Diversidad de levaduras en las fermentaciones abiertas de agave en México

Porfirio Gallegos-Casillas^{1,2}, Luis F. García-Ortega², Adriana Espinosa-Cantú^{1,2,‡}, J. Abraham Avelar-Rivas^{1,2}, Carolina G. Torres-Lagunes^{2,3}, Adrián Cano-Ricardez^{1,2}, Ángela M. García-Acero^{1,4}, Susana Ruiz-Castro², Mayra Flores-Barraza¹, Alejandra Castillo³, Fernando González-Zozaya⁵, América Delgado-Lemus⁶, Francisco Molina-Freaner⁷, Cuauhtémoc Jacques-Hernández⁸, Antonio Hernández-López⁹, Luis Delaye², Xitlali Aguirre-Dugua¹⁰, Manuel R. Kirchmayr¹¹, Lucía Morales^{3,*}, Eugenio Mancera^{2,*} & Alexander DeLuna^{1,*}

¹Unidad de Genómica Avanzada (Langebio), Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Irapuato, México

²Departamento de Ingeniería Genética, Unidad Irapuato, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Irapuato, México

³Laboratorio Internacional de Investigación sobre el Genoma Humano (LIIGH), Universidad Nacional Autónoma de México, Juriquilla, México

⁴Departamento de Ingeniería Química y Ambiental, Universidad Nacional de Colombia, Bogotá, Colombia

⁵Centro INAH Colima, Instituto Nacional de Antropología e Historia, Colima, México

⁶Investigadora independiente, Jesús del Monte, Morelia, México

⁷Departamento de Ecología de la Biodiversidad, Instituto de Ecología, Unidad Hermosillo, Universidad Nacional Autónoma de México, Hermosillo, México

⁸Centro de Biotecnología Genómica, Instituto Politécnico Nacional, Reynosa, México

⁹Escuela Nacional de Estudios Superiores, Unidad León, Universidad Nacional Autónoma de México, León, México

¹⁰Investigadores por México, Consejo Nacional de Humanidades, Ciencias y Tecnologías, CDMX, México

¹¹Unidad de Biotecnología Industrial, Centro de Investigación y Asistencia en Tecnología y Diseño del Estado de Jalisco A.C. (CIATEJ), Zapopan, Jalisco, México

[‡]Adscripción actual: Institute for Biological Physics, University of Cologne, Cologne, Germany

*Autores responsables: Lucía Morales (<u>lmorales@liigh.unam.mx</u>), Eugenio Mancera (<u>eugenio.mancera@cinvestav.mx</u>), Alexander DeLuna (<u>alexander.deluna@cinvestav.mx</u>)

SINOPSIS

- Se aislaron e identificaron 4,524 cepas de levaduras de fermentaciones abiertas de agave en México.
- Se reporta un conjunto central de seis especies de levaduras en todas las regiones productoras.
- Los genomas de *Kazachstania humilis* difieren significativamente de los aislados de la misma especie en otras regiones del mundo.
- Se reportan dos posibles nuevas especies relacionadas con el clado Pichia.

Palabras clave: Diversidad de levaduras; Fermentación de agave; Destilados de agave; Fermentación abierta; Saccharomycetales; *Kazachstania humilis*

RESUMEN

Las levaduras son un grupo diverso de microorganismos fúngicos ampliamente utilizados para producir alimentos y bebidas fermentadas. En México, las fermentaciones abiertas se utilizan para obtener destilados a partir de plantas de agave. A pesar de la prevalencia de esta práctica tradicional en todo el país, las levaduras solo han sido aisladas y estudiadas en un número limitado de destilerías. Con el objetivo de describir sistemáticamente la diversidad de especies de levaduras en las fermentaciones abiertas de agave, hemos creado la colección de cultivos YMX-1.0 mediante el aislamiento de 4,524 cepas de 68 sitios en diversos contextos climáticos, geográficos y biológicos. Usamos espectrometría de masas MALDI-TOF para la clasificación taxonómica y verificamos una fracción de las cepas mediante secuenciación de ITS y D1/D2, lo que también reveló dos posibles nuevas especies de Saccharomycetales. En general, la composición de las comunidades de levaduras mostró cierta relación con variables locales y climáticas, pero un conjunto central de seis especies se aisló de manera recurrente en diversas regiones productoras. Para explorar la variación intraespecífica de las levaduras en las fermentaciones de agave, realizamos la secuenciación del genoma de cuatro aislados de la levadura no convencional Kazachstania humilis. Los genomas de estas cuatro cepas mostraron ser sustancialmente distintos de un aislado europeo de la misma especie, lo que sugiere que podrían pertenecer a poblaciones diferentes. Nuestro trabajo contribuye a la comprensión y a la conservación de un sistema de fermentación abierta de gran importancia cultural y económica, proporcionando un recurso valioso para estudiar la biología y diversidad genética de los microorganismos que viven en la interfaz de ambientes naturales y culturales.

INTRODUCCIÓN

Las levaduras son hongos unicelulares que han desempeñado un papel significativo en la historia humana. Estos microorganismos tienen una amplia distribución geográfica, en parte debido a su uso por parte de los humanos en la producción de alimentos y bebidas fermentadas. Estos microorganismos se encuentran de forma natural en una variedad de climas, ecosistemas y materias primas de fermentación como uvas y cereales (Gallone et al., 2016; Giannakou et al., 2020; Marsit et al., 2017; Molinet & Cubillos, 2020). Los primeros registros arqueológicos de la producción de bebidas fermentadas se remontan a 8,000-3,000 a.C. en China, Sumeria, Irán y Egipto (De Chiara et al., 2022; Marsit et al., 2017; Turker, M. 2015). Mucho antes de que los humanos supieran de su identidad biológica, las levaduras se utilizaron en fermentaciones para producir vino, sidra, licores destilados, pan, queso y otros productos. Las levaduras del orden Saccharomycetales se utilizan en la actualidad en una variedad de aplicaciones biotecnológicas, como la biorremediación, biocatálisis, producción de biocombustibles, producción de proteínas e investigación biomédica (Cubillos, 2016; Johnson & Echavarri-Erasun, 2011; Turker, M. 2015).

La diversidad de levaduras influye en las propiedades de los alimentos y bebidas fermentados. Se sabe que *Saccharomyces cerevisiae* es una especie relevante en el proceso de fermentación, mientras que otras levaduras "no convencionales" tienen el potencial de metabolizar fuentes de carbono poco comunes o de producir metabolitos deseables de interés biotecnológico. Por ejemplo, en la producción de vino, *S. cerevisiae* es responsable de la fermentación alcohólica de la uva, las levaduras *Pichia* producen metabolitos que agregan propiedades organolépticas deseables, mientras que *Brettanomyces bruxellensis* está asociada con sabores no deseados (Burini et al., 2021; Ciani & Comitini, 2011; Gallone et al., 2016; Godoy et al., 2017; Jolly et al., 2014; Kręgiel et al., 2017; Molinet & Cubillos, 2020). Además, las poblaciones de levaduras asociadas a ambientes relacionados con los humanos muestran adaptaciones específicas. Por ejemplo, existen linajes de *S. cerevisiae* en la producción de vino que son tolerantes a la presencia de cobre, sulfatos o concentraciones elevadas de etanol, mientras que otras usadas en la producción de cerveza tienen la capacidad de metabolizar azúcares más complejos como la maltotriosa (Cubillos et al., 2019; De Chiara

et al., 2022; Gallone et al., 2016; Giannakou et al., 2020, 2021; Liti et al., 2009; Peter et al., 2018; Warringer et al., 2011). La domesticación de las levaduras se ha documentado de manera más completa en escenarios extremos de manejo humano, como el uso de cultivos iniciadores de *S. cerevisiae* en la producción de cerveza (Gallone et al., 2016; Molinet & Cubillos, 2020).

