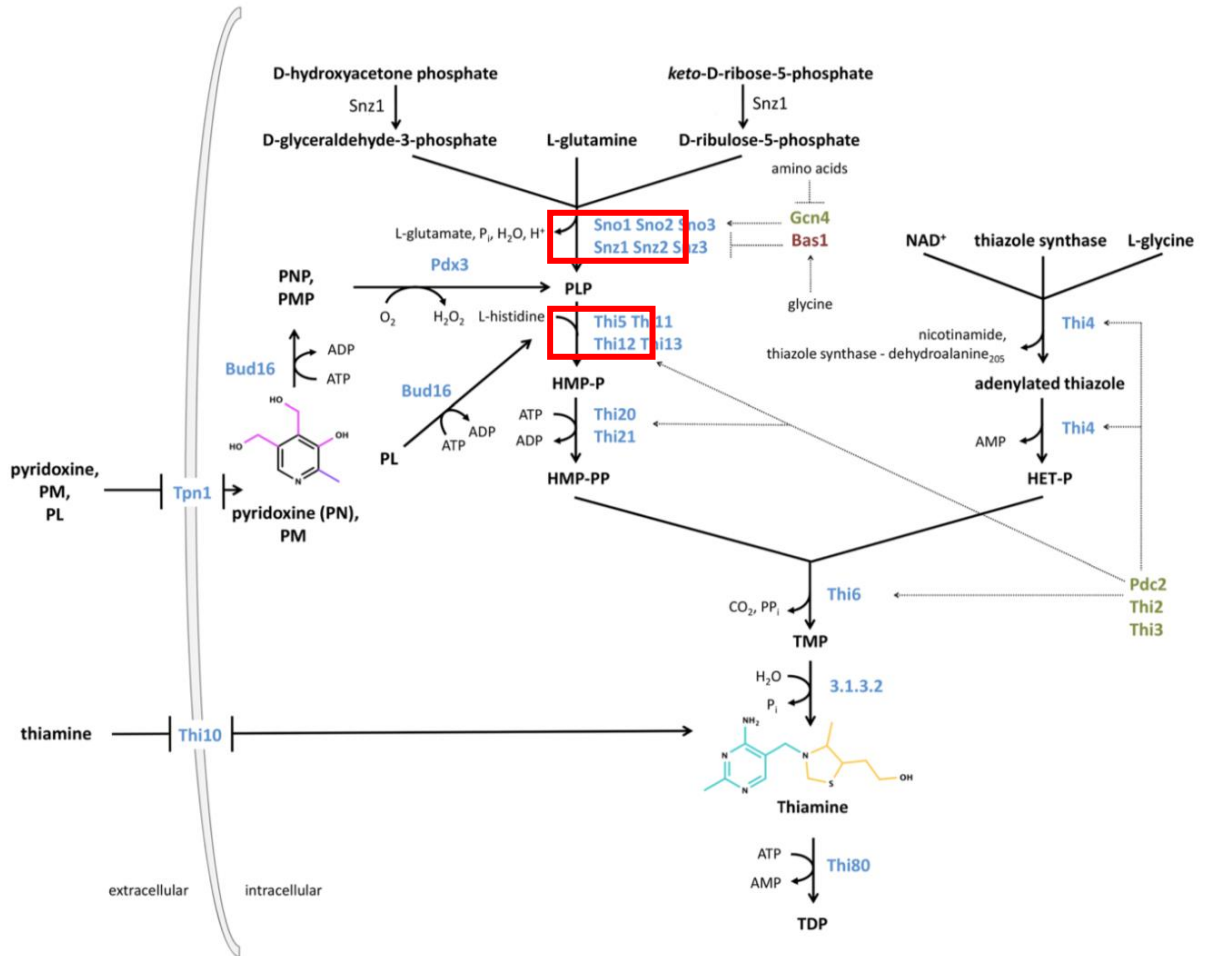


Figura 6.9: Ruta metabólica para la síntesis *de novo* de PLP (piridoxal fosfato) y TDP (difosfato de tiamina) en *S. cerevisiae*. En esta imagen, modificada de Perli, *et al.* se presenta un diagrama de la biosíntesis de vitamina B, donde se muestra resaltado en rojo la participación de los genes enriquecidos según el análisis de ontología genética (Modificado de Perli, *et al.* 2020).

Tabla 6.6: Resultados del análisis de categorías GO. Términos GO enriquecidos con los genes pertenecientes a cada término.

Proceso biológico				
Categoría	Descripción	P-valor	FDR q-valor	Genes
GO:0042724	thiamine-containing compound biosynthetic process	1.45E-09	7.50E-06	THI12, SNZ2, THI11, THI13, THI5, SNZ3
GO:0009228	thiamine biosynthetic process	1.45E-09	3.75E-06	THI12, SNZ2, THI11, THI13, THI5, SNZ3
GO:0006772	thiamine metabolic process	2.67E-09	4.60E-06	THI12, SNZ2, THI11, THI13, THI5, SNZ3
GO:0042723	thiamine-containing compound metabolic process	4.63E-09	5.98E-06	THI12, SNZ2, THI11, THI13, THI5, SNZ3
GO:0042364	water-soluble vitamin biosynthetic process	6.31E-08	6.52E-05	THI12, SNZ2, THI11, THI13, THI5, SNZ3, SNO3, SNO2
GO:0009110	vitamin biosynthetic process	6.31E-08	5.44E-05	THI12, SNZ2, THI11, THI13, THI5, SNZ3, SNO3, SNO2
GO:0044272	sulfur compound biosynthetic process	1.18E-07	8.73E-05	THI12, STR2, SNZ2, THI5, THI11, THI13, SNZ3, IRC7, BAT1
GO:0006767	water-soluble vitamin metabolic process	2.10E-07	1.36E-04	SNZ2, SNZ3, SNO3, SNO2
GO:0006766	vitamin metabolic process	2.10E-07	1.21E-04	THI12, THI11, THI13, THI5
GO:0008614	pyridoxine metabolic process	1.01E-06	5.24E-04	SNZ2, SNZ3, SNO3, SNO2
GO:0042357	thiamine diphosphate metabolic process	1.01E-06	4.76E-04	THI12, THI11, THI13, THI5
GO:0009229	thiamine diphosphate biosynthetic process	1.01E-06	4.36E-04	THI12, THI11, THI13, THI5
GO:0042822	pyridoxal phosphate metabolic process	1.81E-06	7.19E-04	SNZ2, SNZ3, SNO3, SNO2
GO:0042823	pyridoxal phosphate biosynthetic process	1.81E-06	6.67E-04	SNZ2, SNZ3, SNO3, SNO2
GO:0006790	sulfur compound metabolic process	2.30E-06	7.94E-04	THI12, SNZ2, THI11, THI13, THI5, SNZ3, BAT1, IRC7, STR2, IBA57
GO:0042819	vitamin B6 biosynthetic process	2.99E-06	9.65E-04	SNZ2, SNZ3, SNO3, SNO2
GO:0046184	aldehyde biosynthetic process	2.99E-06	9.08E-04	SNZ2, SNZ3, SNO3, SNO2
GO:0042816	vitamin B6 metabolic process	4.66E-06	1.34E-03	SNZ2, SNZ3, SNO3, SNO2
GO:0072528	pyrimidine-containing compound biosynthetic process	5.47E-06	1.49E-03	THI12, SNZ2, THI11, THI13, THI5, SNZ3
GO:0009108	coenzyme biosynthetic process	7.59E-06	1.96E-03	THI12, SNZ2, THI11, THI13, THI5, SNZ3, SNO3, SNO2, HXK1
GO:0072527	pyrimidine-containing compound metabolic process	1.24E-05	3.05E-03	THI12, SNZ2, THI11, THI13, THI5, SNZ3
GO:0051188	cofactor biosynthetic process	2.76E-05	6.47E-03	THI12, SNZ2, THI11, THI13, THI5, SNZ3, SNO3, SNO2, HXK1
GO:0017144	drug metabolic process	1.07E-04	2.40E-02	THI12, SNZ2, THI11, THI13, THI5, SNZ3, SNO3, SNO2, ATP2, HXK1, BAT1
GO:0006081	cellular aldehyde metabolic process	1.23E-04	2.64E-02	SNZ2, SNZ3, SNO3, SNO2, AAD14,
GO:0044283	small molecule biosynthetic process	1.30E-04	2.69E-02	SNZ2, SNZ3, SNO3, SNO2, THI11, THI13, THI5, THI12, HXK1, STR2, IMD2, IRC7, BAT1
GO:0090407	organophosphate biosynthetic process	1.56E-04	3.11E-02	IMD2, THI12, THI13, THI11, THI5, SNO2, SNZ3, SNO3, ATP2, HXK1
GO:0006732	coenzyme metabolic process	1.66E-04	3.18E-02	THI12, SNZ2, THI5, THI13, THI11, SNO2, SNZ3, SNO3, HXK1
GO:0072525	pyridine-containing compound biosynthetic process	3.00E-04	5.54E-02	SNZ2, SNO2, SNZ3, SNO3, HXK1
GO:0051186	cofactor metabolic process	4.35E-04	7.75E-02	THI12, SNZ2, THI13, THI11, THI5, SNO2, SNZ3, SNO3, HXK1, IBA57
Función molecular				
Categoría	Descripción	P-valor	FDR q-valor	Genes
GO:0016843	amine-lyase activity	2.21E-07	5.58E-04	SNZ2, SNO2, SNZ3, SNO3
GO:0036381	pyridoxal 5'-phosphate synthase (glutamine hydrolysing) activity	2.21E-07	2.79E-04	SNZ2, SNZ3, SNO2, SNO3
GO:0016840	carbon-nitrogen lyase activity	1.38E-05	1.16E-02	SNZ2,SNO2,SNZ3,SNO3
GO:0016829	lyase activity	9.58E-04	6.05E-01	SNZ2, SNO2, SNZ3, DDI3, SNO3, IRC7
Componente celular				
Categoría	Descripción	P-valor	FDR q-valor	Genes
GO:1903600	glutaminase complex	7.48E-04	7.89E-01	SNO2, SNO3

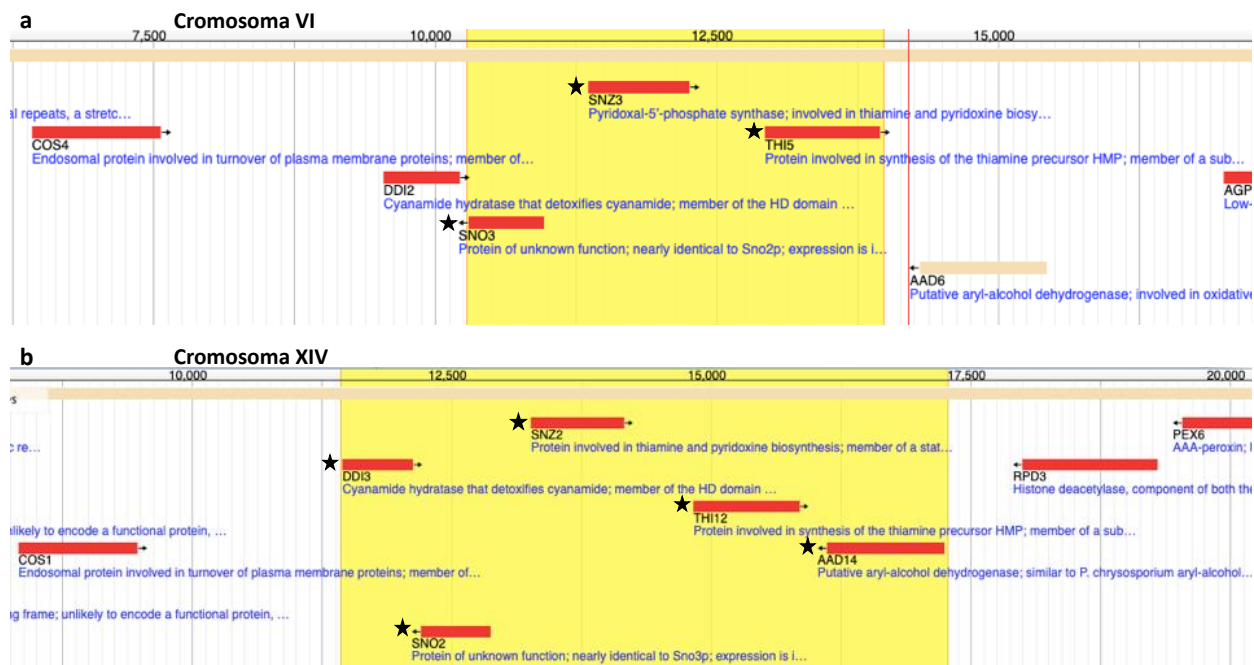


Figura 6.10: Homólogos en *S. cerevisiae* de los genes introgresados de *S. paradoxus* en categorías ontológicas enriquecidas. Vista de la ubicación de los genes homólogos en *S. cerevisiae* de los genes introgresados de *S. paradoxus* que se encuentran en categorías ontológicas enriquecidas, los genes de interés están marcados con una estrella negra y se encuentran dentro del área amarilla; solamente para los cromosomas **a**: VI y **b**: XIV.

6.6.2 Análisis de ontología de genes

Al analizar los 17 genes introgresados de *S. paradoxus* que conforman las categorías GO mencionadas anteriormente es posible ver que estos tienen varias cosas en común, comenzando por su ubicación, es posible formar 5 grupos de genes con ubicaciones comunes (B, E, F, H, I) y otros 4 genes separados espacialmente en su propio grupo (A, C, D, G). Así mismo, estos genes se pueden agrupar por funciones, siendo las principales: la biosíntesis de tiamina y piridoxina (*SNZ2/SNZ3*, grupos E,H,I), el metabolismo de la piridoxina (*SNO2/SNO3*, grupos E,H,I), la síntesis del precursor de tiamina hidroximetilpirimidina (grupos E, H, I), funciones mitocondriales (grupos D, F) y metabolismo de cisteína (grupos B, G) (**Tabla 6.6**).

Uno de los aspectos más destacables de este análisis es que estos 17 genes están presentes casi de forma exclusiva en los aislados mexicanos de agave secuenciados para este estudio, estando presentes en al menos 79% de los aislados mexicanos pero presentes en menos del 18% de los aislados del resto del mundo. El gen con mayor presencia en los aislados del mundo es el *DDI3* (hidratasa de cianamida), presente en el 96% de los aislados mexicanos y sólo en el 17.9% de los aislados del resto del mundo (**Tabla 6.6**).

Grupo	ID <i>S. paradoxus</i>	Gen en <i>S. cerevisiae</i>	Aislados de agave que comparten la introgresión (%) (n=124)	Aislados del mundo que comparten la introgresión (%) (n=161)	Anotación
E	YPS138_10G00030	SNZ2/SNZ3	96.0	1.9	Proteins involved in thiamine and pyridoxine biosynthesis; set of paralogs (SNZ1, SNZ2 and SNZ3); enables GMP synthase (glutamine-hydrolyzing) activity. SNZ2 forms a co-regulated gene pair with SNO2; interacts with Thi11p
H	YPS138_10G03310	SNZ2/SNZ3	96.0	5.8	
I	YPS138_14G00030	SNZ2/SNZ3	91.9	7.7	
H	YPS138_10G03320	SNO2/SNO3	95.2	0.6	Proteins of unknown function; set of paralogs(SNO2/SNO3); expression is induced before the diauxic shift and also in the absence of thiamin; involved in pyridoxine metabolic process
E	YPS138_10G00020	SNO2/SNO3	85.5	0.6	
H	YPS138_10G03300	THI5/THI12	95.2	10.3	Subtelomeric gene family including THI5, THI11, THI12, and THI13; Proteins involved in synthesis of the thiamine precursor hydroxymethylpyrimidine
I	YPS138_14G00040	THI5/THI12	89.5	0.6	
E	YPS138_10G00040	THI5/THI12	87.1	0.0	
A	YPS138_04G00050	THI11/THI13	82.3	0.0	
D	YPS138_08G02480	BAT1	87.1	0.0	Mitochondrial branched-chain amino acid (BCAA) aminotransferase; preferentially involved in BCAA biosynthesis, BAT1 has a paralog, BAT2, that arose from the whole genome duplication, Branched-chain Amino acid Transaminase. Branched-chain amino acid aminotransferase; involved in both the biosynthesis and degradation of branched chain amino acids, such as leucine, isoleucine and valine; localizes to the mitochondrial matrix
F	YPS138_10G02960	ATP2	85.5	0.0	Beta subunit of the F1 sector of mitochondrial F1F0 ATP synthase; which is a large, evolutionarily conserved enzyme complex required for ATP synthesis; F1 translationally regulates ATP6 and ATP8 expression to achieve a balanced output of ATP synthase genes encoded in nucleus and mitochondria; phosphorylated. Subunit of the catalytic core of the F1 sector of mitochondrial F1F0 ATP synthase; ATP synthase complex is localized to the mitochondrial inner membrane
B	YPS138_06G01190	IRC7	92.7	13.5	Cysteine desulphhydrase, enables growth on cysteine as nitrogen source; involved in the production of thiols; Increased Recombination Centers, Cysteine-S-conjugate beta-lyase involved in sulfur metabolism and L-cysteine catabolism, Expression induced by nitrogen limitation in a GLN3, GAT1-dependent manner and by copper levels
G	YPS138_10G03040	STR2	79.8	3.2	Cystathionine gamma-synthase, converts cysteine into cystathionine; STR2 has a paralog
H	YPS138_10G03330	DDI3	96.0	17.9	Cyanamide hydratase that detoxifies cyanamide; identical to DDI2 and Ddi2p, enables cyanamide hydratase activity
I	YPS138_14G00050	AAD14	91.9	0.0	Putative aryl-alcohol dehydrogenase; similar to <i>P. chrysosporium</i> aryl-alcohol dehydrogenase; mutational analysis has not yet revealed a physiological role; members of the AAD gene family comprise three pairs (AAD3 + AAD15, AAD6/AAD16 + AAD4, AAD10 + AAD14) whose two genes are more related to one another than to other members of the family
C	YPS138_08G02550	IMD2	91.1	4.5	Inosine monophosphate dehydrogenase; catalyzes the rate-limiting step in GTP biosynthesis, expression is induced by mycophenolic acid resulting in resistance to the drug, expression is repressed by nutrient limitation; IMD2 has a paralog, YAR073W/YAR075W, that arose from a segmental duplication
F	YPS138_10G02970	IBA57	89.5	0.0	Protein involved in incorporating iron-sulfur clusters into proteins; mitochondrial matrix protein; involved in the incorporation of iron-sulfur clusters into mitochondrial aconitase-type proteins; activates the radical-SAM family members Bio2p and Lip5p; interacts with Ccr4p in the two-hybrid system
B	YPS138_06G01160	HXK1	79.0	4.5	Hexokinase isoenzyme 1; a cytosolic protein that catalyzes phosphorylation of glucose during glucose metabolism; expression is highest during growth on non-glucose carbon sources; glucose-induced repression involves hexokinase Hxk2p; HXK1 has a paralog, HXK2, that arose from the whole genome duplication. Hexokinase; phosphorylates glucose, mannose, and fructose; localized in the cytoplasm but also has minor activity associated with mitochondria

Tabla 6.6. Genes introgresados asociados a las categorías GO sobrerrepresentadas agrupados por función y ubicación. Genes de *S. paradoxus* asociados a categorías GO sobrerrepresentadas agrupados por su función y ubicación dentro del genoma de referencia. Se muestra además el porcentaje de cepas mexicanas y del mundo que comparten la introgresión, y la anotación del gen, en *S. cerevisiae*, obtenida de *Saccharomyces* Genome Database (SGD, <http://www.yeastgenome.org>). Se muestran del mismo color aquellos genes que comparten ubicación y en blanco aquellos que se encuentran de manera individual, además de encontrarse agrupados los genes que poseen funciones similares.

7. Discusión

La asignación a especie utilizando el alineamiento hacia la referencia concatenada fue exitosa. El 100% de los aislados de *S. cerevisiae* tomados de Peter, *et al.* 2018 que usamos como control fueron asignados a esta especie con el método de alineamiento competitivo que implementamos en este trabajo. De igual forma los híbridos publicados en Gallone, *et al.* dieron la señal híbrida esperada (*S. cerevisiae* + *S. uvarum*, *S. eubayanus* o *S. kudriavzevii*, según fuera el caso). Cabe destacar que este método también detectó para ciertos aislados de *S. cerevisiae* un porcentaje de lecturas que alineaban a la referencia de *S. paradoxus*, como consecuencia de las introgresiones. Concluyendo que el análisis de coberturas derivado de un alineamiento a una referencia de genomas concatenados de *Saccharomyces*, hace posible la identificación de genes introgresados e híbridos interespecie.

Uno de los hallazgos más importantes de este trabajo fue descubrir que híbridos y especies parentales de *S. cerevisiae* y *S. paradoxus*, algunas con introgresiones, coexisten en el nicho de fermentación de agave. De los 142 genomas mexicanos que analizamos 87.32% (n=124) fueron de la especie *S. cerevisiae*, 3.52% (n=5) de *S. paradoxus* y 9.16% (n=13) híbridos entre estas dos especies. La presencia de *S. cerevisiae* concuerda con lo que se observó en estudios previos (Kirchmayr, *et al.*, 2017; Lappe-Oliveras, *et al.*, 2008; Garibay Marcelo, 2019), donde se describe a este como el principal organismo responsable de la fermentación de agave para la producción de mezcal. Si bien estos números no reflejan las proporciones reales del ambiente fermentativo de agave debido al enriquecimiento realizado en nuestro protocolo, sí aportan evidencia de su co-existencia en las tinas de fermentación. Los aislados de *S. paradoxus* son menos que los de *S. cerevisiae* e híbridos. Esta observación puede deberse a varias razones no excluyentes entre ellas. En primer lugar, *S. paradoxus* está asociada principalmente a ambientes naturales no domesticados como son la tierra y las cortezas de árboles del género *Quercus* (Sylvester, *et al.*, 2015) y en segundo lugar esta especie es menos tolerante a etanol que *S. cerevisiae*, y nuestro protocolo de enriquecimiento usaba etanol al 5% para enriquecer levaduras *Saccharomyces*. Posiblemente el método de enriquecimiento empleado no favorece su aislamiento. Independientemente de las proporciones en las que se encuentran estas dos especies y sus híbridos en los tanques de fermentación el hecho de haber aislado cepas de estos tres grupos nos indica que coexisten en este medio. Dicha coexistencia aunada a la presencia de genes introgresados, nos indican que la hibridación interespecie en el nicho ocurre con frecuencia y que ninguna de las tres poblaciones ha desplazado a las otras dos. En futuros estudios resultaría interesante indagar en que momento del proceso de fermentación cada una de estas tres poblaciones se divide con mayor velocidad, o si todas lo hacen de igual manera manteniendo su proporción constante y actuando en sinergia.

En este estudio se reveló por primera vez la presencia de híbridos interespecie en este nicho, aún hace falta indagar si los híbridos se forman dentro de las tinas de fermentaciones y/o en los alrededores de la destilería, y de formarse en la tina de fermentación, bajo qué condiciones lo hacen.

S. cerevisiae y *S. paradoxus* son las dos especies más cercanas de la familia *Saccharomyces* y conservan sintenia en la mayor parte de sus genomas, además son conocidas por aparearse entre sí, intercambiando material genético, estos dos factores se ven reflejados en la existencia de introgresiones de *S. paradoxus* en los aislados de *S. cerevisiae*. Una de las ventajas de la metodología aquí empleada para identificar introgresiones yace en el mapeo competitivo seguido de un filtro en la calidad de mapeo, que disminuye los alineamientos espurios y conserva únicamente las lecturas alineadas con buena calidad, eliminando artificios creados por el alineador debido a la similitud entre secuencias. Además, para la definición de genes introgresados, aplicamos un filtro de divergencia mínima entre genes homólogos con el objetivo de eliminar los problemas que pudiesen surgir debido a la similitud entre secuencias. Esto aunado al resto de filtros de cobertura garantizan que la totalidad de lo que se considera una introgresión esté bien cubierta y no solamente un segmento. Una desventaja de este método recae en la calidad de la anotación de *S. paradoxus* y en la definición de las relaciones de ortología en genes que tienen parálogos, ya que aún al aplicar los filtros de divergencia en genes ortólogos se identificaron genes cubiertos en la referencia de *S. paradoxus* asociados a más de un gen en *S. cerevisiae*. Otro punto a mencionar es que, al usar la anotación de las especies parentales para ubicar introgresiones, se asume que estas se encuentran en el mismo lugar que en las cepas de referencia, sin embargo, no tenemos información para respaldar o refutar que esto sea así, se necesitaría realizar análisis adicionales para encontrar las ubicaciones reales de las introgresiones, lo único que constatamos es la presencia de la introgresión en la cepa.

Las secuencias obtenidas de los distintos aislados tienen un patrón que hace que la cobertura aumente en los extremos de cada cromosoma inflando de manera arbitraria la cobertura que está dentro de estas regiones. Al ser esto un patrón que se encuentra a lo largo de varias de nuestras muestras no se puede asegurar que estas regiones realmente tengan un mayor número de copias que el resto del cromosoma. Esto puede provocar que los genes de regiones con cobertura arbitrariamente mayor no sean filtrados. Se necesitan otros métodos no basados solamente en cobertura para asegurarnos que los genes clasificados como introgresados en este trabajo que yacen en regiones subteloméricas en verdad tengan alelos de *S. paradoxus* y no sólo han sido clasificados así por el exceso de cobertura que hay en estas regiones.

Los aislados secuenciados de *S. cerevisiae* provenientes de fermentaciones de agave poseen una gran cantidad de material genético proveniente de *S. paradoxus*, siendo el segundo grupo con mayor cantidad de material genético proveniente de dicha especie (**Figura 6.3 y 6.10**), solo por debajo de los aislados de alpechín los cuales se caracterizan por tener un alto número de introgresiones (D'Angiolo, *et al.*, 2020). Las introgresiones son una fuente importante de variabilidad fenotípica y hay evidencia de su importancia adaptativa a nichos asociados al humano (Peter, *et al.*, 2018).

Como parte de este trabajo se determinó mediante alineamientos competitivos y cobertura qué genes estaban introgresados por cepa y como es que estos se comparten entre cepas. De un universo de 577 genes introgresados observamos que 71 están compartidos por el 80% de las cepas, mientras que la gran mayoría de las introgresiones (373) están compartidas por 10% o menos de las cepas. Que la mayoría de las introgresiones estén compartidas por tan pocas cepas

nos dice que hay una constante pérdida de material genético de *S. paradoxus* y pocos genes logran fijarse en la población general, ya sea por que confieren una ventaja adaptativa o por deriva génica. Análisis que no se detallan en este trabajo indican que las cepas de *S. cerevisiae* de agave comparten un origen monofilético, para determinar si se fijaron debido a que confieren una ventaja sería necesario realizar análisis adicionales donde se evalúe el cambio en la adecuación en un medio similar al del agave en función de su presencia o ausencia.

Una vez que determinamos el universo de genes introgresados decidimos hacer un análisis de enriquecimiento en categorías ontológicas con base en la función molecular, el componente celular y el proceso biológico. Usando los genes introgresados por el 75% de las cepas encontramos que hay un enriquecimiento en varias categorías relacionadas con la biosíntesis de tiamina y que dichas categorías están compuestas por familias de genes parálogos. Es posible que el enriquecimiento de estas categorías se deba a que el gen o los genes introgresados pertenezcan a una familia de parálogos y que el método para la detección de introgresiones utilizado sea muy sensible a la presencia de parálogos y no logre discernir correctamente. Como perspectiva de este trabajo se empleará un método para la detección de introgresiones que ponga especial atención en las familias de parálogos. De comprobarse que el enriquecimiento de las categorías génicas no se debe a la presencia de genes parálogos sería necesario realizar análisis adicionales que evalúen si la tiamina es un factor limitante en el ambiente fermentativo del agave, ya sea porque se encuentra en bajas concentraciones o porque dentro de la microbiota del jugo en fermentación otros microorganismos la estén consumiendo, y que los alelos de *S. paradoxus* brinden una mayor ventaja en bajas concentraciones de tiamina.

Dentro de los genes introgresados que componen las categorías ontológicas enriquecidas encontramos a *BAT1* y a *ATP2*, presentes en el 87% y 85% de las cepas respectivamente. Ambos genes están relacionados con funciones mitocondriales y pertenecen a familias de parálogos. El primero está relacionado con la biosíntesis de aminoácidos de cadena ramificada y el segundo con la síntesis de ATP. Análisis que no se muestran en este trabajo demuestran la presencia de secuencias de *S. paradoxus* en el genoma mitocondrial en las cepas de agave. Como perspectiva de este trabajo sería interesante analizar si estos alelos de *S. paradoxus* son predominantes debido a que resuelven incompatibilidades entre genes nucleares y mitocondriales.

Otros dos genes introgresados son *IRC7* y *STR2*, encontradas en 97% y 79% de las cepas respectivamente. Estos genes pertenecen a dos familias de parálogos y cumplen funciones relacionadas con la cisteína. El primero permite el uso de cisteína como fuente de nitrógeno y el segundo codifica para cisteína-gamma liasa. Hay evidencia que muestra al jugo de fermentación de agave como un medio deficiente en nitrógeno. Como perspectiva para este trabajo, resultaría interesante analizar si los alelos de *S. paradoxus* se conservan de manera predominante debido a que aumentan la tasa de crecimiento de las cepas en un medio deficiente en nitrógeno. Lo anterior podría realizarse de forma experimental, con análisis bioinformáticos previos en dónde se busque si las diferencias nucleotídicas entre las dos especies se encuentran en el sitio catalítico de la enzima.

Otro de los puntos a destacar es la recurrente aneuploidía del brazo derecho del cromosoma IX presente en aproximadamente el 10% de los aislados secuenciados de *S. cerevisiae*, aunque aún no se ha hecho un análisis de genes localizados dentro de esta región genómica, al ser un evento recurrente se podría pensar que dicha aneuploidía provee una ventaja adaptativa a su medio. Análisis adicionales son necesarios para hacer explorar la veracidad de esta afirmación. No obstante, se ha observado que las aneuploidías juegan un papel importante al momento de adaptarse a nuevos nichos. Para resolver esta pregunta, se requerirá delimitar con precisión la región genómica duplicada y hacer análisis adicionales para determinar genes que probablemente confieran una ventaja adaptativa al estar en mayor número de copias.