En México, los destilados de agave se obtienen a partir de la fermentación del jugo y el bagazo de los tallos y bases de las hojas cocidos de las plantas de agave (Agave spp., Asparagaceae). Esta categoría se refiere a todas las bebidas destiladas obtenidas del agave, como el tequila, el mezcal, el bacanora, la raicilla y otras bebidas destiladas con diversos nombres locales. La producción tradicional tiene lugar en destilerías pequeñas o medianas en todo el país, donde la fermentación abierta ocurre "espontáneamente" gracias a los microorganismos del entorno que habitan en una amplia variedad de climas. Se emplea una amplia variedad de técnicas de producción (Arellano-Plaza et al., 2022; CONABIO, 2006) y se utilizan más de cincuenta taxa diferentes de agave (Colunga-García-Marín et al., 2017). En resumen, la producción de destilados de agave implica la cosecha de la planta, la cocción, la molienda, la fermentación y la destilación (Kirchmayr et al., 2017; Nolasco-Cancino et al., 2018). La cocción transforma los carbohidratos complejos como los fructanos en azúcares simples, principalmente fructosa. Este paso también genera compuestos inhibidores para microorganismos como furanos, aldehídos y ácidos orgánicos (Mancilla-Margalli & López, 2002). A través de la molienda, se obtiene pulpa, fibras y jugo, y con la adición de agua, la fermentación abierta puede llevarse a cabo gracias a la actividad de levaduras presentes en el entorno y otros microorganismos (Kirchmayr et al., 2017; Lappe-Oliveras et al., 2008; Nolasco-Cancino et al., 2018; Páez-Lerma et al., 2013).

A pesar de la creciente demanda de aguardientes de agave en los mercados internacionales, la diversidad de levaduras responsables de la fermentación de los agaves sigue siendo en gran medida desconocida. Estudios pioneros realizados por Marc-André Lachance en una destilería tradicional de tequila mostraron que las plantas locales, los insectos y los objetos en la destilería pueden actuar como reservorios de levaduras en la fermentación del agave (Lachance, 1995). En esfuerzos posteriores, se han aislado levaduras de los tanques de

5

fermentación en algunas destilerías tradicionales en los estados de Oaxaca, Durango y Tamaulipas (Kirchmayr et al., 2017; Lappe-Oliveras et al., 2008; Nolasco-Cancino et al., 2018; Páez-Lerma et al., 2013). De estos estudios, se han identificado al menos 43 especies de levaduras y más de 50 especies de bacterias presentes en las fermentaciones de agave. Entre ellas, *S. cerevisiae, Kluyveromyces marxianus* y especies de *Pichia* son las más abundantes, independientemente de la ubicación muestreada o el año del estudio. Estos estudios se han centrado en ubicaciones específicas en algunas regiones productoras y, por lo tanto, no han abarcado la totalidad de la diversidad biogeográfica y cultural asociada a la producción de destilados de agave en México.

El consorcio denominado "Genomas de Levaduras Mx" tiene como objetivo proporcionar una descripción amplia y una comprensión más profunda de la diversidad genómica y funcional de las levaduras en México. Dada la relevancia de estos microorganismos en las fermentaciones de agave y la limitada información disponible sobre su historia natural en tales ambientes, estas comunidades son un buen punto de partida para investigar la diversidad de levaduras en esta región del mundo. Para este propósito, hemos ensamblado la colección de cultivos YMX-1.0, que incluye miles de cepas de levaduras aisladas de tanques de fermentación de agave en 68 destilerías tradicionales en todo el país. La identificación mediante espectrometría de masas MALDI-TOF, seguida de secuenciación de ITS y D1/D2 de un subconjunto de aislados, confirmó la presencia constante de *S. cerevisiae* y reveló la amplia presencia de levaduras no convencionales, incluidas especies previamente no asociadas al sustrato del agave y dos especies candidatas a nuevas especies.

La colección de cultivos YMX-1.0 proporciona un recurso único para iniciar estudios genómicos y poblacionales de levaduras de las fermentaciones de agave en el contexto de una región megadiversa. Como caso de estudio, generamos y analizamos secuencias de genoma preliminares de cuatro aislados de la levadura no convencional *Kazachstania humilis*, lo que demostró que son notablemente diferentes de cepas de otras regiones.

MATERIALES Y MÉTODOS

Las destilerías de agave y la información sobre los lugares de muestreo. Los lugares de muestreo se seleccionaron para incluir destilerías tradicionales representantes de cada una de las siete regiones productoras de destilados de agave en México, según Aguirre-Dugua et al. (2006), con el área del Oeste subdividida (Oeste I y Oeste II) debido a condiciones ambientales contrastantes y diferentes especies de agave empleadas. El trabajo de campo se llevó a cabo entre marzo de 2018 y diciembre de 2020. Se recolectaron muestras de 68 destilerías distribuidas en 13 estados de México (Colima, Durango, Estado de México, Guanajuato, Guerrero, Jalisco, Michoacán, Oaxaca, Puebla, San Luis Potosí, Sonora, Tamaulipas y Zacatecas) (Figura 1A). Solo se incluyeron en el muestreo el mezcal, la raicilla, el bacanora y otros destilados tradicionales de agave, dejando de lado el tequila, que se produce en su mayoría en destilerías industrializadas (véase **Dataset S1**). Los parámetros biogeográficos se obtuvieron mediante la toma de las coordenadas geográficas del sitio sobre la cartografía de ráster correspondiente. Los datos de elevación se obtuvieron a una resolución de escala de 1 km a través del conjunto de datos GTOPO30 del Servicio Geológico de los Estados Unidos (earthexplorer.usgs.gov; entidades GT30W100N40 y GT30W140N40).



Figura 1. La colección de cultivos YMX-1.0 se generó con levaduras aisladas de diversas condiciones ambientales y de producción. (A) Los sitios de muestreo fueron 68 destilerías tradicionales de agave en todo México (puntos negros); las áreas en color indican las siete regiones donde se producen bebidas destiladas de agave en el país. (B) Gráficos de dispersión que muestran la diversidad de parámetros climáticos de los sitios de muestreo agrupados por región: temperatura anual promedio (arriba a la izquierda), isoterma (arriba a la derecha), precipitación anual promedio (abajo a la derecha) y altitud sobre el nivel del mar (abajo a la izquierda). (C) Se recolectaron muestras de tanques de fermentación activa; las imágenes muestran ejemplos de recipientes, incluyendo (i)

tanques de cemento, (ii) barriles de madera, (iii) recipientes de piel de vaca, (iv) barriles de plástico, (v) ollas de barro y (vi) barriles de acero. (**D**) Diagrama de flujo que muestra la generación de la colección y los análisis iniciales realizados en este trabajo. Las muestras de agave fermentado se enriquecieron en medio de cultivo líquido y se sembraron en Agar Nutritivo WL. Las colonias aisladas se cultivaron y almacenaron en placas de microtitulación de 96 pocillos y se les asignó una especie putativa mediante la espectrometría de masas MALDI-TOF. La asignación de especies se validó mediante secuenciación de ITS para un subconjunto de aislados. La secuenciación de D1/D2 e ITS se utilizó para identificar posibles nuevas especies y la secuenciación del genoma para evaluar la diversidad intraespecífica (consulte Métodos).

Los datos ambientales se obtuvieron a través de las capas climáticas de WorldClim (worldclim.org/bioclim) con una resolución de 1 km (30 segundos de arco). Las variables analizadas incluyen la Precipitación Anual Media (wcbio_12), la Temperatura Anual Media (wcbio_01) y la Isotermalidad (wcbio_03); los climas se clasificaron como Aw (cálido subhúmedo), (A)C (semi-cálido), C(w) (templado) y BS (árido y semiárido) con base en (García, 1998). Para cada destilería, se registró información como coordenadas geográficas, tipo de horno, tipo de molienda y otras características estructurales del edificio. Los metadatos asociados con cada aislado se proporcionan en el **Dataset S1**.

Recolección y manejo de muestras de fermentación de agave. Las muestras se recogieron de destilerías que tenían al menos un tanque de fermentación activo en el momento de la visita. Cuando había más de un tanque activo disponible, se utilizaba la información proporcionada por los productores sobre la especie de agave y la fase de fermentación para seleccionar los tanques a muestrear, es decir, se muestreaba uno de cada agave o etapa.