Aunque se pudiera pensar que únicamente se han preservado los haplotipos favorables mediante selección natural, es necesario recordar que esta no es la única fuerza evolutiva que está actuando, por ejemplo, no podemos descartar el efecto de la deriva génica dentro de nuestras poblaciones de *S. cerevisiae*. Recordemos que la deriva génica es más fuerte cuando el tamaño efectivo de la población es reducido (Masel, 2011) y que al ocurrir la hibridación hay un aislamiento sexual que disminuye drásticamente su capacidad para reproducirse (D'Angiolo, *et al.*, 2020; Peris, *et al.*, 2022), por lo que la fijación inicial de alelos puede ser un evento meramente fortuito y no un producto de la adaptación.

Todos nuestros aislados *Saccharomyces* tienen ya sea entre 1% y 5% de lecturas correspondientes a introgresiones o más del 33% de las lecturas alineando hacia un segundo genoma. Esta observación indica que no estamos recuperando un continuo en las proporciones del genoma de una segunda especie, en este caso *S. paradoxus*, en los aislados de fermentaciones de agave. Encontramos tanto híbridos como levaduras con introgresiones conviviendo en los tanques de fermentación, sin muestrearse etapas intermedias, algo similar ha sido reportado por D'Angiolo y colaboradores (2020) en los aislados de alpechín. Desconocemos la razón por la que individuos en estas etapas no fueron aislados. Como perspectiva de este trabajo sería interesante indagar qué pasa cuando un híbrido empieza a perder gran parte del subgenoma menor representado y cómo es que se va dando este proceso.

8. Conclusiones

Para el desarrollo de esta tesis se aislaron, secuenciaron y analizaron 142 aislados obtenidos de tinajas de fermentación de agave para la producción de mezcal ubicadas en 12 estados diferentes de la República Mexicana. Encontramos levaduras del género *Saccharomyces*, particularmente *S. paradoxus* y *S. cerevisiae*, así como otros microorganismos y destacamos la presencia de 13 híbridos interespecie *S. cerevisiae* x *S. paradoxus*.

Al enfocarnos únicamente en los aislados de *S. cerevisiae* se encontró una fuerte presencia de introgresiones de la especie *S. paradoxus* que van del 2% al 5% de las lecturas alineadas, contrastando con las cepas del resto del mundo reportadas por Peter y colaboradores (2018), las cuales exceptuando algunos nichos específicos poseen hasta 1%. De igual forma, los genes introgresados, aunque no se comparten entre todas las cepas, son en promedio más frecuentes que en las cepas del resto del mundo, a excepción de las cepas aisladas de alpechín.

El análisis de ontología genética sobre los genes introgresados reveló el enriquecimiento de genes relacionados con la biosíntesis de tiamina. Aunque aún es temprano para sacar conclusiones sobre este punto, destaca que estos alelos provenientes de *S. paradoxus* prevalecen en la mayoría de los aislados mexicanos mientras que en las cepas del resto del mundo no.

La amplia gama de estructuras genómicas observadas en los híbridos secuenciados durante este estudio poblacional sugiere múltiples hibridaciones antiguas que dieron origen a las introgresiones presentes en la mayoría de los aislados secuenciados. En conjunto, los resultados presentados muestran la clara presencia de material genético originario de *S. paradoxus* en los aislados de *S. cerevisiae* pertenecientes a fermentaciones de agave, así como distintos híbridos, abriendo la puerta a la posibilidad de la hibridación como mecanismo de adaptación al nicho de agave.

9. Perspectivas

Aunque el trabajo realizado para esta tesis brinda una perspectiva más amplia acerca de la diversidad genómica de las levaduras involucradas en el proceso de fermentación de agave para la producción de mezcal, aún hay mucho trabajo por hacer para comprender en mayor medida la historia evolutiva y diversidad de estas levaduras.

Un paso crucial para entender la estructura genómica de los híbridos sería poder secuenciar los híbridos utilizando tecnologías de secuenciación de lecturas largas. Esto nos brindaría información sobre la forma real en que se han combinado ambos genomas, sin tomar como base las posiciones de los genomas de referencia.

También queda pendiente determinar las ploidías de los aislados secuenciados mediante FACS para observar cómo es que las ploidías se distribuyen geográficamente y si es que esta aporta una ventaja adaptativa.

Además, es importante fenotipar los aislados mexicanos, lo cual nos permitiría determinar si las introgresiones de *S. paradoxus* les proporcionan una ventaja al momento de crecer en el jugo de fermentación de agave en comparación con los aislados del resto del mundo.

Entendiendo que el proceso de elaboración de mezcal es en su mayoría artesanal y que en muchos casos la fermentación se deja en manos de la naturaleza, sería importante muestrear y analizar aislados naturales provenientes de la vecindad de las tinajas de fermentación para observar si existen diferencias al compararlas con sus semejantes aisladas del jugo de fermentación de agave. Además de secuenciar y analizar un mayor número de aislados previamente recolectados para obtener un mayor número de híbridos y cepas de *S. paradoxus* con la finalidad de tener una filogenia más completa.

Elaborar un método de introgresiones sensible a relaciones de paralogía ayudará a establecer si las categorías enriquecidas encontradas actualmente son en realidad debido a que hay parálogos entre los genes compartidos entre cepas, o son debido a que en realidad se encuentran sobrerrepresentadas. El hecho de que estas introgresiones se encuentren en regiones subteloméricas es un factor para tomar en cuenta para análisis futuros.

10. Referencias bibliográficas

- 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>
- Casas, A., Torres-Guevara, J., & Parra, F. (2017). *Domesticación en el Continente Americano Volumen 2. Investigación para el manejo sustentable de recursos genéticos en el nuevo mundo*.
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Dalvie, N. C., Rodriguez-Aponte, S. A., Hartwell, B. L., Tostanoski, L. H., Biedermann, A. M., Crowell, L. E., Kaur, K., Kumru, O. S., Carter, L., Yu, J., Chang, A., McMahan, K., Courant, T., Lebas, C., Lemnios, A. A., Rodrigues, K. A., Silva, M., Johnston, R. S., Naranjo, C. A., ... Love, J. C. (2021). Engineered SARS-CoV-2 receptor binding domain improves manufacturability in yeast and immunogenicity in mice. *Proceedings of the National Academy of Sciences*, 118(38), e2106845118. <https://doi.org/10.1073/pnas.2106845118>
- 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), 420–425. <https://doi.org/10.1038/s41586-020-2889-1>
- de Queiroz, K. (2005). Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences*, 102(suppl_1), 6600–6607. <https://doi.org/10.1073/pnas.0502030102>
- Drillon, G., Carbone, A., & Fischer, G. (2014). SynChro: A Fast and Easy Tool to Reconstruct and Visualize Synteny Blocks along Eukaryotic Chromosomes. *PLoS ONE*, 9(3), e92621. <https://doi.org/10.1371/journal.pone.0092621>
- Financiera Rural. (2011). *Monografía del Mezcal*. Financiera Rural.
- Gallone, B., Steensels, J., Mertens, S., Dzialo, M. C., Gordon, J. L., Wauters, R., Theßeling, F. A., Bellinazzo, F., Saels, V., Herrera-Malaver, B., Prahl, T., White, C., Hutzler, M., Meußdoerffer, F., Malcorps, P., Souffriau, B., Daenen, L., Baele, G., Maere, S., & Verstrepen, K. J. (2019). Interspecific hybridization facilitates niche adaptation in beer yeast. *Nature Ecology & Evolution*, 3(11), 1562–1575. <https://doi.org/10.1038/s41559-019-0997-9>
- Garibay Marcelo, M. D. L. A. (2019). *Caracterización de la diversidad de levaduras presentes durante el proceso de fermentación del mezcal en el estado de Guerrero*. Universidad Autónoma de Guerrero.
- Giannakou, K., Cotterrell, M., & Delneri, D. (2020). Genomic Adaptation of *Saccharomyces* Species to Industrial Environments. *Frontiers in Genetics*, 11, 916. <https://doi.org/10.3389/fgene.2020.00916>
- 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*, 274(5287), 546–567. <https://doi.org/10.1126/science.274.5287.546>
- Heasley, L. R., & Argueso, J. L. (2022). Genomic characterization of a wild diploid isolate of *Saccharomyces cerevisiae* reveals an extensive and dynamic landscape of structural variation. *Genetics*, 220(3), iyab193. <https://doi.org/10.1093/genetics/iyab193>
- Kirchmayr, M. R., Segura-García, L. E., Lappe-Oliveras, P., Moreno-Terrazas, R., de la Rosa, M., & Gschaedler Mathis, A. (2017). Impact of environmental conditions and process modifications on microbial diversity, fermentation efficiency and chemical profile during the fermentation of Mezcal in Oaxaca. *LWT - Food Science and Technology*, 79, 160–169. <https://doi.org/10.1016/j.lwt.2016.12.052>
- Korostin, D., Kulemin, N., Naumov, V., Belova, V., Kwon, D., & Gorbachev, A. (2020). Comparative analysis of novel MGISEQ-2000 sequencing platform vs Illumina HiSeq 2500 for whole-genome sequencing. *PLOS ONE*, 15(3), e0230301. <https://doi.org/10.1371/journal.pone.0230301>
- Krogerus, K., Preiss, R., & Gibson, B. (2018). A Unique *Saccharomyces cerevisiae* × *Saccharomyces uvarum* Hybrid Isolated from Norwegian Farmhouse Beer: Characterization and Reconstruction. *Frontiers in Microbiology*, 9, 2253. <https://doi.org/10.3389/fmicb.2018.02253>
- Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biology*, 5(2), R12. <https://doi.org/10.1186/gb-2004-5-2-r12>
- Kurtzman, C. P., Fell, J. W., & Boekhout, Teun. (2011). *The Yeasts*. Elsevier Science. <http://www.totalboox.com/book/id-8426120699017155524>
- Lappe-Oliveras, P., Moreno-Terrazas, R., Arrizón-Gaviño, J., Herrera-Suárez, T., García-Mendoza, A., & Gschaedler-Mathis, A. (2008). Yeasts associated with the production of Mexican alcoholic nondistilled and distilled *Agave* beverages. *FEMS Yeast Research*, 8(7), 1037–1052. <https://doi.org/10.1111/j.1567-1364.2008.00430.x>
- Li, H. (2014). *Aligning sequence reads, clone sequences and assembly con*gs with BWA-MEM* (p. 0 Bytes). figshare. <https://doi.org/10.6084/M9.FIGSHARE.963153.V1>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Liti, G., Warringer, J., Blomberg, A. (2017). Isolation and Laboratory Domestication of Natural Yeast Strains. *Cold Spring Harb. Protoc.*, 2017(8):pdb.prot089052. <https://doi:10.1101/pdb.prot089052>. PMID: 28765292.
- Masel, J. (2011). Genetic drift. *Curr. Biol.*, 21(20):R837-838. <https://doi:10.1016/j.cub.2011.08.007>.
- 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/MMBR.00022-12>
- Peris, D., Ubbelohde, E. J., Kuang, M. C., Kominek, J., Langdon, Q. K., Adams, M., Koshalek, J. A., Hulfachor, A. B., Opulente, D. A., Hall, D. J., Hyma, K., Fay, J. C., Leducq, J.-B., Charron, G., Landry, C. R., Libkind, D., Gonçalves, C., Gonçalves, P., Sampaio, J. P., ... Hittinger, C. T.

- (2023). Macroevolutionary diversity of traits and genomes in the model yeast genus *Saccharomyces*. *Nat Commun.*, 14(1):690. <https://doi.org/10.1038/s41467-023-36139-2>.
- Perli, T., Wronska, A. K., Ortiz-Merino, R. A., Pronk, J. T., & Daran, J. (2020). Vitamin requirements and biosynthesis in *Saccharomyces cerevisiae*. *Yeast*, 37(4), 283–304. <https://doi.org/10.1002/yea.3461>
- 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), 339–344. <https://doi.org/10.1038/s41586-018-0030-5>
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. <https://www.R-project.org/>
- Rossum, G. van, & Drake, F. L. (2010). *The Python language reference* (Release 3.0.1 [Repr.]). Python Software Foundation.
- Shen, X.-X., Opulente, D. A., Kominek, J., Zhou, X., Steenwyk, J. L., Buh, K. V., Haase, M. A. B., Wisecaver, J. H., Wang, M., Doering, D. T., Boudouris, J. T., Schneider, R. M., Langdon, Q. K., Ohkuma, M., Endoh, R., Takashima, M., Manabe, R., Čadež, N., Libkind, D., ... Rokas, A. (2018). Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell*, 175(6), 1533-1545.e20. <https://doi.org/10.1016/j.cell.2018.10.023>
- Sicard, D., & Legras, J.-L. (2011). Bread, beer and wine: Yeast domestication in the *Saccharomyces sensu stricto* complex. *Comptes Rendus Biologies*, 334(3), 229–236. <https://doi.org/10.1016/j.crv.2010.12.016>
- Steensels, J., Gallone, B., Voordeckers, K., & Verstrepen, K. J. (2019). Domestication of Industrial Microbes. *Current Biology*, 29(10), R381–R393. <https://doi.org/10.1016/j.cub.2019.04.025>
- Sylvester, K., Wang, Q.-M., James, B., Mendez, R., Hulfachor, A. B., & Hittinger, C. T. (2015). Temperature and host preferences drive the diversification of *Saccharomyces* and other yeasts: A survey and the discovery of eight new yeast species. *FEMS Yeast Research*, 15(3). <https://doi.org/10.1093/femsyr/fov002>
- Winzeler, E. A., Castillo-Davis, C. I., Oshiro, G., Liang, D., Richards, D. R., Zhou, Y., & Hartl, D. L. (2003). Genetic Diversity in Yeast Assessed with Whole-Genome Oligonucleotide Arrays. *Genetics*, 163(1), 79–89. <https://doi.org/10.1093/genetics/163.1.79>
- Wolfe, K. H. (2015). Origin of the Yeast Whole-Genome Duplication. *PLOS Biology*, 13(8), e1002221. <https://doi.org/10.1371/journal.pbio.1002221>
- Zhang, H., Skelton, A., Gardner, R. C., & Goddard, M. R. (2010). *Saccharomyces paradoxus* and *Saccharomyces cerevisiae* reside on oak trees in New Zealand: Evidence for migration from Europe and interspecies hybrids: New Zealand *Saccharomyces* ecology. *FEMS Yeast Research*, 10(7), 941–947. <https://doi.org/10.1111/j.1567-1364.2010.00681.x>

11. Bibliografía

- Kurtzman, Cletus, J. W. Fell, and Teun Boekhout. 2011. *The Yeasts: A Taxonomic Study*. Elsevier.
- Ayyildiz, Dilara, and Silvano Piazza. 2019. "Introduction to Bioinformatics." *Methods in Molecular Biology* 1986: 1–15.
- Baxevanis, Andreas D., Gary D. Bader, and David S. Wishart. 2020. *Bioinformatics*. John Wiley & Sons.
- Chan, Bertram K. C. 2018. "Data Analysis Using R Programming." *Advances in Experimental Medicine and Biology* 1082: 47–122.
- Gentleman, Robert. 2008. *R Programming for Bioinformatics*. CRC Press.
- Peng, Roger D. 2016. *R Programming for Data Science*. Leanpub Victoria, BC, Canada.

12. Material Suplementario

Tabla Suplementaria 1

Tabla suplementaria 1. Lista de genomas usados

ID	Seq_Date	Lib_Prepare	Platform	ID_SeqProvider	Ploidy	Geo_Region	Isolation	Group	Clasif
DK001c10	4/20/20	Mod_MPFrance	BGI	DD2a	NA	Oaxaca	Santiago Matatlan	BGI_1.5	oursSACE_TESIS
DK001c27	4/20/20	Mod_MPFrance	BGI_1	02-D4	NA	Oaxaca	Santiago Matatlan	BGI_1	oursSACE_TESIS
DK002c39	4/20/20	Mod_MPFrance	BGI_1	02-E2	NA	Oaxaca	San Pedro Totolepan	BGI_1	oursHYB
DK003c30	4/20/20	Mod_MPFrance	BGI_1	05-E8	NA	San Luis Potosi	Laguna Seca	BGI_1	oursHYB
DK003c56	4/20/20	Mod_MPFrance	BGI_1	02-E6	NA	San Luis Potosi	Laguna Seca	BGI_1	oursSACE_TESIS
DK016c5	4/20/20	Mod_MPFrance	BGI_1	02-E7	NA	Guerrero	Eduardo Neri	BGI_1	oursSACE_TESIS
DK017c3	4/20/20	Mod_MPFrance	BGI_1	02-E10	NA	Guerrero	Chilapa	BGI_1	oursKLMA
DK017c4	4/20/20	Mod_MPFrance	BGI_1	05-D8	NA	Guerrero	Chilapa	BGI_1	oursSACE_TESIS
DK018c4	4/20/20	Mod_MPFrance	BGI_1	02-E12	NA	Guerrero	Huitzoco	BGI_1	oursSACE_TESIS
DK019c1	4/20/20	Mod_MPFrance	BGI_1	05-D6	NA	Michoacan	Charo	BGI_1	oursSACE_TESIS
DK030c8	4/20/20	Mod_MPFrance	BGI_1	02-B11	NA	Jalisco	Altos sur	BGI_1	oursUNKNOWN
DS002c10	4/20/20	Mod_MPFrance	BGI_1	02-G5	NA	Sonora	Huasabas	BGI_1	oursSACE_TESIS
DS002c4	4/20/20	Mod_MPFrance	BGI_1	G1 petri	NA	Sonora	Huasabas	BGI_1	oursSACE_TESIS
DS002c5	4/20/20	Mod_MPFrance	BGI_1	G4 petri	NA	Sonora	Huasabas	BGI_1	oursSACE_TESIS
DS002c6	4/20/20	Mod_MPFrance	BGI_1	G6 petri	NA	Sonora	Huasabas	BGI_1	oursSACE_TESIS
JS002c1	NA	NA	HiSeq2000	DBVPG4410	2	Italy	Wine	1011	others
JS008c1	NA	NA	HiSeq2000	SBD	2	Northern Europe	Bakery	1011	others
JS014c1	NA	NA	HiSeq2000	CEV622	2	French Guiana	Human feces	1011	others
JS022c1	NA	NA	HiSeq2000	DBVPG1631	2	Spain	Grape must treated with sulfur dioxide	1011	others
JS025c1	NA	NA	HiSeq2000	CBS4903	2	Spain	Alpechin (waste from olive oil production)	1011	1011alp
JS032c1	NA	NA	HiSeq2000	CBS7539	2	Plovdiv, Bulgaria	Turbid beer	1011	others
JS033c1	NA	NA	HiSeq2000	CBS1602	2	NA	Must of Blanchard pears	1011	others
JS038c1	NA	NA	HiSeq2000	SJ5111	3	Tuchih, Nantou, Taiwan	Leaf of Acacia dealbata	1011	others
JS065c1	NA	NA	HiSeq2000	CBS6308	2	South Yorkshire, UK	Brewer yeast, Yorkshire type	1011	others
JS070c1	NA	NA	HiSeq2000	CBS7993	1	Rio de Janeiro, Tijuca National Forest, Cushvara River, Brazil	Fresh water	1011	others
JS080c1	NA	NA	HiSeq2000	CLIB382_1b	1	Ireland	Beer	1011	others
JS086c1	NA	NA	HiSeq2000	HN10	2	Wuzhi Mountain, Hainan province, China	Rotten wood	1011	others
JS089c1	NA	NA	HiSeq2000	2163	2	Castellon, Spain	Quercus faginea, 30C	1011	others
JS094c1	NA	NA	HiSeq2000	L-958	2	Mendoza area, Argentina	Wine	1011	others
JS098c1	NA	NA	HiSeq2000	UWOPS03-433.3	2	Malaysia	Nectar, bertam palm	1011	others
JS104c1	NA	NA	HiSeq2000	CECT10266	1	Spain	Tanning liquor	1011	1011alp
JS109c1	NA	NA	HiSeq2000	906	2	Mexico	Must from various Agave sp.	1011	1011mex
JS114c1	NA	NA	HiSeq2000	CEV651	2	French Guiana	Fruits	1011	others
JS119c1	NA	NA	HiSeq2000	MAJ_G	2	Majunga, Madagascar	Plant material	1011	others
JS127c1	NA	NA	HiSeq2000	ES2M03-7A	2	Taiwan	Fruiting body of Russula sp.	1011	others
JS150c1	NA	NA	HiSeq2000	CEV625	2	French Guiana	Human feces	1011	others
JS157c1	NA	NA	HiSeq2000	CBS3093	1	Spain	Alpechin (waste from olive oil production)	1011	1011alp
JS160c1	NA	NA	HiSeq2000	CLIB192_1b	1	France	Bakery	1011	others
JS161c1	NA	NA	HiSeq2000	CBS2247	1	South Africa	Grape must	1011	others
JS163c1	NA	NA	HiSeq2000	SA.1.5.	2	Brazil	Bioethanol	1011	others
JS165c1	NA	NA	HiSeq2000	HE012	2	French Guiana	Human feces	1011	others
JS166c1	NA	NA	HiSeq2000	CLIB1082	2	Upper Normandy, France	Cider brewery, dry cider	1011	others
JS171c1	NA	NA	HiSeq2000	CEV620	2	French Guiana	Human feces	1011	others

Tabla suplementaria 1. Lista de genomas usados

J5191c1	NA	NA	HiSeq2000	CBS2910	1 Portugal	Human feces	1011	1011alp
J5194c1	NA	NA	HiSeq2000	ATCC_38618:2-2	2 NA	Concentrated orange juice	1011	others
J5204c1	NA	NA	HiSeq2000	CBS1586	1 NA	Pear must	1011	others
J5208c1	NA	NA	HiSeq2000	CEV621	2 French Guiana	Human feces	1011	others
J5210c1	NA	NA	HiSeq2000	CEV617	2 French Guiana	Human feces	1011	others
J5211c1	NA	NA	HiSeq2000	CECT10171-5B	2 Spain	Larvae of <i>Dacus oleae</i>	1011	1011alp
J5229c1	NA	NA	HiSeq2000	L-960	2 Mendoza area, Argentina	Wine	1011	others
J5235c1	NA	NA	HiSeq2000	ZP_611	2 Canada, Varcouwer,	<i>Quercus robur</i>	1011	others
J5240c1	NA	NA	HiSeq2000	LCBG-3Y8	2 Tamaulipas, Mexico	Agave spp. fermentation during rustic mezcal production	1011	1011mex
J5246c1	NA	NA	HiSeq2000	HE015	1 French Guiana	Human feces	1011	others
J5247c1	NA	NA	HiSeq2000	TL22952.2	1 France	Cheese	1011	others
J5252c1	NA	NA	HiSeq2000	Ksc2-2B	2 Japan	Tree exudate	1011	others
J5254c1	NA	NA	HiSeq2000	MC6	2 Umbria, Italy	Grape	1011	others
J5262c1	NA	NA	HiSeq2000	RIB6006	2 Japan	Sake	1011	others
J5263c1	NA	NA	HiSeq2000	HE005	2 French Guiana	Human feces	1011	others
J5270c1	NA	NA	HiSeq2000	HE016	2 French Guiana	Human feces	1011	others
J5273c1	NA	NA	HiSeq2000	LCBG-3Y3	2 Tamaulipas, Mexico	Agave spp. fermentation during rustic mezcal production	1011	1011mex
J5274c1	NA	NA	HiSeq2000	DBY/PG6302	2 Spain	Alpechin (waste from olive oil production)	1011	1011alp
J5277c1	NA	NA	HiSeq2000	CEV652	2 French Guiana	Fruits	1011	others
J5281c1	NA	NA	HiSeq2000	SM.8.7.BR1	2 Brazil	Bioethanol	1011	others
J5288c1	NA	NA	HiSeq2000	CBS4507	2 NA	Brewer's yeast, English top yeast	1011	others
J5293c1	NA	NA	HiSeq2000	CBS403	2 West Africa	Ginger beer, from Zingiber officinale	1011	others
J5294c1	NA	NA	HiSeq2000	CLIB1413	2 Laos	Rice wine, Oenology	1011	others
J5308c1	NA	NA	HiSeq2000	DBY/PG1919	3 Italy	Caverns	1011	others
J5310c1	NA	NA	HiSeq2000	HE009	2 French Guiana	Human feces	1011	others
J5311c1	NA	NA	HiSeq2000	HE010	2 French Guiana	Human feces	1011	others
J5312c1	NA	NA	HiSeq2000	YJM269_1b	1 Austria	Red Blauer Portugieser grapes	1011	others
J5319c1	NA	NA	HiSeq2000	CBS5155	2 South Armenia	Fermenting Grape must	1011	others
J5321c1	NA	NA	HiSeq2000	UCD_11-601	2 Davis, CA, UCSD Food Science, Pilot plant.	Olive fermentation, 1-gallon, Bucket 5B brine, Sicilian style olives, inoculated with <i>Saccharomyces cerevisiae</i> from 2009 olives.	1011	1011alp
J5326c1	NA	NA	HiSeq2000	CEV616	2 French Guiana	Human feces	1011	others
J5329c1	NA	NA	HiSeq2000	NPA03.1	2 Nigeria	Palm wine	1011	others
J5335c1	NA	NA	HiSeq2000	CEV615	2 French Guiana	Human feces	1011	others
J5345c1	NA	NA	HiSeq2000	CBS1190	2 Germany	Wine	1011	others
J5346c1	NA	NA	HiSeq2000	CBS3000	1 Denmark	Palm wine	1011	others
J5348c1	NA	NA	HiSeq2000	B-17	2 Georgia	Wine conserved in amphora	1011	others
J5356c1	NA	NA	HiSeq2000	CBS2188	2 Japan	Fruit of <i>Eryobotrya japonica</i>	1011	others
J5368c1	NA	NA	HiSeq2000	CLIB1077	2 Upper Normandy, France	Cider brewery, dry cider	1011	others
J5372c1	NA	NA	HiSeq2000	N38-4A	2 Botanical Garden, Novosibirsk, Siberia, Russia	Exudate of <i>Quercus robur</i>	1011	others
J5375c1	NA	NA	HiSeq2000	BI20	2 Dongling Mountain, Beijing, China	Bark from <i>Quercus wutaishanica</i>	1011	others
J5378c1	NA	NA	HiSeq2000	CBS6914	1 Spain	White wine	1011	others
J5382c1	NA	NA	HiSeq2000	A-18	4 Ghana	Sorghum beer	1011	others
J5385c1	NA	NA	HiSeq2000	FTPW35	2 Burkina Faso	Palm wine	1011	others
J5387c1	NA	NA	HiSeq2000	CEV649	2 French Guiana	Agouti paca (RO)	1011	others
J5389c1	NA	NA	HiSeq2000	CBS1196	2 Indonesia	Arak Distillery	1011	others

Tabla suplementaria 1. Lista de genomas usados

J5402c1	NA	NA	CBS1428			1 NA	Bakery	1011 others
J5403c1	NA	NA	SJ5117		2 Austria	2	Probiotic	1011 others
J5410c1	NA	NA	CBS5112		2 Spain	2	Grape must	1011 others
J5417c1	NA	NA	N3.00-7A		2 Botanical garden, Blagoveshchensk, Russia	2	Grape berries of wild <i>Vitus amurensis</i>	1011 others
J5418c1	NA	NA	HN6		2 Bawangling Mountain, Hainan province, China	2	Rotten wood	1011 others
J5419c1	NA	NA	CLIB563		2 Normandy, France	2	Dairy, cheese (camembert)	1011 others
J5421c1	NA	NA	LCBG-3D6		2 Tamaulipas, Mexico	2	Agave spp. fermentation during rustic mezcal production	1011 1011mex
J5423c1	NA	NA	CEY647	997	2 Hungary	2	Apricot pulp	1011 others
J5440c1	NA	NA	DBVPG1581		2 French Guiana	2	Molossus molossus (bat)	1011 others
J5445c1	NA	NA	N51-1A		2 Italy	2	Grape must	1011 others
J5454c1	NA	NA	CBS2165a		2 East	2	Exudate of <i>Quercus mongolica</i>	1011 others
J5460c1	NA	NA	LCBG-3D4		2 England	2	Brewers	1011 others
J5466c1	NA	NA	HE004		2 Tamaulipas, Mexico	2	Agave spp. fermentation during rustic mezcal production	1011 1011mex
J5471c1	NA	NA	MTF2554		2 French Guiana	2	Human feces	1011 others
J5474c1	NA	NA		2187	2 West Africa	2	Cocoa bean fermentation	1011 others
J5497c1	NA	NA	CBS5378	908	2 Hungary	2	Forest soil, 30C	1011 others
J5499c1	NA	NA	CBS1239		2 Jalisco, Mexico	2	Raicilla (Mexican distilled spirit)	1011 1011mex
J5502c1	NA	NA	CBS1388a		2 Spain	2	Alpechin (waste from olive oil production)	1011 1011alp
J5510c1	NA	NA	RP.10.13		4 NA	4	Distillery	1011 others
J5525c1	NA	NA	HE007		1 NA	1	Unknown	1011 others
J5527c1	NA	NA	SX3		2 Brazil	2	Bioethanol	1011 others
J5542c1	NA	NA	CBS1252		2 French Guiana	2	Human feces	1011 others
J5553c1	NA	NA	VNL3		2 Quiling Mountain, Shaanxi province, China	2	Bark from <i>Carya</i> sp.	1011 others
J5561c1	NA	NA	RIB1003		1 France	1	Bakery	1011 others
J5562c1	NA	NA	MC5		3 Ha Noi, North Vietnam	3	Unknown	1011 others
J5574c1	NA	NA	Y10_1b		2 Japan	2	Sake	1011 others
J5576c1	NA	NA	RP11.4.14		2 Ascoli Piceno, Italy	2	Cooked wine (vino cotto)	1011 others
J5579c1	NA	NA	K1_1		1 Philippines	1	Coconut	1011 others
J5581c1	NA	NA	CECT10331:9-1		2 Brazil	2	Bioethanol	1011 others
J5585c1	NA	NA	EXF-7145		1 Japan	1	Sake	1011 others
J5594c1	NA	NA	HE006		2 Spain	2	Flowers of <i>Centaurea alba</i>	1011 others
J5599c1	NA	NA	EMI14S01-3B		2 Serbia	2	<i>Quercus robur</i>	1011 others
J5611c1	NA	NA	UCD_05-780		2 French Guiana	2	Human feces	1011 others
J5612c1	NA	NA	HE008		2 Taiwan	2	Soil	1011 others
J5615c1	NA	NA	FTPW3		Winters, CA Wolfskill		Beetle from infested <i>Prunus bokhariensis</i> , walk plate on RBGA. No mold on plate. <i>Prunus bokhariensis</i> Type Botanky E-1-18	1011 others
J5646c1	NA	NA	CLOCA_20-221		2 USDA Prunus germplasm collection,	2	Human feces	1011 others
J5649c1	NA	NA	21-4-122		2 French Guiana	2	Palm wine	1011 others
J5654c1	NA	NA	CEY624		2 Burkina Faso	2	Flower from <i>Heliconia</i> sp.	1011 others
J5657c1	NA	NA	CH06		2 Yasuni, Orellana	2	Blossoms of apricot tree (<i>Prunus armeniaca</i>)	1011 others
J5660c1	NA	NA	SG60		2 Male Zaluzie, Slovakia	2	Human feces	1011 others
	NA	NA			5 Abengourou, Ivory Coast	5	Pearl millet beer	1011 others
	NA	NA			2 Italy	2	Wine	1011 others