La fase de fermentación para cada muestra se determinó dividiendo el tiempo de fermentación en curso en el momento del muestreo por la duración total estimada de la fermentación, según lo reportado por el productor. El tiempo total de fermentación se dividió en tres fases definidas como terciles. El número total de tanques muestreados fue de 201, variando de uno a cinco por destilería, con la excepción de diez tanques que se muestrearon solo una vez. El muestreo se realizó recogiendo mosto de agave a la distancia de un brazo desde el borde del tanque y a una profundidad de 25-30 cm desde la superficie, utilizando una pipeta serológica estéril de 25 mL. El pH se midió con bandas colorimétricas (Civeq

CVQ2051) y la temperatura del tanque se determinó con un termómetro de aguja a 10 a 15 cm de la superficie. El mosto de agave recogido se utilizó para llenar dos crioviales preetiquetados de 4 mL con glicerol a una concentración final del 12.5%, que se congelaron de inmediato en nitrógeno líquido. Todas las muestras se mantuvieron en nitrógeno líquido hasta que se transfirieron a un congelador a -80 °C en el laboratorio. Además, se recopilaron datos del sitio y de la muestra, que incluyen el material del tanque de fermentación, el nombre de la especie de agave en cada tanque, la temperatura de la muestra y el pH (consulte **Dataset S1**).

Aislamiento de levaduras y establecimiento de la colección de cultivos YMX-1.0. Las muestras se descongelaron y se utilizaron 500 μ L para inocular un medio de aislamiento de *Saccharomyces* (3 g/L de extracto de levadura, 3 g/L de extracto de malta, 5 g/L de bacto peptona, 10 g/L de glucosa, 1 ml/L de HCL, 200 mg/L de cloramfenicol), tal como se describe en (Liti et al., 2017). La concentración de etanol en el medio de aislamiento se redujo al 6% v/v, imitando lo que suele encontrarse en las fermentaciones de agave y también para promover el aislamiento de especies de levadura no convencionales con menor tolerancia al etanol. Un total de 158 de las 201 muestras (86.7%) de 147 tanques diferentes mostraron signos de fermentación después de dos hasta siete días; esas muestras se diluyeron 1:100 y 1:1000 y se sembraron en Agar Nutriente Laboratorio Wallerstein (WL) (Sigma 17222), lo que permite la detección de variaciones morfológicas y fisiológicas (Verdugo Valdez et al., 2011).

Las muestras se incubaron durante cinco días a 25 °C, seguidos de dos días a 4 °C para aumentar las diferencias morfológicas. Las colonias que mostraban morfologías diferentes se seleccionaron al azar de la placa que mostraba el mayor número de colonias aisladas; se transfirieron al menos 12 y hasta 60 colonias a 150 µL de medio YPD con 12.5% de glicerol en placas de 96 pozos para microtitulación (Corning CLS3367) y se almacenaron a -80°C. Para la mayoría de las muestras (144 de 158), se aislaron entre 24 y 36 cepas. Es importante mencionar que solo aquellas cepas utilizadas para la extracción de ADN y secuenciación de ITS o D1/D2 se volvieron a sembrar dos veces para obtener colonias aisladas. La colección

de cultivos YMX-1.0 contiene 4,524 cepas distribuidas en cincuenta y tres placas para microtitulación.

Identificación de cepas mediante espectrometría de masas MALDI-TOF. Todas las cepas se analizaron según lo descrito en (García-Gamboa et al., 2021). En resumen, las placas de microtitulación se descongelaron, los arreglos de cepas se imprimieron en placas de agar YPD y se incubaron durante 48 h a 30 °C. La biomasa se aplicó directamente sobre placas de acero pulido de destino MALDI MSP 96; se agregó 1 µL de ácido fórmico al 70% a cada punto hasta que estuvo completamente seco, luego se cubrieron los puntos con 1 μ L de una solución de matriz HCCA (ácido α-ciano-4-hidroxycinámico) al 10 mg/mL, 50% de acetonitrilo y 2.5% de ácido trifluoroacético como disolvente orgánico. Las placas se procesaron en un instrumento MICROFLEX LT (Bruker Daltonics) y se adquirieron huellas dactilares de biomasa para cada punto con el software flexControl 3.4 con el método MBT_FC.par. Se recopilaron cuarenta disparos láser en seis posiciones aleatorias dentro del punto, generando automáticamente espectros de masas en el rango de 2,000 a 20,000 Da. Los archivos de salida se compararon con las bases de datos BDAL v.8 o v.10 (consulte el **Dataset S1**). No se obtuvieron picos de masa para 462 (10.2%) muestras; las 4,524 muestras restantes, 3,063 (67.7%) superaron un Score Value estricto de 1.7 que se utilizó para la asignación de especies. Un Score Value más relajado de 1.5 resultó en la identificación de 3,503 (77.4%) de las cepas.

Pruebas estadísticas de asociaciones entre la composición de la comunidad y los metadatos de las muestras. Se realizó una prueba de Análisis de Similitudes no paramétrica (ANOSIM) (Clarke, 1993; biblioteca vegan en R, Oksanen et al., 2022) con muestras o destilerías como unidades de muestreo; las comunidades se codificaron a nivel taxonómico de especie o género. Las pruebas se realizaron con la disimilitud de presencia o ausencia entre las unidades de muestreo. Se utilizaron los índices de Jaccard como distancia y los factores de agrupación estaban relacionados con las condiciones de fermentación (**Tabla S1**). Las ubicaciones, es decir, las destilerías, se agruparon usando como factores la región, el grupo climático y el tipo de molienda (tres pruebas simultáneas, α =0.05/3). Para las muestras individuales como factores de agrupación, se consideraron otros siete factores adicionales, a

saber, localidad, fase de fermentación, material del tanque, duración total esperada de la fermentación, especie de agave, pH y temperatura del tanque (diez pruebas simultáneas, α =0.05/10). Para la fase de fermentación y el grupo climático, se identificaron los taxa que diferían entre los grupos mediante un análisis de patrones multinivel utilizando 9999 permutaciones para evaluar la significancia del valor de asociación (p<0.05, índice de correlación biserial puntual r.g.) de cada especie con respecto a cada grupo (biblioteca indicspecies en R; De Cáceres et al., 2010). Se evaluó si la diversidad de levaduras era similar en todas las fases de fermentación mediante un modelo de efectos mixtos, utilizando la diversidad de Simpson por muestra como variable dependiente, la fase de fermentación como efecto fijo, las localidades dentro de las regiones como efectos aleatorios anidados, y la temperatura y el pH del tanque como efectos aleatorios cruzados (biblioteca lme4 en R; Bates et al., 2015). Este modelo se comparó con un modelo nulo en el que se descartó el efecto de la fase de fermentación mediante una prueba de razón de verosimilitud.

Identificación de cepas mediante secuenciación de ITS y D1/D2. Se seleccionó un panel de 474 cepas, incluyendo cuatro de las especies putativas más abundantes identificadas por MALDI-TOF. Los aislados se volvieron a sembrar dos veces para obtener cultivos axénicos, a partir de los cuales se extrajo ADN genómico con hidróxido de sodio (Sylvester et al., 2015). La región ITS/5.8 rDNA ITS se amplificó utilizando cebadores fúngicos universales ITS-1 e ITS-4, como se describe anteriormente (Sylvester et al., 2015), y se secuenciaron utilizando los mismos cebadores. Las secuencias se filtraron y procesaron utilizando un umbral de recorte de 0.01, utilizando un algoritmo de corte Mott modificado (M1) implementado en el paquete *sangeranalyseR* para R (Chao et al., 2021).

Para un subconjunto de aislados que no proporcionaron una identificación confiable por MALDI-TOF, la identidad taxonómica se asignó analizando las secuencias tanto de la región ITS-5.8S mencionada anteriormente como de los dominios variables D1/D2 que se amplificaron y secuenciaron utilizando los cebadores NL1 y NL4. Las secuencias se ensamblaron, editaron y alinearon con MEGA11 (Tamura et al., 2021) y se compararon con las secuencias de levaduras anotadas en la base de datos de GenBank utilizando BLAST, como se describió anteriormente (Sayers et al., 2020). Se utilizó un enfoque de secuenciación

de baja profundidad para la asignación taxonómica de los aislados YMX003144, YMX000711 y YMX000142, que fueron asignados a dos posibles nuevas especies. En resumen, el ADN genómico se purificó utilizando el kit de purificación de ADN MasterPure según las recomendaciones del fabricante y se secuenció utilizando la plataforma Illumina NextSeq (2 x 150 pb). Las lecturas de baja calidad se eliminaron con *fastp* v0.20.0 (Chen et al., 2018). Las lecturas restantes se ensamblaron con *SPAdes* v3.12.0 (Bankevich et al., 2012). Las secuencias de nucleótidos de las regiones ITS y D1/D2 se recuperaron de los ensamblajes con el paquete *biostrings* de R (Pagès et al., 2023) utilizando las secuencias de los cebadores correspondientes.

Secuenciación y análisis de genomas. Se llevó a cabo la secuenciación del genoma de tres cepas de *Kz. humilis* utilizando las plataformas Illumina (2x300 pb) y BGI (2x150 pb). En total, se generaron 22.6 millones de lecturas de Illumina (un promedio de 7.54 millones) y 93.87 millones de lecturas de la plataforma BGI (un promedio de 31.29 millones). Las lecturas crudas se filtraron por calidad utilizando *fastp* v0.20.0 (Chen et al., 2018). Después del filtrado, se retuvieron 96.87 millones de lecturas de alta calidad (~83.16%; 18.81 millones de Illumina y 78.05 millones de BGI). Se realizaron ensamblajes de novo de las lecturas filtradas utilizando *SPAdes* v3.12.0 (Bankevich et al., 2012) con los parámetros por defecto.

Los ensamblajes iniciales se mejoraron posteriormente con RagTag v2.1.0 (Alonge et al., 2022), utilizando el ensamblaje del genoma YMX004033 (García-Ortega et al., 2022) como referencia. Los genomas ensamblados, incluyendo los genomas de referencia CLIB 1323 y YMX004033, fueron anotados utilizando el programa *MAKER* v2.31.8 (Cantarel et al., 2008). Las estadísticas del ensamblaje se calcularon utilizando *Quast* (Gurevich et al., 2013). Los valores del promedio de identidad nucleotídica (ANI) entre las cepas y la cepa tipo CBS 1323 se determinaron mediante *FastANI* (Jain et al., 2018) y se representaron gráficamente con los paquetes *pheatmap* (Kolde, 2012). Las secuencias nucleotídicas de los genes ITS, D1/D2 LSU rRNA, RPB1, RPB2 y EF-1 α fueron descargadas de GenBank. Si las secuencias no estaban disponibles, se obtuvieron a partir de secuencias de genomas completos utilizando el paquete *biostrings* de R (Pagès et al., 2023). Las alineaciones para el árbol filogenético se realizaron con *Clustal* (Larkin et al., 2007). Estas alineaciones se concatenaron y se

eliminaron las posiciones con brechas. La alineación resultante se utilizó para inferir un árbol de Máxima Verosimilitud con RaxML (Stamatakis, 2014). Se empleó el modelo GTR+GAMMA para la sustitución de nucleótidos y se realizaron 1,000 réplicas de remuestreo (bootstrap replicates). La variación estructural entre los genomas CLIB1323 y YMX004033 se identificó mediante comparaciones de los ensamblajes de genomas completos de ambos utilizando el software SyRI (Goel et al., 2019). Ambos ensamblajes se organizaron entre sí utilizando RagTag, que organizó los 16 contigs de CLIB 1323 en 13 andamios y los 21 contigs de YMX004033 en 15 andamios sin pérdida de información. Los genomas ensamblados se alinearon utilizando MUMmer para identificar pares de andamios homólogos (Delcher et al., 2003), lo que resultó en un conjunto de 13 andamios comparables para cada cepa. Este conjunto de andamios mantuvo el 100% de la secuencia original de CLIB 1323 y el 98.63% de la secuencia de YMX004033. Para identificar los genes codificados dentro de las regiones mal alineadas que excedían 1 Kb entre los dos genomas, superpusimos estas regiones con las anotaciones de genes utilizando la función findOverlaps del paquete GenomicRanges de R (Lawrence et al., 2013). Los ortólogos entre las cepas de *Kz. humilis* se determinaron utilizando OrthoFinder (Emms & Kelly, 2019).

RESULTADOS

Una nueva colección exhaustiva de cepas de levaduras aisladas de fermentaciones de agave.

La colección de cultivos YMX-1.0 consta de 4,524 cepas de levaduras aisladas de 68 destilerías artesanales de agave en México. Cada aislado está asociado a metadatos sobre la identificación a nivel de especie, información biogeográfica, prácticas de producción, el pH y la temperatura del tanque al momento del muestreo (consulte **Dataset S1**). El trabajo de campo se llevó a cabo entre 2018 y 2020, y se proporciona una descripción detallada en Materiales y Métodos.

Los sitios de muestreo fueron destilerías tradicionales ubicadas en trece estados en las siete regiones de México donde se producen destilados de agave: Noroeste, Noreste, Oeste I, Oeste

II, Cuenca del Balsas, Altiplano Central y Centro-sur (**Figura 1A**). Se excluyeron las destilerías de tequila porque en la actualidad se utilizan monocultivos axénicos comúnmente, para la fermentación controlada en estos entornos industrializados. La distancia máxima entre las destilerías muestreadas fue de 2,033 km (1,263 millas) desde Baviácora, Sonora, hasta San Luis Amatlán, Oaxaca. Los sitios de muestreo representaron al menos cuatro tipos generales de clima: cálido subhúmedo, semi-cálido, semiárido y templado. La temperatura ambiental anual en los sitios de muestreo variaba entre 14 y 24 °C, con una isotermalidad que oscilaba entre 49 y 78. La precipitación media anual se registra desde los 389 L/m2 en San Luis de la Paz, Guanajuato (Altiplano Central) hasta 1,510 L/m2 en Cabo Corrientes, Jalisco (Oeste II). Las altitudes desde 395 metros sobre el nivel del mar (msnm) en San Pedro Ures, Sonora (Noroeste) hasta 2,508 msnm en San Felipe, Guanajuato (Altiplano Central) (**Figura 1B**).

La diversidad de prácticas de producción se reflejaba en los tipos de tanques de fermentación, que estaban construidos de madera, cemento, plástico, piedra, arcilla, acero o cuero de vaca (**Figura 1C**). Las muestras de fermentación de agave se congelaban con nitrógeno y glicerol en el campo y se utilizaban en el laboratorio para aislar cepas de levaduras para su posterior análisis (**Figura 1D**; consulte los Métodos).

Identificación y distribución de las levaduras aisladas de fermentaciones de agave.

Para proporcionar una descripción taxonómica completa de la colección YMX-1.0, buscamos caracterizar todas las cepas aisladas mediante espectrometría de masas MALDI-TOF (consulte **Dataset S1**). Un número minoritario de muestras (465; 10.3%) no generó picos de espectrometría de masas y 996 (22.0%) muestras proporcionaron espectros sin una identificación confiable en la base de datos (valor de corte del Puntaje <1.7). Es importante destacar que 3,063 cepas (67.7%) fueron clasificadas utilizando este valor de corte estricto, correspondiendo a 18 especies diferentes de levaduras. La colección también incluyó una pequeña fracción de aislados bacterianos (n=39). Al utilizar un valor de corte más permisivo de 1.5, el número de aislados clasificados aumentó a 3,503 (77.4%), representando 23 especies de levaduras. Del conjunto clasificado de manera estricta, las especies putativas más abundantes fueron *S. cerevisiae* (55.4%), *Pichia kudriavzevii* (14.2%), *Kluyveromyces*

marxianus (8.6%), Pichia manshurica (6.6%), Pichia kluyveri (3.7%) y Torulaspora delbrueckii (2.7%) (Figura 2A; Tabla 1). Además, al menos el 1% de las cepas en la colección se asignaron a Hanseniaspora lachancei, Pichia membranifaciens y Hanseniaspora valbyensis.