Tabla suplementaria 1. Lista de genomas usados

J5670c1	NA	NA	HiSeq2000	N19-1C		Gelendzhik region, Caucasus, Russia	Exudate of Quercus sp.	1011	others
J5671c1	NA	NA	HiSeq2000	DBVPG1895		2 Italy	Grape must treated with pesticides	1011	others
J5677c1	NA	NA	HiSeq2000	FTPW32		2 Burkina Faso	Palm wine	1011	others
J5678c1	NA	NA	HiSeq2000		2162	2 Hungary	Forest soil_30C	1011	others
J5682c1	NA	NA	HiSeq2000	21-4-105		2 Bratislava, Deniska Nova Ves, Slovakia	River water/Morava	1011	others
J5688c1	NA	NA	HiSeq2000	DBVPG1085		3 Corato, Italy	Grape must	1011	others
J5691c1	NA	NA	HiSeq2000	HE001		2 French Guiana	Human feces	1011	others
J5719c1	NA	NA	HiSeq2000	CB51236-a		2 France	Bakery	1011	others
J5722c1	NA	NA	HiSeq2000	CB51171		4 Oranjeboom brewery, Rotterdam, Netherlands	Brewers top yeast	1011	others
J5728c1	NA	NA	HiSeq2000	HE013		2 French Guiana	Human feces	1011	others
J5733c1	NA	NA	HiSeq2000	FIMA_3		2 Italy	Wine	1011	others
J5735c1	NA	NA	HiSeq2000	CECT1462		4 UK	Beer	1011	others
J5737c1	NA	NA	HiSeq2000	UCD_09-448		2 Davis, CA, UCD	Olives from Middle of season 1. on RBCA 11.25.09 Streaked onto PDA 11/30/09	1011	1011alp
J5740c1	NA	NA	HiSeq2000	CB58292		2 Sweden	Seawater	1011	others
J5741c1	NA	NA	HiSeq2000	DBVPG1790		2 Turku, Finland	Soil	1011	others
J5748c1	NA	NA	HiSeq2000	TayTa		2 Long An, South Vietnam	Unknown	1011	others
J5754c1	NA	NA	HiSeq2000	HE020		2 French Guiana	Human	1011	others
J5756c1	NA	NA	HiSeq2000	UCD_06-645		4 Davis, CA, UCD	Female olive fly, surface sterilized, smashed, plated on RBCA	1011	others
J5757c1	NA	NA	HiSeq2000	ES2M03		2 Dongpu, Sinyi Township, Nantou, Taiwan	Fruiting body of Russula sp.	1011	others
J5759c1	NA	NA	HiSeq2000	HE011		2 French Guiana	Human feces	1011	others
J5760c1	NA	NA	HiSeq2000	DBVPG1918		2 Malia	Grape must	1011	others
J5771c1	NA	NA	HiSeq2000	HN15		2 Wuzhi Mountain, Hainan province, China	Rotten wood	1011	others
J5772c1	NA	NA	HiSeq2000	E4		2 Umbria, Italy		1011	others
J5773c1	NA	NA	HiSeq2000	UCD_40-346		2 Fresno area, California	Spoiled Adriatic orchard figs, (1939)	1011	others
J5774c1	NA	NA	HiSeq2000	ZP_567		2 Portugal, Castelo de Vide,	Quercus pyrenaica	1011	others
J5781c1	NA	NA	HiSeq2000	YPS133		2 Pennsylvania, USA	Soil beneath Quercus alba	1011	others
J5783c1	NA	NA	HiSeq2000	CB51540		1 Switzerland	Tibi grains	1011	others
J5788c1	NA	NA	HiSeq2000		1560	2 Acetuna, Spain	Manzanilla-Alorena, olive (Noe), 2003	1011	1011alp
J5804c1	NA	NA	HiSeq2000	DJ66		3 Yoboki, Djibouti	Palm wine	1011	others
J5806c1	NA	NA	HiSeq2000	LCBG-Mosca3		2 Tamaulipas, Mexico	Agave spp. fermentation during rustic mezcäl production	1011	1011mex
J5807c1	NA	NA	HiSeq2000	SJ5113		4 Tuchi, Nantou, Taiwan	Leaf of Wendlandia formosana	1011	others
J5817c1	NA	NA	HiSeq2000	YJM990		2 Italy	Vagina	1011	others
J5818c1	NA	NA	HiSeq2000	CEV618		2 French Guiana	Human feces	1011	others
J5825c1	NA	NA	HiSeq2000	CB54458		1 France	Red wine	1011	others
J5827c1	NA	NA	HiSeq2000	D-4		2 Georgia	Wine	1011	others
J5835c1	NA	NA	HiSeq2000	CEV619		2 French Guiana	Human feces	1011	others
J5846c1	NA	NA	HiSeq2000	RP11.4.11		2 Brazil	Bioethanol	1011	others
J5849c1	NA	NA	HiSeq2000	Y55_1b		1 France	Wine	1011	others
J5861c1	NA	NA	HiSeq2000	CEV653		2 French Guiana	Fruits	1011	others
J5862c1	NA	NA	HiSeq2000	CCY_21-4-106		4 Devinska Nova Ves, Slovakia	River water (Morava)	1011	others
J5863c1	NA	NA	HiSeq2000	LCBG-3D2		2 Tamaulipas, Mexico	Agave spp. fermentation during rustic mezcäl production	1011	1011mex
J5880c1	NA	NA	HiSeq2000	LCBG-4Y3		2 Tamaulipas, Mexico	Agave spp. fermentation during rustic mezcäl production	1011	1011mex
J5881c1	NA	NA	HiSeq2000	Y3_1b		1 NA	Palm wine	1011	others

Tabla suplementaria 1. Lista de genomas usados

J585c1	NA	NA	HiSeq2000	RP11.4-1	2	Brazil	Bioethanol	1011	others
J5889c1	NA	NA	HiSeq2000	CEV623	2	French Guiana	Human feces	1011	others
J5898c1	NA	NA	HiSeq2000	DBVPG1659	2	Campo de Santiago, La Mancha, Spain	Grape must	1011	others
J5903c1	NA	NA	HiSeq2000	BS6333	1	Costa Rica	Rotting banana (from <i>Musa sapientum</i>)	1011	others
J5906c1	NA	NA	HiSeq2000	CECT10497	2	Spain	Alpechin (waste from olive oil production)	1011	1011alp
J5910c1	NA	NA	HiSeq2000	EXF-7197	2	Montenegro	Quercus sp.	1011	others
J5915c1	NA	NA	HiSeq2000	CBS3093_1b	1	Spain	Alpechin (waste from olive oil production)	1011	1011alp
J5922c1	NA	NA	HiSeq2000	CEV650	2	French Guiana	Agouti paca (RO)	1011	others
J5925c1	NA	NA	HiSeq2000	CBS3081	1	Spain	Alpechin	1011	1011alp
J5928c1	NA	NA	HiSeq2000	L-1198	2	Valle de Ica, Peru	Heterozygote, 12% Sporulation Efficiency	1011	others
J5933c1	NA	NA	HiSeq2000	CEV646	2	French Guiana	Chicken	1011	others
J5937c1	NA	NA	HiSeq2000	CBS2909	2	NA	Human feces	1011	1011alp
J5939c1	NA	NA	HiSeq2000	DBVPG6590	1	Brazil	Alcohol production tank	1011	others
KV001c1	NA	NA	HiSeq2500	BE124	1	France	Sc_Se	Verstrepen	others
KV002c1	NA	NA	MiSeqv3	TUM55B	1	France	Su_Se	Verstrepen	others
KV003c1	NA	NA	MiSeqv3	BE107	1	France	Sc_Sk	Verstrepen	others
KV004c1	NA	NA	MiSeqv3	ABL1608	1	France	Su_Se	Verstrepen	others
XA017c1	4/20/20	Mod_MPFrance	BGI_1	01-G9	NA	Estado de Mexico	Palmar de Guadalupe	BGI_1	oursSACE_TESIS
XA020c1	4/20/20	Mod_MPFrance	BGI	X020.1	NA	Estado de Mexico	Palmar de Guadalupe	BGI_1.5	oursHYB
XA022c2	4/20/20	Mod_MPFrance	BGI_1	01-H4	NA	Estado de Mexico	Palmar de Guadalupe	BGI_1	oursHYB
XA023c1	4/20/20	Mod_MPFrance	BGI_1	01-G7	NA	Estado de Mexico	Palmar de Guadalupe	BGI_1	oursSACE_TESIS
XA024c2	4/20/20	Mod_MPFrance	BGI	X024.2	NA	Estado de Mexico	Palma de Guadalupe	BGI_1.5	oursSACE_TESIS
XA030c1	4/20/20	Mod_MPFrance	BGI_1	01-G5	NA	Estado de Mexico	San Jose Chalmita	BGI_1	oursSACE_TESIS
XA031c2	4/20/20	Mod_MPFrance	BGI_1	01-G10	NA	Estado de Mexico	San Jose Chalmita	BGI_1	oursSACE_TESIS
XA033c2	4/20/20	Mod_MPFrance	BGI_1	01-G6	NA	Estado de Mexico	Piedra Parrada	BGI_1	oursPIKU
XA055c2	4/20/20	Mod_MPFrance	BGI_1	01-B7	NA	Oaxaca	Mihuatv'n	BGI_1	oursSACE_TESIS
XA057c4	4/20/20	Mod_MPFrance	BGI_1	01-C6	NA	Oaxaca	San Isidro Guishe	BGI_1	oursSACE_TESIS
XA063c8	4/20/20	Mod_MPFrance	BGI_1	18 A8	NA	Oaxaca	Santa Catarina de Minas	BGI_1	oursKLMA
XA070c1	4/20/20	Mod_MPFrance	BGI_1	01-H9	NA	Oaxaca	Jayacatv'n	BGI_1	oursSACE_TESIS
XA072c4	4/20/20	Mod_MPFrance	BGI_1	01-C7	NA	Oaxaca	Jayacatv'n	BGI_1	oursSACE_TESIS
XA074c2	4/20/20	Mod_MPFrance	BGI_1	01-C12	NA	Oaxaca	Santa Catarina Ixcatv'n	BGI_1	oursSACE_TESIS
XA076c20	4/20/20	Mod_MPFrance	BGI_1	22 B8	NA	Guerrero	Apeltanca	BGI_1	oursKLMA
XA081c6	4/20/20	Mod_MPFrance	BGI_1	01-C1	NA	Guajuato	San Felipe	BGI_1	oursSACE_TESIS
XA082c2	4/20/20	Mod_MPFrance	BGI_1	01-D1	NA	Guajuato	San Felipe	BGI_1	oursHYB
XA082c24	4/20/20	Mod_MPFrance	BGI_1	21 B12	NA	Guajuato	San Felipe	BGI_1	oursKLMA
XA083c1	4/20/20	Mod_MPFrance	BGI_1	01-D11	NA	Guajuato	San Felipe	BGI_1	oursSACE_TESIS
XA083c6	4/20/20	Mod_MPFrance	BGI_1	05-B8	NA	Guajuato	San Felipe	BGI_1	oursSACE_TESIS
XA096c3	4/20/20	Mod_MPFrance	BGI_1	01-D12	NA	Guajuato	San Luis de la Paz	BGI_1	oursSAPA
XA097c15	4/20/20	Mod_MPFrance	BGI_1	05-B7	NA	Guajuato	San Luis de la Paz	BGI_1	oursSACE_TESIS
XA097c4	4/20/20	Mod_MPFrance	BGI_1	01-D9	NA	Guajuato	San Luis de la Paz	BGI_1	oursSACE_TESIS
XA098c2	4/20/20	Mod_MPFrance	BGI_1	01-D10	NA	Guajuato	San Felipe	BGI_1	oursSACE_TESIS
XA107c1	4/20/20	Mod_MPFrance	BGI_1	01-E9	NA	Jalisco	Las Guasimas Cabo Corrientes	BGI_1	oursSACE_TESIS
XA108c1	4/20/20	Mod_MPFrance	BGI_1	01-F3	NA	Jalisco	Las Guasimas Cabo Corrientes	BGI_1	oursSACE_TESIS
XA121c14	4/20/20	Mod_MPFrance	BGI_1	05-C7	NA	Tamaulipas	San Carlos	BGI_1	oursSACE_TESIS
XA121c18	4/20/20	Mod_MPFrance	BGI_1	05-C8	NA	Tamaulipas	San Carlos	BGI_1	oursSACE_TESIS
XA121c20	4/20/20	Mod_MPFrance	BGI_1	05-C9	NA	Tamaulipas	San Carlos	BGI_1	oursSACE_TESIS
XA121c21	4/20/20	Mod_MPFrance	BGI_1	05-C10	NA	Tamaulipas	San Carlos	BGI_1	oursSACE_TESIS
XA121c6	4/20/20	Mod_MPFrance	BGI_1	05-C6	NA	Tamaulipas	San Carlos	BGI_1	oursSACE_TESIS
XA123c10	4/20/20	Mod_MPFrance	BGI_1	22 C10	NA	Tamaulipas	San Carlos	BGI_1	oursKLMA