Figure 2

Figura 2. La distribución de las levaduras en México revela un núcleo de seis especies y variación local de especies poco comunes. Utilizando la espectrometría de masas MALDI-TOF, se asignaron 3,063 aislados de la colección de cultivos YMX-1.0 a especies putativas. (A) El gráfico de barras muestra la frecuencia de las especies del núcleo; otras especies comprendían el 8.8% de las cepas identificadas. (B) Las barras apiladas muestran la distribución de las especies de levaduras aisladas en cada una de las siete regiones productoras; el número de aislados de cada región se muestra entre paréntesis. Solo se muestran por separado las seis especies del núcleo y el resto de las especies se agrupan como otras (Otrs). Se indican los índices de diversidad de Shannon-Weiner (H') y Simpson (D') para cada región (arriba). (C) Las barras apiladas muestran las frecuencias de las regiones productoras donde se aislaron las cinco especies

prevalentes, pero más raras. Estas especies se aislaron con frecuencias del 0.5-1.0%, pero estaban presentes en al menos el 5% de los sitios de muestreo. El número de aislados de cada especie se muestra entre paréntesis y se muestra información de *S. cerevisiae* como referencia. (*D*) El gráfico muestra las frecuencias de especies que están significativamente asociadas con diferentes fases de fermentación (análisis de patrón multinivel, p<0.05); todas las demás levaduras se agrupan y se representan en una sola serie de datos (línea gris). Las tres fases de fermentación se definieron por el tercil del tiempo en el que se tomó cada muestra en relación con el tiempo total de fermentación.

Con base en los datos de clasificación de MALDI-TOF, a pesar de la gran diversidad de condiciones ambientales, observamos que seis especies - S. cerevisiae, P. kudriavzevii, K. marxianus, P. manshurica, P. kluyveri y T. delbrueckii — se encontraron en el 20-88% de los sitios de muestreo y estuvieron presentes en al menos cinco de las siete regiones productoras. Estas levaduras podrían considerarse las especies centrales de la fermentación de agave. Si solo se tienen en cuenta las especies presentes en todas las regiones productoras, S. cerevisiae, P. kudriavzevii y K. marxianus constituirían el núcleo de estas comunidades de levaduras. Las otras doce especies se encontraron en menos de cinco regiones productoras y juntas representaron el 8.8% de los 3,024 aislados de levaduras con asignaciones de especies en nuestra colección (Figura 2A). Calculamos los índices de Shannon-Weiner (H') y Simpson (D') para cada región, lo que mostró que la diversidad de especies de levaduras variaba desde H'=1.15 en el Noroeste hasta H'=2.11 en la región del Centro-sur (Figura 2B). La dominancia de las especies de levaduras más abundantes fue marcada en el Noroeste (D=0.51) y fue menos evidente en otras regiones. Es importante destacar que los aislados de otras levaduras ampliamente distribuidas como P. kudriavzevii fueron poco comunes en el Altiplano Central. P. manshurica fue rara en el Noroeste. En contraste, los aislados de T. delbrueckii y H. lachanceii eran comunes en la región del Centro-sur, aunque T. delbrueckii estaba presente en seis de las siete regiones productoras, mientras que H. lachanceii solo estaba presente en las regiones del Centro-sur y Oeste I. Otras especies se aislaron con menor frecuencia (0.5-1%), pero estaban distribuidas en diferentes sitios de muestreo y regiones (Figura 2C; Tabla 1).

Especie	Número de cepas aisladas	Frecuencia en destilerías (n=68)	Frecuencia en regiones productoras (n=7)
Saccharomyces cerevisiae*	1,676	62	7
Pichia kudriavzevii *	430	42	7
Kluyveromyces marxianus*	259	31	7
Pichia manshurica *	199	34	6
Pichia kluyveri	111	22	5
Torulaspora delbrueckii	83	14	6
Hanseniaspora lachancei	73	8	2
Pichia membranifaciens	56	8	2
Hanseniaspora valbyensis	38	2	2
Kazachstania humilis	24	7	4
Zygosaccharomyces balii	23	5	4
Hanseniaspora opuntiae	20	6	4
Wickerhamomyces anomalus	18	6	4
[Candida] boidinii / Ogataea clade	5	2	2
Pichia fermentans	4	3	3
Pichia occidentalis	2	2	1
Wickerhamomyces subpelliculosus	2	2	1
Saccharomyces paradoxus	1	1	1

Tabla 1. Levaduras Saccharomycetales de fermentaciones de agave de la colección de cultivos YMX-1.0 identificadas por biotipificación MALDI-TOF¹

¹ Se obtuvieron datos MALDI-TOF para 4,062 (89.8%) muestras; la identidad de las especies se asignó para 3,063 aislados (67.7%) usando un valor de corte estricto (Score Value>1.7)

^{*} Una fracción menor de los aislados fue asignada a una especie distinta cuando se analizaron las secuencias de ITS (Tabla S1).

Aunque un conjunto central de especies fue común en todas las fermentaciones de agave, según ANOSIM, la composición de las comunidades de levaduras en las muestras de fermentación tuvo diferencias estadísticamente significativas entre localidades, grupos climáticos, fases de fermentación y duración total de la fermentación (**Tabla 2**). Mientras tanto, las regiones productoras, los procedimientos de molienda, las especies de agave, el pH

del tanque y la temperatura del tanque no mostraron diferencias estadísticamente significativas. Es notable que en todos los casos las magnitudes de los efectos fueron pequeñas (r<0.25), lo que indica que la variación en la composición de las comunidades asociada a estas variables es pequeña. Estos resultados fueron en su mayoría insensibles al nivel taxonómico utilizado para los análisis (especies o géneros); solo el material del tanque mostró una asociación a nivel de género que no se observó al considerar las especies (Tabla 2). Por el contrario, al analizar las comunidades a nivel de las destilerías, no encontramos diferencias estadísticamente significativas en la composición de las comunidades de levaduras en ninguno de los grupos probados (regiones productoras, grupos climáticos o tipo de molienda utilizado, **Tabla 2**). Los valores pequeños del estadístico r de esta prueba a nivel de las muestras junto con las diferencias no estadísticamente significativas a nivel de las destilerías podrían reflejar la limitación de nuestro diseño de muestreo para desentrañar completamente las variables que influyen en la composición de las comunidades de levaduras en estas fermentaciones. Sin embargo, en general, las similitudes de las clasificaciones dentro de las fases de fermentación y las localidades sugieren que los factores locales y de corto plazo parecen ser los determinantes más importantes de la composición de las comunidades de levaduras en la fermentación de agave.

	Por muestra *		Por ubicación (destilería)*	
	Especies	Géneros	Especies	Géneros
Región productora	<i>r</i> =0.07 N.S.	<i>r</i> =0.09 N.S.	<i>r</i> =0.08 N.S.	<i>r</i> =0.04 N.S.
Clima	r=0.06 p<0.001	r=0.06 p<0.005	<i>r</i> =0.06 N.S.	<i>r</i> =0.05 N.S.
Tipo de molienda	<i>r</i> =0.06 N.S.	<i>r</i> =0.05 N.S.	<i>r</i> =0.08 N.S.	<i>r</i> =0.08 N.S.
Localidad	r=0.23 p<0.001	r=0.2 p<0.001		
Fase de fermentación	r=0.055 p<0.001	r=0.2 p<0.001		
Material del tanque	<i>r</i> =0.07 N.S.	r=0.11 p<0.005		
Duración total de la fermentación	r=0.12 p<0.001	r=0.14 p<0.001		
Especie de agave	<i>r</i> =0.04 N.S.	<i>r</i> =-0.06 N.S.		
рН	<i>r</i> =0.01 N.S.	<i>r</i> =0.01 N.S.		
Temperatura del tanque	<i>r</i> =0.001 N.S.	<i>r</i> =0.01 N.S.		

Tabla 2. Análisis de similitudes de la composición de las comunidades a partir de los metadatos de las muestras y las localidades ¹

¹ Estadístico r de la prueba de ANOSIM (*p-value* calculado con 9999 permutaciones, librería *vegan* en R) usando la distancia de Jaccard en una matriz de presencia/ausencia.

* Valores significativos aplicando la corracción de Bonferroni α =0.05/10 =0.005 para muestras y α =0.05/3 =0.0167 para destilerías. N.S., no significativo.

Para evaluar más a fondo los cambios en la composición de las comunidades de levaduras a lo largo del tiempo, primero estimamos el índice de diversidad de Simpson en las diferentes fases de fermentación. El índice disminuyó con el tiempo de fermentación, pero el cambio no fue estadísticamente significativo después de controlar la variación geográfica (localidades dentro de las regiones) y las condiciones del tanque (pH y temperatura)

(p=0.059, prueba de razón de verosimilitud). Sin embargo, la disimilitud de Bray-Curtis sí mostró una disminución estadísticamente significativa en función de la fase de fermentación, aunque la magnitud del cambio fue pequeña (r=0.06, p<0.001, ANOSIM). El análisis de patrones multinivel indicó que las primeras fases de fermentación estaban asociadas con una abundancia relativamente alta de *K. marxianus* (r.g.=0.276, p=0.002), *P. kluyveri* (r.g.=0.217, p=0.0162), y *P. membranifaciens* (r.g.=0.197, p=0.037), mientras que las fases tardías estaban enriquecidas en *Z. bailii* (fase III, r.g.=0.163, p=0.013) y *S. cerevisiae* (fases II y III, r.g.=0.317, p=0.0185) (**Figura 2D**). En conclusión, la diversidad no disminuyó significativamente a medida que avanzaba la fermentación, pero algunas especies sí cambiaron su abundancia con el tiempo. En general, estos resultados sugieren que hay un núcleo común de especies de levaduras en las fermentaciones de agave y que las variables locales dan forma a la composición de las levaduras en las fermentaciones muestreadas.

Validación de la tipificación MALDI-TOF mediante secuenciación de ITS

Para verificar la clasificación de especies basada en la tipificación MALDI-TOF, secuenciamos las regiones espaciadoras internas transcritas del ADN ribosomal (región ITS) de un subconjunto de 474 cepas (**Tabla S1**). Las secuencias ITS revelaron que 390 (82.3%) de los aislados fueron identificados correctamente al inicio mediante la tipificación MALDI-TOF a nivel de especie (**Figura 3**). Las discrepancias entre las identificaciones mediante estos enfoques probablemente se deban a una combinación de factores; la precisión de la identificación MALDI-TOF depende de la especie analizada, el procesamiento de la muestra y la completitud de la base de datos de referencia. La frecuencia más alta de identificaciones precisas se obtuvo para las cepas clasificadas como *S. cerevisiae* (88.1%) y *Kl. marxianus* (87.6%), mientras que la precisión más baja fue para *P. manshurica* (58.3%); 25 de las 72 cepas putativas de *P. manshurica* (~35%) correspondían a otras especies de *Pichia*. Es importante señalar que una pequeña fracción (7.6%) de los aislamientos putativos de *S. cerevisiae* fueron asignados a su especie hermana, *Saccharomyces paradoxus*, mediante secuenciación de ITS, lo que aumentó el número total de aislamientos de *S. paradoxus* a diez en la colección YMX-1.0.





Figura 3. Validación de la tipificación por MALDI-TOF mediante secuenciación del ITS. Un subconjunto de 474 aislados asignados a las cuatro especies más abundantes fue confirmado mediante secuenciación del ITS; el gráfico muestra la frecuencia de clasificaciones correctas (gris oscuro) e incorrectas (gris claro). Las especies correctas basadas en la secuenciación del ITS se proporcionan en la Tabla S1.

Candidatos a nuevas especies de levaduras en el ambiente de fermentación del agave.

Muchos de los aislados de YMX-1.0 con espectros de alta calidad (996, 22.0%) no fueron asignados a ninguna especie en la base de datos usando un umbral de confianza alto (ScoreValue>=1.7). Este número seguía siendo considerable cuando se usaba un umbral más laxo de >=1.5, lo que resultaba en 439 aislamientos con identificaciones poco confiables. Para explorar la diversidad de levaduras en este conjunto de cepas no identificadas, secuenciamos los dominios D1/D2 y la región ITS de 22 aislamientos sin clasificación MALDI-TOF de Oaxaca (región Centro-sur) y Durango (región Oeste I). Estos sitios distantes fueron elegidos porque fueron algunos de los más muestreados y son considerablemente diferentes en cuanto a clima y prácticas de producción. El análisis de las secuencias mostró que las cepas aisladas pertenecían a tres géneros, a saber, *Saccharomyces*, *Torulaspora y Pichia*. Se identificaron seis especies, cuatro de las cuales eran conocidas previamente, mientras que dos eran candidatas a nuevas especies (**Tabla 3**).

Especie	IDs de las cepas ^a	Región secuenciad a	Especie relacionada (accession GenBank)	Identidad ^b
<i>Pichia</i> sp. 1 (candidato a nueva especie)	YMX003144* YMX003126 YMX000711* YMX000073 YMX003130 YMX000413 YMX000140 YMX000144 YMX000082 YMX000716 YMX000709	ITS	<i>Candida ethanolica / Pichia</i> clade (LT854882)	376/403 (93%)
		D1/D2	<i>Candida ethanolica / Pichia</i> clade (NG_055105)	554/559 (99%)
<i>Pichia</i> sp. 2 (candidato a nueva especie)	YMX000142*	ITS	Pichia fermentans (MK394169)	378/447 (85%)
		D1/D2	Pichia fermentans (KY108804)	532/564 (94%)
Pichia manshurica	YMX001313	D1/D2	Pichia manshurica (MK394164)	400/400 (100%)
Pichia kudriavzevii	YMX000083	ITS	Pichia kudriavzevii (MH545928)	333/334 (99%)
		D1/D2	Pichia kudriavzevii (MH545928)	548/548 (100%)
Torulaspora delbrueckii	YMX001445 YMX001809 YMX001801 YMX001802 YMX001816	ITS	Torulaspora delbrueckii (MK394138)	601/601 (100%)
		D1/D2	Torulaspora delbrueckii (HE616749)	586/586 (100%)
Saccharomyces paradoxus	YMX002318 YMX002310 YMX003022	ITS	Saccharomyces cariocanus (KY104990)	571/571 (100%)
		D1/D2	Saccharomyces cariocanus (KY109235)	465/465 (100%)

Tabla 3. Identificación de 22 aislados Saccharomycetales de fermentación de agave basada en secuencias de las regiones ITS y dominio D1/D2 de la subunidad grande del gen rRNA

^a Todas las cepas fueron reestriadas para obtener colonias individuales antes de la secuenciación de rDNA y los nuevos IDs fueron asignados a las cepas reestriadas. (consulte Dataset S1).

^b Las secuencias para búsqueda de las 22 cepas tenían 100% de cobertura.

* Las secuencias para búsqueda fueron obtenidas de secuencias genómicas de baja profundidad (ver Materiales y Métodos).

Tres cepas de este conjunto de 22 aislados que no tenían asignación de especie por MALDI-TOF, mostraron una alta identidad de ITS con una cepa aislada de la mosca de la fruta anteriormente clasificada como S. cariocanus; esta especie ha sido reclasificada como S. paradoxus debido a la baja divergencia de secuencia entre las cepas (Liti et al., 2006). Además, cinco cepas de la región Centro-sur fueron identificadas como T. delbrueckii, una especie previamente reportada en fermentaciones abiertas de agave (Muñoz-Miranda et al., 2022). En cuanto al clado de *Pichia*, un aislado fue identificado como *P. kudriavzevii*, y otro como P. manshurica (Tabla 3). Las secuencias de consulta de once aislados en las dos regiones de producción mostraron ser idénticas y representaron un nuevo candidato a especie nueva, Pichia sp. 1, estrechamente relacionado con el clado [Candida] ethanolica / Pichia. El candidato a nueva especie Pichia sp. 1 (YMX003144 y otros) difirió por 27 sustituciones en sus secuencias ITS y cinco sustituciones en los dominios D1/D2 de C. ethanolica. Otro aislado de este género representa otro candidato a nueva especie, Pichia sp. 2 YMX000142. El pariente más cercano a esta otra nueva especie candidata fue P. fermentans, que mostró 32 sustituciones en los dominios D1/D2 y 69 sustituciones en la región ITS en comparación con el nuevo aislado. Estas dos nuevas especies candidatas serán descritas formalmente en trabajos futuros. Estos resultados sugieren que el ambiente de fermentación del agave, muestreado de forma subconjunta, alberga especies fúngicas novedosas que aún no se han descrito formalmente.