Tabla suplementaria 1. Lista de genomas usados

XAI12366	4/20/20	Mod_MPFrance	BGI_1	01-D7	NA	Tamaulipas	San Carlos	BGI_1	oursKLIMA
XAI124C1	4/20/20	Mod_MPFrance	BGI_1	01-G1	NA	Tamaulipas	San Carlos	BGI_1	oursSACE_TESIS
XAI125C5	4/20/20	Mod_MPFrance	BGI_1	05-B4	NA	Tamaulipas	San Carlos	BGI_1	oursSACE_TESIS
XAI126C1	4/20/20	Mod_MPFrance	BGI_1	01-D3	NA	Tamaulipas	San Carlos	BGI_1	oursSACE_TESIS
XAI126C5	4/20/20	Mod_MPFrance	BGI_1	05-B5	NA	Tamaulipas	San Carlos	BGI_1	oursSACE_TESIS
XAI126C7	4/20/20	Mod_MPFrance	BGI_1	05-B6	NA	Tamaulipas	San Carlos	BGI_1	oursSACE_TESIS
XAI131C1	4/20/20	Mod_MPFrance	BGI_1	01-E10	NA	Puebla	Santa Mani#a Xoyatla	BGI_1	oursSACE_TESIS
XAI131C10	4/20/20	Mod_MPFrance	BGI_1	05-B11	NA	Puebla	Santa Mani#a Xoyatla	BGI_1	oursSACE_TESIS
XAI131C13	4/20/20	Mod_MPFrance	BGI_1	05-C2	NA	Puebla	Santa Mani#a Xoyatla	BGI_1	oursSACE_TESIS
XAI131C18	4/20/20	Mod_MPFrance	BGI_1	05-C3	NA	Puebla	Santa Mani#a Xoyatla	BGI_1	oursSACE_TESIS
XAI131C2	4/20/20	Mod_MPFrance	BGI_1	05-B9	NA	Puebla	Santa Mani#a Xoyatla	BGI_1	oursSACE_TESIS
XAI131C20	4/20/20	Mod_MPFrance	BGI_1	05-C4	NA	Puebla	Santa Mani#a Xoyatla	BGI_1	oursSACE_TESIS
XAI131C4	4/20/20	Mod_MPFrance	BGI_1	05-B10	NA	Puebla	Santa Mani#a Xoyatla	BGI_1	oursSACE_TESIS
XAI132C2	4/20/20	Mod_MPFrance	BGI_1	01-F4	NA	Puebla	San Nicolas Huajuapam	BGI_1	oursSACE_TESIS
XAI133C17	4/20/20	Mod_MPFrance	BGI_1	05-D5	NA	Puebla	San Nicolas Huajuapam	BGI_1	oursPIKU
XAI133C5	4/20/20	Mod_MPFrance	BGI_1	01-F1	NA	Puebla	San Nicolas Huajuapam	BGI_1	oursSACE_TESIS
XAI134C1	4/20/20	Mod_MPFrance	BGI_1	01-E5	NA	Puebla	San Nicolas Huajuapam	BGI_1	oursSACE_TESIS
XAI134C18	4/20/20	Mod_MPFrance	BGI_1	26 F6	NA	Puebla	San Nicolas Huajuapam	BGI_1	oursKLIMA
XAI137C1	4/20/20	Mod_MPFrance	BGI	X137.1	NA	Oaxaca	San Juan del Rv#o	BGI_1.5	oursSACE_TESIS
XAI139C1	4/20/20	Mod_MPFrance	BGI_1	X139.1	NA	Oaxaca	San Juan del Rv#o	BGI_1.5	oursSACE_TESIS
XAI140C14	4/20/20	Mod_MPFrance	BGI_1	26 D2	NA	Oaxaca	San Juan del Rv#o	BGI_1	oursKLIMA
XAI140C2	4/20/20	Mod_MPFrance	BGI	X140.2	NA	Oaxaca	San Juan del Rv#o	BGI_1.5	oursSACE_TESIS
XAI142C1	4/20/20	Mod_MPFrance	BGI_1	01-H3	NA	Oaxaca	San Juan del Rv#o	BGI_1	oursSACE_TESIS
XAI143C7	4/20/20	Mod_MPFrance	BGI_1	01-E12	NA	Oaxaca	San Luis Amatlan	BGI_1	oursSACE_TESIS
XAI144C3	4/20/20	Mod_MPFrance	BGI_1	01-F2	NA	Oaxaca	San Luis Amatlan	BGI_1	oursSACE_TESIS
XAI145C3	4/20/20	Mod_MPFrance	BGI_1	01-E7	NA	Oaxaca	San Luis Amatlan	BGI_1	oursSACE_TESIS
XAI146C1	4/20/20	Mod_MPFrance	BGI_1	01-E8	NA	Oaxaca	San Agust#n Amatengo	BGI_1	oursSACE_TESIS
XAI147C2	4/20/20	Mod_MPFrance	BGI_1	01-F6	NA	Oaxaca	San Agust#n Amatengo	BGI_1	oursSACE_TESIS
XAI151C1	4/20/20	Mod_MPFrance	BGI_1	32 E1	NA	Jalisco	Las Guasimas Cabo Corrientes	BGI_1	oursKLIMA
XAI151C6	4/20/20	Mod_MPFrance	BGI_1	01-F7	NA	Jalisco	Las Guasimas Cabo Corrientes	BGI_1	oursSACE_TESIS
XAI153C1	4/20/20	Mod_MPFrance	BGI_1	01-G4	NA	Jalisco	Mascota	BGI_1	oursSACE_TESIS
XAI154C1	4/20/20	Mod_MPFrance	BGI_1	01-G8	NA	Jalisco	Mascota	BGI_1	oursSACE_TESIS
XAI156C1	4/20/20	Mod_MPFrance	BGI_1	01-H5	NA	Jalisco	La estancia	BGI_1	oursSACE_TESIS
XAI157C21	4/20/20	Mod_MPFrance	BGI_1	33 F9	NA	Jalisco	La Estancia	BGI_1	oursKLIMA
XAI162C1	4/20/20	Mod_MPFrance	BGI_1	02-A1	NA	Jalisco	La estancia	BGI_1	oursUNKNOWN
XAI181C5	4/20/20	Mod_MPFrance	BGI_1	02-A2	NA	Durango	Nombre de Dios	BGI_1	oursUNKNOWN
XAI182C2	4/20/20	Mod_MPFrance	BGI_1	02-A3	NA	Durango	Nombre de Dios	BGI_1	oursHYB
XAI183C1	4/20/20	Mod_MPFrance	BGI_1	02-A4	NA	Durango	Nombre de Dios	BGI_1	oursHYB
XAI184C6	4/20/20	Mod_MPFrance	BGI_1	02-A5	NA	Durango	Nombre de Dios	BGI_1	oursHYB
XAI186C3	4/20/20	Mod_MPFrance	BGI_1	02-A7	NA	Durango	Nombre de Dios	BGI_1	oursSACE_TESIS
XAI187C1	4/20/20	Mod_MPFrance	BGI_1	02-A8	NA	Durango	Nombre de Dios	BGI_1	oursSACE_TESIS
XAI189C1	4/20/20	Mod_MPFrance	BGI_1	02-A9	NA	Durango	Nombre de Dios	BGI_1	oursSACE_TESIS
XAI190C1	4/20/20	Mod_MPFrance	BGI_1	02-A10	NA	Durango	Nombre de Dios	BGI_1	oursSACE_TESIS
XAI190C17	4/20/20	Mod_MPFrance	BGI_1	31 F5	NA	Durango	Nombre de Dios	BGI_1	oursKLIMA
XAI191C2	4/20/20	Mod_MPFrance	BGI_1	02-A11	NA	Durango	Nombre de Dios	BGI_1	oursSACE_TESIS
XAI192C4	4/20/20	Mod_MPFrance	BGI_1	02-A12	NA	Durango	Nombre de Dios	BGI_1	oursHYB
XAI195C1	4/20/20	Mod_MPFrance	BGI	X195.1	NA	Durango	Yonora, Mezquital	BGI_1.5	oursSACE_TESIS
XAI196C1	4/20/20	Mod_MPFrance	BGI_1	02-B2	NA	Durango	Yonora Mezquital	BGI_1	oursSACE_TESIS
XAI197C1	4/20/20	Mod_MPFrance	BGI_1	02-B3	NA	Durango	Yonora Mezquital	BGI_1	oursSACE_TESIS
XAI198C9	4/20/20	Mod_MPFrance	BGI_1	02-B4	NA	Durango	Yonora Mezquital	BGI_1	oursPIKU

Tabla suplementaria 1. Lista de genomas usados

XA199c1	4/20/20	Mod_MPFrance	BGI_1	02-B5	NA	Durango	Temohaya Mezquital	BGI_1	oursPIKU
XA200c2	4/20/20	Mod_MPFrance	BGI_1	02-B6	NA	Durango	Temohaya Mezquital	BGI_1	oursSACE_TESIS
XA201c1	4/20/20	Mod_MPFrance	BGI_1	02-B7	NA	Durango	Temohaya Mezquital	BGI_1	oursSACE_TESIS
XA211c1	4/20/20	Mod_MPFrance	BGI_1	02-B8	NA	Zacatecas	Huitzila	BGI_1	oursSACE_TESIS
XA251c2	4/20/20	Mod_MPFrance	BGI_1	01-A1	NA	Oaxaca	Matativ'n	BGI_1	oursSACE_TESIS
XA252c3	4/20/20	Mod_MPFrance	BGI_1	01-A9	NA	Oaxaca	Matativ'n	BGI_1	oursSACE_TESIS
XA254c23	4/20/20	Mod_MPFrance	BGI_1	05-D2	NA	Michoacan	Piedras de lumbre	BGI_1	oursSACE_TESIS
XA255c1	4/20/20	Mod_MPFrance	BGI_1	01-A2	NA	Oaxaca	Matativ'n	BGI_1	oursSACE_TESIS
XA256c15	4/20/20	Mod_MPFrance	BGI_1	01-B1	NA	Michoacan	Rv'ro de Parras	BGI_1	oursHYB
XA257c1	4/20/20	Mod_MPFrance	BGI_1	35 G1	NA	Michoacan	Los Lobos	BGI_1	oursKLIMA
XA258c4	4/20/20	Mod_MPFrance	BGI_1	01-A10	NA	Oaxaca	Matativ'n	BGI_1	oursSACE_TESIS
XA261c23	4/20/20	Mod_MPFrance	BGI_1	01-A7	NA	Oaxaca	Matativ'n	BGI_1	oursSACE_TESIS
XB001c4	4/20/20	Mod_MPFrance	BGI_1	01-A5	NA	Oaxaca	Matativ'n	BGI_1	oursSACE_TESIS
XB002c8	Dec-18	NextSeq	NextSeq	G506-M8-NextSeq8	NA	2 Guanajuato	Mezcal	ymeZ	oursSACE_TESIS
XB003c8	4/20/20	Mod_MPFrance	BGI_1	M3M006	NA	San Luis Potosi	San Luis Potosi	BGI_1	oursHYB
XB003s8	Dec-18	NextSeq	NextSeq	G501-M6-NextSeq3	NA	3 SLP	Mezcal	ymeZ	oursHYB
XB005c7	Dec-18	NextSeq	NextSeq	G511-M7-NextSeq13	NA	SLP	Mezcal	ymeZ	oursPIKU
XB006c2	4/20/20	Mod_MPFrance	BGI_1	01-B11	NA	San Luis Potosi	San Luis Potosi	BGI_1	oursSACE_TESIS
XB007c9	Dec-18	NextSeq	NextSeq	G514-M21-NextSeq16	NA	2 Guanajuato	Mezcal	ymeZ	oursSAPA
XB008c4	4/20/20	Mod_MPFrance	BGI_1	01-E1	NA	Guanajuato	San Luis de la Paz	BGI_1	oursHYB
XB011c1	4/20/20	Mod_MPFrance	BGI_1	01-D4	NA	Michoacan	Etucuario	BGI_1	oursSACE_TESIS
XB011c2	Dec-18	MiSeq	MiSeq	G530-M28-MiSeq8	NA	1.8 Michoacan	Mezcal	ymeZ	oursSACE_TESIS
XB011s1	Dec-18	MiSeq	MiSeq	G534-M30-MiSeq12	NA	1.9 Michoacan	Mezcal	ymeZ	oursSACE_TESIS
XB041c1	4/20/20	Mod_MPFrance	BGI_1	01-A3	NA	Oaxaca	Santa Catarina de Albarradas	BGI_1	oursSACE_TESIS
XB041c3	Dec-18	MiSeq	MiSeq	G528-M12-MiSeq6	NA	4 Oaxaca	Mezcal	ymeZ	oursSACE_TESIS
XB042c1	4/20/20	Mod_MPFrance	BGI_1	01-B2	NA	Oaxaca	Santa Catarina de Albarradas	BGI_1	oursSACE_TESIS
XB044c3	Dec-18	NextSeq	NextSeq	G519-M11-NextSeq21	NA	2 Oaxaca	Mezcal	ymeZ	oursSACE_TESIS
XB044c4	4/20/20	Mod_MPFrance	BGI_1	01-B3	NA	Oaxaca	Matativ'n	BGI_1	oursSACE_TESIS
XB045c4	4/20/20	Mod_MPFrance	BGI_1	01-A4	NA	Oaxaca	Matativ'n	BGI_1	oursSACE_TESIS
XB047c2	4/20/20	Mod_MPFrance	BGI_1	01-H7	NA	Oaxaca	Matativ'n	BGI_1	oursSACE_TESIS
XB048c1	4/20/20	Mod_MPFrance	BGI_1	01-B4	NA	Oaxaca	Matativ'n	BGI_1	oursSACE_TESIS
XB049c2	Dec-18	MiSeq	MiSeq	G527-M24-MiSeq5	NA	1.7 Oaxaca	Mezcal	ymeZ	oursSACE_TESIS
XB050c1	4/20/20	Mod_MPFrance	BGI_1	01-C9	NA	Oaxaca	Matativ'n	BGI_1	oursSACE_TESIS
XB051c4	4/20/20	Mod_MPFrance	BGI_1	01-D5	NA	Oaxaca	Matativ'n	BGI_1	oursSACE_TESIS
XB052c1	4/20/20	Mod_MPFrance	BGI_1	01-B9	NA	Oaxaca	Matativ'n	BGI_1	oursSACE_TESIS
XB053c1	4/20/20	Mod_MPFrance	BGI_1	01-B6	NA	Oaxaca	Matativ'n	BGI_1	oursSACE_TESIS
XB054c5	4/20/20	Mod_MPFrance	BGI_1	01-C5	NA	Oaxaca	Matativ'n	BGI_1	oursSACE_TESIS
XB054c8	Dec-18	NextSeq	NextSeq	G518-M10-NextSeq20	NA	3 Oaxaca	Mezcal	ymeZ	oursSACE_TESIS
XB056c1	4/20/20	Mod_MPFrance	BGI_1	01-B5	NA	Oaxaca	Mihuatv'n	BGI_1	oursSACE_TESIS
XB056c3	Dec-18	NextSeq	NextSeq	G507-M9-NextSeq9	NA	2 Oaxaca	Mezcal	ymeZ	oursSACE_TESIS
XB063c2	4/20/20	Mod_MPFrance	BGI_1	01-C2	NA	Oaxaca	Santa Catarina de Minas	BGI_1	oursSACE_TESIS
XB064c2	4/20/20	Mod_MPFrance	BGI_1	01-C3	NA	Oaxaca	Santa Catarina de Minas	BGI_1	oursSACE_TESIS
XB065c1	4/20/20	Mod_MPFrance	BGI_1	01-C4	NA	Oaxaca	Santa Catarina de Minas	BGI_1	oursSACE_TESIS
XB066c7	Dec-18	NextSeq	NextSeq	G509-M4-NextSeq11	NA	2 Oaxaca	Mezcal	ymeZ	oursSACE_TESIS
XB067c1	4/20/20	Mod_MPFrance	BGI_1	01-E3	NA	Oaxaca	Jayacatv'n	BGI_1	oursSAPA
XB069c1	4/20/20	Mod_MPFrance	BGI_1	01-H8	NA	Oaxaca	Jayacatv'n	BGI_1	oursSACE_TESIS
XB069s1	Dec-18	NextSeq	NextSeq	G516-M25-NextSeq18	NA	2 Oaxaca	Mezcal	ymeZ	oursSACE_TESIS
XB071c2	Dec-18	NextSeq	NextSeq	G508-M13-NextSeq10	NA	2 Oaxaca	Mezcal	ymeZ	oursSACE_TESIS
XB073c4	Dec-18	NextSeq	NextSeq	G524-M14-MiSeq2	NA	2 Oaxaca	Mezcal	ymeZ	oursSACE_TESIS
XB073c8	4/20/20	Mod_MPFrance	BGI_1	01-G11	NA	Oaxaca	Santa Catarina Ixcatv'n	BGI_1	oursSACE_TESIS