Los análisis genómicos revelan una diversidad genética intraespecífica considerable en *Kazachstania humilis*.

La colección de cultivos YMX-1.0 proporciona un recurso biológico único para llevar a cabo estudios genómicos y poblacionales sobre las diversas especies de levaduras encontradas en las fermentaciones de agave. En nuestro estudio, la levadura *Kz. humilis* (sin. *Candida humilis, Candida milleri*) se aisló con baja frecuencia, con 24 aislamientos en la colección completa, pero estaba presente en siete destilerías (10%), abarcando cuatro de las siete regiones productoras (**Tabla 1**). Para comprender mejor la diversidad genómica de *Kz. humilis* en las fermentaciones de agave, secuenciamos y analizamos el genoma completo de cuatro cepas de regiones distantes. Este conjunto incluyó la cepa YMX004033 previamente

secuenciada (García-Ortega et al., 2022). Los ensambles resultantes se compararon con la secuencia del genoma de la cepa tipo CLIB 1323 aislada de una masa madre en Francia (Jacques et al., 2016).

Los tamaños de genomas provisionales de los tres nuevos aislados de *Kz. humilis* secuenciados en este estudio (YMX00387, YMX000554 y YMX003162) variaron de 17.1 a 19 Mbp, que son tamaños más grandes que el de las referencias del genoma YMX004033 y CLIB 1323 (13.9 Mbp) (**Tabla S2**). Las diferencias en los tamaños estimados de los genomas pueden surgir de ensambles de lecturas cortas de genomas diploides heterocigotos. En tales casos, los haplotipos heterocigotos podrían estar presentes como dos contigs separados en lugar de estar fusionados en el ensamble haploide. Además, los ensambles de lecturas cortas tienden a tener más scaffolds, en este caso un promedio de 229 scaffolds, en comparación con los obtenidos con lecturas largas como YMX004033 (21 scaffolds) y CLIB 1323 (16 scaffolds).

Estas discrepancias derivadas de diferentes tecnologías de secuenciación han sido documentadas anteriormente (Korlach et al., 2017). El número promedio de genes predichos para las cuatro cepas de fermentación de agave fue 5,333, lo cual es más alto que los 4,295 predichos para la cepa tipo. Sin embargo, este valor es acorde con el número de genes reportados para otras especies dentro del género, incluyendo *Kz. barnettii* (5,276), *Kz. exigua* (5,416), *Kz. unispora* (5,287), *Kz. saulgeensis* (5,329), *Kz. africana* (5,375) y *Kz. naganishii* (5,319). Aún es posible que estas cifras sobreestimen el contenido real de genes debido a la generación de duplicados que resultan de ensamblar genomas haploides a partir de genomas no haploides, especialmente en regiones de mayor heterocigosidad (Ko et al., 2022).

Para delinear la posición filogenética de las cuatro cepas de *Kz. humilis* de la fermentación de agave, recuperamos las secuencias nucleotídicas de cinco marcadores genéticos frecuentemente utilizados (ITS, D1/D2 LSU rRNA, RPB1, RPB2 y EF-1 α) (Schoch et al., 2012) de los ensambles de genomas. Después de concatenarlos, se construyó un árbol filogenético basado en la alineación con secuencias de marcadores moleculares disponibles públicamente de especies de *Kazachstania* (**Figura 4A**). El árbol filogenético reconstruido

coincidió con las filogenias reportadas previamente para estas especies (Jacques et al., 2016). Como se esperaba, las cepas de las fermentaciones de agave se agruparon dentro de un clado único. Notablemente, la cepa tipo *Kz. humilis* CLIB 1323 junto con otras dos cepas, CBS 5658 (aislada de la cerveza Bantu en Sudáfrica) y NRRL Y-7245 (de masa madre en Estados Unidos), formaron un clado vecino.



Figura 4. Genómica comparativa de los aislados de *Kazachstania humilis.* (*A*) Árbol filogenético de máxima verosimilitud para las especies de *Kazachstania* basado en la alineación de los genes concatenados ITS, D1/D2 LSU rRNA, RPB1, RPB2 y EF-1a (2,863 posiciones en total); *Zygosaccharomyces rouxii* CBS 732 se utilizó como grupo externo. Los cuatro aislados de fermentación de agave se muestran en rojo. Los demás aislados de *K. humilis* se obtuvieron de diversas fuentes; CLIB 1323 y NRLL Y-7245 son de masa madre en Francia y Estados Unidos, respectivamente, mientras que CBS 5658 es

de cerveza Bantu en Sudáfrica. Los números en las ramas son el soporte bootstrap basado en 1,000 réplicas; solo se muestran los números superiores al 50%. La escala de distancia genética representa las sustituciones por nucleótido. La matriz (abajo a la derecha) muestra los valores de identidad promedio de nucleótidos (ANI) en todo el genoma para las cepas de *Kz. humilis* con secuencias de genoma disponibles. (**B**) Comparación entre los genomas de la cepa tipo *Kz. humilis* CLIB 1323 (líneas negras) y la cepa YMX004033 (líneas rojas). La figura muestra regiones sinténicas (gris), duplicadas (cian), translocadas (verde) e invertidas (naranja) mayores de 10 Kb. La densidad de SNP de YMX04033 en comparación con CLIB 1323 se muestra en barras grises. Las regiones duplicadas que resultan en ganancias de copias de genes se indican con triángulos magenta, mientras que las regiones duplicadas con elementos repetitivos de baja complejidad se muestran como cuadrados morados. Los nombres de los genes duplicados se muestran al lado del genoma donde se observa la duplicación.

Para cuantificar la relación genética entre las cepas de Kz. humilis de agave y la cepa tipo a nivel del genoma, determinamos la identidad promedio de nucleótidos (ANI) entre todas las cepas con secuencias de genomas disponibles. Los valores de ANI oscilaron entre el 96.5% y el 99.6% (Figura 4A, inset), los cuales se encontraron dentro del rango de diversidad esperado para cepas de levadura de la misma especie (Gostinčar, 2020). El valor de ANI más alto se observó entre YMX000554 y YMX003162 (99.56%), ambas son cepas de levadura de agave de la misma región productora. La cepa tipo CLIB 1323 tuvo el valor de ANI mediano más bajo al ser comparada con todas las cepas de las fermentaciones de agave (96.68%), un valor cercano al esperado entre dos especies diferentes. Aunque se necesita analizar más aislados de la especie, estas diferencias sugieren la existencia de una población distinta de Kz. humilis en el agave. Para descubrir variaciones estructurales dentro de los genomas de estas cepas, realizamos un análisis de sintenia entre las dos secuencias de genoma ensambladas a partir de datos de lecturas largas: la cepa tipo de Kz. humilis CLIB 1323 (Jacques et al., 2016) de masa madre en Francia y el aislado de fermentación de agave YMX004033 de la región productora Oeste II en México (García-Ortega et al., 2022) (Tabla S3). El análisis de sintenia a lo largo de 13 scaffolds consistentes reveló aproximadamente 11.8 Mb de regiones de sintenia (~85.9% del genoma CLIB 1323), 66 translocaciones, 137 duplicaciones en CLIB 1323 y 71 duplicaciones en YMX004033, 57,063 INDELs y 50,3408 SNVs. Además, SyRI identificó 0.24 Mb de secuencias de ADN de CLIB 1323 y 0.69 Mb de YMX004033 que no se alinearon con el otro genoma respectivo (Figura 4B).