Tabla suplementaria 1. Lista de genomas usados

X8075c7	4/20/20	Mod_MPFrance	BGI_1	01-D8	NA	Oaxaca	Santa Catarina Ixcativ'n	BGI_1	oursSACE_TESIS
X8076c1	4/20/20	Mod_MPFrance	BGI_1	01-D6	NA	Guerrero	Apetiv'n	BGI_1	oursSACE_TESIS
X8076c9	Dec-18	Dec-18	MiSeq	G529-M26-MiSeq7	1.8	Guerrero	Mezcal	ymez	oursSACE_TESIS
X8077c1	4/20/20	Mod_MPFrance	BGI_1	01-B12	NA	Guerrero	Apetiv'n	BGI_1	oursSAPA
X8077c4	Dec-18	Dec-18	NextSeq	G515-M22-NextSeq17	2	Guerrero	Mezcal	ymez	oursSACE_TESIS
X8077c5	Dec-18	Dec-18	NextSeq	G500-M27-NextSeq2	1.8	Guerrero	Mezcal	ymez	oursSACE_TESIS
X8094c1	Dec-18	Dec-18	NextSeq	G510-M16-NextSeq12	2	Jalisco	Mezcal	ymez	oursSACE_TESIS
X8094c2	4/20/20	Mod_MPFrance	BGI_1	01-C10	NA	Jalisco	San Juan Espanatica	BGI_1	oursSACE_TESIS
X8094s2	Dec-18	Dec-18	NextSeq	G513-M19-NextSeq15	2	Jalisco	Mezcal	ymez	oursSACE_TESIS
X8250c6	Dec-18	Dec-18	NextSeq	G517-M29-NextSeq19	2	Michoacan	Mezcal	ymez	oursSACE_TESIS
X8251c1	4/20/20	Mod_MPFrance	BGI_1	01-A12	NA	Michoacan	Morelia Pino Bonito	BGI_1	oursSACE_TESIS
X8252c1	4/20/20	Mod_MPFrance	BGI_1	01-A8	NA	Michoacan	Morelia Pino Bonito	BGI_1	oursSACE_TESIS
X8253c6	Dec-18	Dec-18	NextSeq	G512-M18-NextSeq14	2	Oaxaca	Mezcal	ymez	oursSACE_TESIS
X8254c1	4/20/20	Mod_MPFrance	BGI_1	01-H11	NA	Michoacan	Piedras de lumbre	BGI_1	oursSACE_TESIS
X8256c3	4/20/20	Mod_MPFrance	BGI_1	01-A6	NA	Oaxaca	Matativ'n	BGI_1	oursSACE_TESIS
X8257c6	Dec-18	Dec-18	MiSeq	G533-M17-MiSeq11	2.7	Michoacan	Mezcal	ymez	oursSAPA
X8258c4	4/20/20	Mod_MPFrance	BGI_1	01-A11	NA	Oaxaca	Matativ'n	BGI_1	oursSACE_TESIS

Tabla Suplementaria 2

Tabla suplementaria 2. Lista de lecturas alineadas a cada especie de la referencia concatenada.

AISLADO	SACE	SAPA	SAMI	SAJU	SAKU	SAAR	SAUV	SAEU	KLMA	PIKU	Total	Especie
DK001c10	96.18%	3.16%	0.10%	0.21%	0.03%	0.06%	0.00%	0.26%	0.00%	0.00%	2.34E+09	SACE
DK001c27	96.10%	2.98%	0.46%	0.07%	0.03%	0.12%	0.00%	0.23%	0.00%	0.00%	2.04E+09	SACE
DK002c39	32.73%	66.71%	0.06%	0.12%	0.06%	0.06%	0.01%	0.24%	0.01%	0.00%	2.22E+09	SACE-SAPA
DK003c30	32.83%	66.65%	0.05%	0.08%	0.05%	0.07%	0.01%	0.25%	0.01%	0.00%	2.12E+09	SACE-SAPA
DK003c56	96.33%	3.25%	0.07%	0.06%	0.02%	0.05%	0.00%	0.21%	0.01%	0.01%	2.11E+09	SACE
DK016c5	96.08%	3.18%	0.34%	0.07%	0.02%	0.04%	0.00%	0.27%	0.00%	0.00%	2.05E+09	SACE
DK017c3	0.09%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.88%	0.01%	2.19E+09	KLMA
DK017c4	96.53%	3.05%	0.06%	0.06%	0.03%	0.03%	0.00%	0.23%	0.01%	0.00%	2.15E+09	SACE
DK018c4	96.53%	2.90%	0.08%	0.07%	0.03%	0.04%	0.00%	0.29%	0.04%	0.00%	2.02E+09	SACE
DK019c1	96.84%	2.99%	0.08%	0.05%	0.02%	0.02%	0.00%	0.00%	0.00%	0.00%	2.15E+09	SACE
DK030c8	8.60%	7.12%	6.92%	8.03%	8.81%	9.67%	8.14%	9.09%	20.22%	13.41%	6.32E+07	INCONCLUSO
DS002c10	96.83%	0.85%	0.30%	0.08%	1.83%	0.10%	0.00%	0.01%	0.00%	0.00%	2.15E+09	SACE
DS002c4	96.63%	2.93%	0.27%	0.03%	0.01%	0.12%	0.00%	0.00%	0.00%	0.00%	2.18E+09	SACE
DS002c5	96.27%	3.16%	0.36%	0.04%	0.01%	0.15%	0.00%	0.00%	0.00%	0.00%	2.06E+09	SACE
DS002c6	96.13%	3.25%	0.39%	0.04%	0.01%	0.17%	0.00%	0.00%	0.00%	0.00%	2.22E+09	SACE
XA017c1	96.35%	3.03%	0.24%	0.07%	0.02%	0.06%	0.00%	0.22%	0.00%	0.00%	2.12E+09	SACE
XA020c1	70.73%	28.38%	0.30%	0.09%	0.03%	0.07%	0.01%	0.39%	0.00%	0.00%	2.30E+09	SACE-SAPA
XA022c2	64.89%	34.52%	0.38%	0.05%	0.03%	0.11%	0.01%	0.00%	0.01%	0.01%	1.82E+09	SACE-SAPA
XA023c1	96.12%	3.13%	0.30%	0.08%	0.02%	0.07%	0.00%	0.27%	0.01%	0.00%	1.77E+09	SACE
XA024c2	96.01%	3.19%	0.33%	0.08%	0.03%	0.07%	0.00%	0.28%	0.00%	0.00%	1.91E+09	SACE
XA030c1	96.50%	2.91%	0.17%	0.08%	0.03%	0.05%	0.01%	0.25%	0.01%	0.01%	2.02E+09	SACE
XA031c2	96.70%	2.61%	0.37%	0.19%	0.02%	0.09%	0.00%	0.00%	0.01%	0.01%	2.05E+09	SACE
XA033c2	0.04%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.95%	2.09E+09	PIKU
XA055c2	96.45%	3.06%	0.19%	0.21%	0.03%	0.04%	0.00%	0.00%	0.01%	0.00%	2.19E+09	SACE
XA057c4	95.61%	3.74%	0.11%	0.12%	0.04%	0.10%	0.00%	0.26%	0.01%	0.01%	2.11E+09	SACE
XA063c8	0.01%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.96%	0.01%	2.23E+09	KLMA
XA070c1	96.26%	2.81%	0.51%	0.07%	0.03%	0.04%	0.00%	0.26%	0.01%	0.01%	2.01E+09	SACE
XA072c4	95.82%	3.32%	0.26%	0.23%	0.03%	0.04%	0.00%	0.27%	0.02%	0.01%	2.20E+09	SACE
XA074c2	96.13%	3.02%	0.46%	0.25%	0.02%	0.10%	0.00%	0.00%	0.01%	0.01%	2.18E+09	SACE
XA076c20	0.03%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.96%	0.00%	2.25E+09	KLMA
XA081c6	96.80%	2.74%	0.24%	0.07%	0.02%	0.12%	0.00%	0.00%	0.01%	0.00%	2.23E+09	SACE
XA082c2	52.99%	43.53%	0.47%	0.33%	2.06%	0.38%	0.01%	0.23%	0.01%	0.00%	2.07E+09	SACE-SAPA
XA082c24	0.05%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.94%	0.00%	2.27E+09	KLMA
XA083c1	96.41%	2.88%	0.36%	0.10%	0.03%	0.19%	0.00%	0.00%	0.02%	0.00%	2.20E+09	SACE
XA083c6	96.33%	3.08%	0.37%	0.13%	0.03%	0.05%	0.00%	0.00%	0.00%	0.00%	1.98E+09	SACE
XA096c3	0.52%	98.49%	0.07%	0.56%	0.04%	0.26%	0.03%	0.02%	0.02%	0.00%	2.17E+09	SAPA
XA097c15	96.10%	2.87%	0.22%	0.10%	0.07%	0.12%	0.04%	0.26%	0.10%	0.11%	1.76E+09	SACE
XA097c4	96.14%	3.03%	0.26%	0.09%	0.03%	0.10%	0.00%	0.32%	0.02%	0.01%	2.17E+09	SACE
XA098c2	96.60%	2.90%	0.16%	0.05%	0.03%	0.04%	0.00%	0.20%	0.02%	0.01%	2.00E+09	SACE
XA107c1	95.86%	2.87%	0.73%	0.09%	0.03%	0.06%	0.00%	0.35%	0.01%	0.01%	2.12E+09	SACE

Tabla suplementaria 2. Lista de lecturas alineadas a cada especie de la referencia concatenada.

XA108c1	94.27%	4.80%	0.36%	0.06%	0.04%	0.16%	0.00%	0.27%	0.02%	0.01%	2.18E+09	SACE
XA121c14	94.41%	2.60%	0.39%	0.26%	0.10%	0.18%	0.06%	0.23%	0.11%	1.67%	1.65E+09	SACE
XA121c18	95.90%	2.95%	0.46%	0.28%	0.03%	0.13%	0.00%	0.23%	0.00%	0.02%	2.10E+09	SACE
XA121c20	97.32%	2.23%	0.17%	0.12%	0.02%	0.05%	0.00%	0.07%	0.01%	0.00%	2.09E+09	SACE
XA121c21	91.29%	4.90%	0.38%	0.24%	0.10%	0.18%	0.06%	0.25%	0.11%	2.48%	1.66E+09	SACE
XA121c6	97.28%	2.23%	0.18%	0.13%	0.02%	0.06%	0.00%	0.09%	0.00%	0.00%	2.15E+09	SACE
XA123c10	0.06%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.93%	0.00%	2.25E+09	KLMA
XA123c6	0.05%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.93%	0.00%	2.09E+09	KLMA
XA124c1	97.96%	1.11%	0.11%	0.23%	0.02%	0.31%	0.00%	0.24%	0.01%	0.00%	2.27E+09	SACE
XA125c5	97.40%	1.32%	0.51%	0.30%	0.03%	0.15%	0.00%	0.27%	0.00%	0.01%	2.12E+09	SACE
XA126c1	97.89%	1.37%	0.19%	0.12%	0.02%	0.22%	0.00%	0.17%	0.01%	0.01%	2.21E+09	SACE
XA126c5	98.90%	0.67%	0.14%	0.09%	0.03%	0.13%	0.01%	0.01%	0.02%	0.01%	2.07E+09	SACE
XA126c7	97.04%	2.44%	0.19%	0.12%	0.02%	0.06%	0.00%	0.11%	0.01%	0.00%	2.11E+09	SACE
XA131c1	95.71%	3.30%	0.53%	0.08%	0.03%	0.06%	0.00%	0.28%	0.00%	0.00%	2.12E+09	SACE
XA131c10	96.34%	3.07%	0.30%	0.06%	0.02%	0.04%	0.00%	0.15%	0.01%	0.01%	2.24E+09	SACE
XA131c13	96.40%	2.91%	0.38%	0.05%	0.03%	0.03%	0.00%	0.19%	0.01%	0.00%	2.15E+09	SACE
XA131c18	96.48%	2.74%	0.43%	0.06%	0.03%	0.04%	0.00%	0.21%	0.01%	0.01%	2.21E+09	SACE
XA131c2	96.26%	3.03%	0.37%	0.06%	0.02%	0.05%	0.00%	0.18%	0.02%	0.01%	2.17E+09	SACE
XA131c20	95.90%	2.75%	0.78%	0.09%	0.04%	0.06%	0.00%	0.37%	0.01%	0.01%	2.08E+09	SACE
XA131c4	96.24%	3.09%	0.35%	0.07%	0.02%	0.05%	0.00%	0.17%	0.01%	0.01%	2.16E+09	SACE
XA132c2	96.08%	3.15%	0.27%	0.07%	0.03%	0.15%	0.00%	0.23%	0.01%	0.01%	2.14E+09	SACE
XA133c17	0.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.89%	2.19E+09	PIKU
XA133c5	95.78%	2.95%	0.47%	0.08%	0.03%	0.06%	0.00%	0.24%	0.01%	0.38%	1.93E+09	SACE
XA134c1	96.06%	2.87%	0.54%	0.06%	0.03%	0.13%	0.00%	0.30%	0.00%	0.00%	1.67E+09	SACE
XA134c18	0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.97%	0.01%	2.22E+09	KLMA
XA137c1	96.84%	2.55%	0.32%	0.08%	0.02%	0.18%	0.00%	0.00%	0.00%	0.00%	2.20E+09	SACE
XA139c1	97.02%	2.41%	0.23%	0.06%	0.02%	0.13%	0.00%	0.00%	0.01%	0.13%	2.15E+09	SACE
XA140c14	0.04%	0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.93%	0.00%	2.28E+09	KLMA
XA140c2	95.99%	2.80%	0.62%	0.07%	0.03%	0.15%	0.00%	0.34%	0.01%	0.00%	2.05E+09	SACE
XA142c1	96.03%	3.09%	0.28%	0.16%	0.04%	0.12%	0.00%	0.27%	0.00%	0.00%	1.86E+09	SACE
XA143c7	95.69%	3.61%	0.16%	0.11%	0.04%	0.09%	0.00%	0.29%	0.01%	0.01%	2.08E+09	SACE
XA144c3	95.86%	3.43%	0.15%	0.10%	0.04%	0.10%	0.00%	0.29%	0.02%	0.01%	2.00E+09	SACE
XA145c3	96.00%	3.30%	0.12%	0.18%	0.03%	0.05%	0.00%	0.31%	0.01%	0.00%	2.07E+09	SACE
XA146c1	96.26%	2.79%	0.28%	0.18%	0.02%	0.15%	0.00%	0.31%	0.00%	0.00%	2.08E+09	SACE
XA147c2	96.17%	3.08%	0.40%	0.10%	0.03%	0.21%	0.00%	0.00%	0.00%	0.00%	2.13E+09	SACE
XA151c1	0.02%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.96%	0.01%	2.16E+09	KLMA
XA151c6	94.32%	4.74%	0.35%	0.06%	0.04%	0.16%	0.00%	0.30%	0.02%	0.00%	2.12E+09	SACE
XA153c1	97.12%	2.64%	0.07%	0.11%	0.02%	0.02%	0.00%	0.00%	0.00%	0.00%	2.19E+09	SACE
XA154c1	95.80%	3.36%	0.33%	0.09%	0.03%	0.06%	0.00%	0.32%	0.00%	0.00%	2.11E+09	SACE
XA156c1	96.58%	2.92%	0.32%	0.05%	0.02%	0.11%	0.00%	0.00%	0.00%	0.00%	2.10E+09	SACE
XA157c21	0.07%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.92%	0.00%	2.16E+09	KLMA

Tabla suplementaria 2. Lista de lecturas alineadas a cada especie de la referencia concatenada.

XA162c1	16.67%	0.46%	0.01%	0.03%	0.01%	0.01%	0.00%	0.00%	82.81%	0.01%	2.06E+09	INCONCLUSO
XA181c5	37.49%	20.63%	0.25%	0.04%	0.02%	0.07%	0.01%	0.00%	0.02%	41.46%	2.14E+09	INCONCLUSO
XA182c2	63.66%	35.83%	0.32%	0.04%	0.03%	0.09%	0.01%	0.00%	0.01%	0.00%	2.27E+09	SACE-SAPA
XA183c1	62.86%	36.40%	0.51%	0.04%	0.03%	0.15%	0.01%	0.00%	0.00%	0.00%	2.11E+09	SACE-SAPA
XA184c6	62.82%	36.48%	0.46%	0.05%	0.04%	0.13%	0.01%	0.00%	0.01%	0.00%	2.18E+09	SACE-SAPA
XA186c3	95.56%	3.67%	0.46%	0.07%	0.01%	0.21%	0.00%	0.00%	0.01%	0.00%	2.17E+09	SACE
XA187c1	96.33%	3.43%	0.06%	0.06%	0.04%	0.06%	0.00%	0.01%	0.00%	0.00%	2.23E+09	SACE
XA189c1	95.46%	3.99%	0.33%	0.05%	0.02%	0.13%	0.00%	0.00%	0.00%	0.00%	2.20E+09	SACE
XA190c1	95.65%	3.94%	0.25%	0.04%	0.02%	0.09%	0.00%	0.00%	0.00%	0.00%	2.21E+09	SACE
XA190c17	0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.96%	0.01%	2.11E+09	KLIMA
XA191c2	95.47%	3.96%	0.36%	0.05%	0.02%	0.13%	0.00%	0.00%	0.00%	0.00%	2.18E+09	SACE
XA192c4	48.73%	50.70%	0.30%	0.10%	0.03%	0.08%	0.02%	0.01%	0.04%	0.00%	2.08E+09	SACE-SAPA
XA195c1	96.44%	3.10%	0.28%	0.04%	0.02%	0.10%	0.00%	0.00%	0.00%	0.00%	2.10E+09	SACE
XA196c1	96.39%	3.10%	0.30%	0.05%	0.02%	0.11%	0.00%	0.00%	0.03%	0.01%	2.22E+09	SACE
XA197c1	95.79%	3.71%	0.31%	0.05%	0.02%	0.11%	0.00%	0.00%	0.00%	0.00%	2.18E+09	SACE
XA198c9	2.83%	1.72%	1.42%	1.28%	1.51%	1.57%	1.18%	1.39%	2.40%	84.71%	3.66E+07	PIKU
XA199c1	2.02%	1.49%	1.38%	1.20%	1.26%	1.59%	1.03%	1.24%	1.98%	86.81%	1.76E+08	PIKU
XA200c2	95.88%	3.49%	0.37%	0.06%	0.01%	0.17%	0.00%	0.00%	0.01%	0.01%	2.17E+09	SACE
XA201c1	96.29%	3.16%	0.35%	0.05%	0.02%	0.12%	0.00%	0.00%	0.01%	0.01%	2.20E+09	SACE
XA211c1	95.65%	3.70%	0.39%	0.06%	0.01%	0.18%	0.00%	0.00%	0.00%	0.00%	2.13E+09	SACE
XA251c2	96.56%	2.56%	0.24%	0.15%	0.02%	0.21%	0.00%	0.19%	0.05%	0.00%	1.95E+09	SACE
XA252c3	96.44%	2.66%	0.44%	0.07%	0.02%	0.14%	0.00%	0.23%	0.00%	0.00%	2.09E+09	SACE
XA254c23	97.03%	2.28%	0.20%	0.08%	0.01%	0.25%	0.00%	0.14%	0.01%	0.00%	2.04E+09	SACE
XA255c1	96.78%	2.76%	0.23%	0.07%	0.02%	0.13%	0.00%	0.00%	0.00%	0.00%	1.75E+09	SACE
XA256c15	57.01%	42.41%	0.18%	0.02%	0.02%	0.08%	0.00%	0.18%	0.08%	0.00%	1.98E+09	SACE-SAPA
XA257c1	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.98%	0.00%	2.23E+09	KLIMA
XA258c4	96.62%	2.57%	0.44%	0.10%	0.03%	0.23%	0.00%	0.00%	0.01%	0.00%	1.71E+09	SACE
XA261c23	96.53%	2.55%	0.27%	0.16%	0.03%	0.23%	0.00%	0.20%	0.02%	0.00%	2.03E+09	SACE
XB001c4	96.68%	2.15%	0.34%	0.21%	0.04%	0.34%	0.00%	0.23%	0.01%	0.01%	2.12E+09	SACE
XB002c8	96.34%	2.41%	0.57%	0.54%	0.08%	0.06%	0.00%	0.00%	0.00%	0.00%	1.43E+09	SACE
XB003c8	63.06%	36.68%	0.09%	0.02%	0.02%	0.05%	0.00%	0.07%	0.01%	0.00%	2.00E+09	SACE-SAPA
XB003s8	35.78%	63.66%	0.35%	0.02%	0.02%	0.03%	0.02%	0.13%	0.00%	0.00%	9.96E+08	SACE-SAPA
XB005c7	4.59%	1.44%	0.96%	0.95%	0.86%	1.27%	0.59%	0.78%	1.43%	87.11%	1.03E+08	PIKU
XB006c2	96.12%	3.58%	0.10%	0.10%	0.03%	0.06%	0.00%	0.00%	0.01%	0.00%	1.88E+09	SACE
XB007c9	1.58%	96.23%	0.73%	0.24%	0.08%	0.80%	0.03%	0.31%	0.00%	0.00%	1.14E+09	SAPA
XB008c4	52.22%	45.72%	0.33%	0.28%	0.99%	0.36%	0.00%	0.09%	0.00%	0.00%	1.98E+09	SACE-SAPA
XB011c1	96.55%	2.83%	0.24%	0.28%	0.03%	0.05%	0.00%	0.00%	0.00%	0.00%	2.10E+09	SACE
XB011c2	96.80%	2.37%	0.32%	0.39%	0.04%	0.06%	0.00%	0.00%	0.00%	0.00%	1.94E+09	SACE
XB011s1	96.70%	2.32%	0.42%	0.41%	0.06%	0.08%	0.00%	0.00%	0.00%	0.00%	3.24E+08	SACE
XB041c1	95.10%	3.90%	0.57%	0.22%	0.03%	0.17%	0.00%	0.00%	0.01%	0.00%	2.02E+09	SACE
XB041c3	97.07%	2.59%	0.22%	0.06%	0.03%	0.03%	0.00%	0.00%	0.00%	0.00%	6.85E+08	SACE

Tabla suplementaria 2. Lista de lecturas alineadas a cada especie de la referencia concatenada.

XB042c1	16.11%	6.44%	5.83%	6.40%	6.54%	9.69%	6.88%	7.02%	22.29%	12.81%	5.53E+07	INCONCLUSO
XB044c3	96.64%	2.19%	0.36%	0.17%	0.03%	0.36%	0.00%	0.25%	0.00%	0.00%	1.75E+09	SACE
XB044c4	96.93%	2.49%	0.28%	0.08%	0.03%	0.17%	0.00%	0.00%	0.00%	0.00%	2.13E+09	SACE
XB045c4	96.25%	2.41%	0.34%	0.23%	0.01%	0.42%	0.00%	0.33%	0.00%	0.00%	2.05E+09	SACE
XB047c2	96.09%	2.83%	0.62%	0.12%	0.03%	0.29%	0.00%	0.00%	0.00%	0.00%	2.06E+09	SACE
XB048c1	96.35%	2.67%	0.56%	0.11%	0.03%	0.27%	0.00%	0.00%	0.00%	0.00%	2.09E+09	SACE
XB049c2	96.26%	2.72%	0.14%	0.39%	0.04%	0.03%	0.00%	0.41%	0.00%	0.00%	3.24E+08	SACE
XB050c1	95.01%	2.73%	1.28%	0.08%	0.05%	0.29%	0.00%	0.54%	0.01%	0.01%	1.95E+09	SACE
XB051c4	95.88%	2.59%	0.84%	0.06%	0.04%	0.19%	0.00%	0.39%	0.01%	0.01%	1.94E+09	SACE
XB052c1	96.09%	2.62%	0.53%	0.27%	0.03%	0.06%	0.00%	0.39%	0.00%	0.00%	2.03E+09	SACE
XB053c1	96.27%	2.57%	0.86%	0.05%	0.04%	0.19%	0.00%	0.01%	0.00%	0.00%	2.09E+09	SACE
XB054c5	96.00%	3.02%	0.33%	0.11%	0.03%	0.10%	0.00%	0.39%	0.00%	0.01%	1.92E+09	SACE
XB054c8	96.81%	2.48%	0.24%	0.08%	0.04%	0.06%	0.00%	0.28%	0.00%	0.00%	1.46E+09	SACE
XB056c1	95.50%	2.84%	0.90%	0.12%	0.04%	0.09%	0.00%	0.49%	0.00%	0.01%	1.92E+09	SACE
XB056c3	94.64%	3.02%	1.01%	0.53%	0.07%	0.10%	0.00%	0.62%	0.00%	0.01%	1.70E+09	SACE
XB063c2	84.26%	2.59%	0.18%	0.28%	0.03%	0.06%	0.00%	0.28%	0.00%	12.31%	2.08E+09	INCONCLUSO
XB064c2	96.50%	2.74%	0.19%	0.07%	0.04%	0.08%	0.00%	0.35%	0.00%	0.02%	2.11E+09	SACE
XB065c1	96.17%	2.85%	0.35%	0.06%	0.04%	0.06%	0.00%	0.45%	0.01%	0.01%	2.11E+09	SACE
XB066c7	96.85%	2.28%	0.33%	0.06%	0.05%	0.03%	0.00%	0.39%	0.00%	0.01%	9.04E+08	SACE
XB067c1	0.49%	98.37%	0.11%	0.53%	0.04%	0.37%	0.05%	0.02%	0.00%	0.00%	2.17E+09	SAPA
XB069c1	96.32%	2.78%	0.50%	0.10%	0.03%	0.25%	0.00%	0.00%	0.00%	0.00%	2.04E+09	SACE
XB069s1	96.19%	2.13%	1.03%	0.12%	0.06%	0.03%	0.00%	0.43%	0.00%	0.01%	1.23E+09	SACE
XB071c2	95.85%	3.32%	0.36%	0.28%	0.10%	0.09%	0.00%	0.00%	0.00%	0.00%	1.48E+09	SACE
XB073c4	95.55%	2.27%	1.08%	0.81%	0.06%	0.23%	0.00%	0.00%	0.00%	0.00%	1.09E+09	SACE
XB073c8	95.75%	3.17%	0.47%	0.09%	0.03%	0.07%	0.00%	0.40%	0.00%	0.01%	1.99E+09	SACE
XB075c7	98.82%	0.30%	0.35%	0.26%	0.02%	0.07%	0.00%	0.18%	0.01%	0.00%	2.17E+09	SACE
XB076c1	95.07%	3.60%	0.65%	0.10%	0.04%	0.06%	0.00%	0.48%	0.01%	0.01%	2.02E+09	SACE
XB076c9	96.18%	3.06%	0.29%	0.24%	0.09%	0.14%	0.00%	0.00%	0.00%	0.00%	4.37E+08	SACE
XB077c1	1.39%	96.41%	0.82%	0.22%	0.12%	0.74%	0.01%	0.28%	0.01%	0.00%	2.19E+09	SAPA
XB077c4	96.97%	2.76%	0.11%	0.09%	0.03%	0.03%	0.00%	0.00%	0.00%	0.00%	1.29E+09	SACE
XB077c5	96.73%	2.80%	0.20%	0.15%	0.06%	0.05%	0.00%	0.00%	0.00%	0.00%	1.11E+09	SACE
XB094c1	94.95%	2.71%	1.22%	0.48%	0.07%	0.05%	0.00%	0.50%	0.00%	0.01%	1.40E+09	SACE
XB094c2	94.56%	3.32%	1.04%	0.42%	0.05%	0.20%	0.00%	0.40%	0.00%	0.01%	2.06E+09	SACE
XB094s2	95.09%	3.15%	1.09%	0.10%	0.05%	0.06%	0.00%	0.45%	0.00%	0.01%	1.81E+09	SACE
XB250c6	96.30%	2.62%	0.66%	0.07%	0.04%	0.04%	0.00%	0.27%	0.00%	0.01%	1.25E+09	SACE
XB251c1	95.58%	2.74%	0.60%	0.15%	0.01%	0.55%	0.00%	0.35%	0.00%	0.00%	2.00E+09	SACE
XB252c1	95.42%	2.80%	0.64%	0.16%	0.01%	0.60%	0.00%	0.36%	0.00%	0.00%	2.03E+09	SACE
XB253c6	95.65%	2.59%	0.57%	0.25%	0.05%	0.54%	0.00%	0.33%	0.00%	0.01%	1.04E+09	SACE
XB254c1	95.08%	2.74%	0.79%	0.20%	0.01%	0.72%	0.00%	0.45%	0.00%	0.00%	2.03E+09	SACE
XB256c3	96.01%	2.66%	0.40%	0.22%	0.03%	0.37%	0.00%	0.31%	0.01%	0.00%	2.02E+09	SACE
XB257c6	1.24%	97.94%	0.05%	0.59%	0.06%	0.01%	0.01%	0.09%	0.00%	0.00%	2.84E+08	SAPA

Tabla suplementaria 2. Lista de lecturas alineadas a cada especie de la referencia concatenada.

XB258c4	95.79%	2.72%	0.76%	0.09%	0.03%	0.22%	0.00%	0.36%	0.01%	0.00%	2.00E+09	SACE
---------	--------	-------	-------	-------	-------	-------	-------	-------	-------	-------	----------	------

SACE	
SAPA	
SACE-SAPA	
KLMA	
PIKU	
INCONCLUSO	
KV	