Estas regiones no alineadas estaban asociadas con 159 genes en YMX004033 y once genes en CLIB1323. De estos, 20 genes de YMX004033 y un gen no caracterizado de CLIB 1323 no tenían homólogos en sus respectivas secuencias. Además, cuatro de los 21 genes presentes en dos contigs de YMX004033, que no se consideraron en el análisis de sintenia (ver Métodos), también carecían de ortólogos en el genoma CLIB1323. Los genes únicos para la cepa de agave YMX004033 incluían homólogos de AZR1, CDD1, egt-2, EHT1, FLO9, FLO10, GPM3, MOG1, OPT1, PIN2, PHO3, PDX3, SOA1, SUC4, STB3, Tf2-6, THI5, ygeX, YCT1, YJR142W, YKL068W-A y tres ORFs dudosos.

Estas variaciones estructurales importantes también demostraron que las cepas de *Kz. humilis* de agave y masa madre están claramente diferenciadas a nivel genómico, con una posible compatibilidad genética reducida entre ellas. La variación estructural también incluyó duplicaciones segmentarias y dispersas, algunas de las cuales abarcaban marcos de lectura abierta completos. En conjunto, estos hallazgos indicaron que las cepas aisladas de fermentaciones de agave son genéticamente distintas, destacando el potencial de la colección de cultivos YMX-1.0 para avanzar en la investigación sobre la genómica de poblaciones de levaduras no convencionales, fundamentales para la producción de alimentos y bebidas en todo el mundo.

DISCUSIÓN

La diversidad de levaduras asociadas con la fermentación del agave ha permanecido poco explorada, a pesar del amplio interés biológico e industrial en las levaduras Saccharomycetales y de una creciente demanda de los licores de agave en los mercados globales. En este estudio, hemos identificado las especies de levaduras presentes en este entorno relacionado con los humanos mediante el aislamiento sistemático de miles de cepas de las destilerías tradicionales en todo México. Cada aislado de la colección de cultivos YMX-1.0 está vinculado a metadatos de la muestra de fermentación y al sitio de aislamiento, incluyendo información biogeográfica, pH y temperatura en el momento de la recolección, y prácticas de producción.

Nuestro trabajo amplía considerablemente estudios previos de comunidades de levaduras realizados en ubicaciones específicas en algunas de las regiones productoras de agave (Kirchmayr et al., 2017; Lachance, 1995; Nolasco-Cancino et al., 2018; Páez-Lerma et al., 2013), proporcionando una visión extensa y comparativa de la diversidad de levaduras que habitan en este entorno. Es importante reconocer que la descripción presentada aquí se basa únicamente en cepas cultivadas y aisladas. Además, los resultados probablemente estén sesgados por las condiciones específicas de enriquecimiento y el enfoque MALDI-TOF utilizado para la asignación de especies. Debido a estas limitaciones, elegimos probar la asociación de la composición de la comunidad solo con un puñado de variables de metadatos utilizando la presencia y ausencia de las especies. Este es un enfoque conservador, ya que no se tienen en cuenta las diferencias en la abundancia relativa de especies dentro de las muestras al probar las asociaciones entre variables de metadatos y una comunidad dada. Además, no incluimos metadatos para los cuales el muestreo fue considerablemente desigual. Este es el caso de las localidades (municipios), ya que solo aproximadamente la mitad de nuestras muestras provienen de localidades con más de una destilería muestreada. A pesar de las limitaciones y el enfoque conservador, la naturaleza sistemática y la extensión de nuestro esfuerzo de muestreo ofrecen una oportunidad sin precedentes para contrastar la diversidad de especies de levaduras aisladas de fermentaciones de agave en diversos contextos biogeográficos y culturales. Esto permitió sugerir que factores como la destilería o el tiempo de fermentación tienen un efecto moderado, pero estadísticamente significativo en la composición de las comunidades de levaduras. Estudios futuros que utilicen enfoques de metagenómica proporcionarán una descripción más completa de las levaduras en estas fermentaciones y pueden revelar características únicas de las comunidades microbianas de regiones productoras específicas. Además, el recurso de la colección de cultivos y nuestros hallazgos iniciales probablemente impulsarán estudios impulsados por hipótesis centrados, por ejemplo, en comprender las características fisicoquímicas específicas que resultan en las diferencias locales de las comunidades de levaduras.

La colección de cultivos YMX1.0 también abre nuevas vías de investigación para comprender aspectos cruciales de la biología de las levaduras, como la diversidad genómica y funcional dentro de las poblaciones, lo que no sería posible utilizando enfoques que no se

basan en organismos cultivados. Por ejemplo, la duplicación del gen ZWF1 encontrada en un aislado de *Kz. humilis* secuenciado en este estudio y ausente en cepas de otras fuentes podría estar asociada con la tolerancia a furanos. Se ha demostrado que la sobreexpresión de genes de la vía de la pentosa-fosfato, como ZWF1, subyace a la tolerancia al hidroximetilfurfural en *S. cerevisiae* (Gencturk & Ulgen, 2022; Gorsich et al., 2006; Park et al., 2011). Del mismo modo, el conjunto de genes que están presentes o ausentes específicamente en los aislados de fermentación de agave podría ser la base genética candidata de la adaptación al nicho. La disponibilidad de una gran colección de cepas aisladas de fermentaciones de agave permitirá ahora poner a prueba experimentalmente este tipo de hipótesis.

En este estudio, las cepas de levadura fueron clasificadas taxonómicamente principalmente mediante la biotipificación MALDI-TOF utilizando la transferencia directa de biomasa (García-Gamboa et al., 2021), lo que permite una identificación rápida directamente desde colonias individuales. Este enfoque resulta en tasas más altas de asignación incorrecta de especies en comparación con los métodos basados en secuenciación. Evaluamos el alcance de este problema al secuenciar la región ITS de un subconjunto de 474 aislados y mostramos que se asignaron especies dispares en aproximadamente el 19% de los casos. Es importante destacar que García-Gamboa et al. (2021) también utilizaron extractos de levadura para la clasificación MALDI-TOF, logrando una identificación correcta del 100% de los aislados de P. kluyveri en comparación con la secuenciación ITS. En nuestro estudio, la clasificación incorrecta estuvo fuertemente sesgada hacia una de las especies probadas (P. manshurica), lo que subraya que las bases de datos de referencia utilizadas para la biotipificación por espectrometría de masas aún deben mejorarse para especies menos estudiadas. Esto es especialmente relevante para la caracterización de las fermentaciones de agave, en las que se ha demostrado que las especies de *Pichia* son comunes. La secuenciación de ITS y D1/D2 también reveló la presencia de S. paradoxus en las fermentaciones de agave. A diferencia de su especie hermana S. cerevisiae, esta especie es relativamente menos común en entornos fermentativos (Dashko et al., 2016).

Es importante destacar que nuestro estudio reveló dos nuevos candidatos a especies relacionados con el clado de *Pichia*, uno de los cuales parecía estar ampliamente presente en dos de las regiones productoras analizadas. Será necesario realizar más investigaciones para caracterizar estos aislados. Estos resultados sugieren que la colección de cultivos YMX-1.0 puede albergar nuevas especies de hongos asociadas con este ambiente fermentativo poco estudiado.

La colección de cultivos YMX-1.0 ofrece un recurso poderoso para la genómica de poblaciones de levaduras del orden Saccharomycetales. Analizar las numerosas cepas de especies modelo como *S. cerevisiae* o *Kl. marxianus* ayudará a entender su evolución e historia natural en la interfaz de ambientes naturales y relacionados con los humanos. De hecho, el exhaustivo estudio del genoma de *S. cerevisiae* realizado por Peter et al., (2018) ya ha revelado amplias introgresiones de la especie hermana *S. paradoxus* en el linaje asociado a fermentación de agave, una característica que solo se ve a un nivel similar en otro linaje de *S. cerevisiae*. En el caso de *Kl. marxianus*, se ha demostrado que el genoma de una cepa aislada del jugo de agave forma un clado específico, bastante distante de todas las demás cepas secuenciadas de esta especie (Ortiz-Merino et al., 2018).

La colección de cultivos YMX-1.0 también es un recurso poderoso para estudiar la diversidad de especies de levaduras no convencionales. Aquí, informamos que Kz. humilis prospera en las fermentaciones de agave. Esta es una de las especies dominantes en los sourdoughs tipo I (Carbonetto et al., 2020; García-Ortega et al., 2022; Wittwer et al., 2022) y uno de los hongos más abundantes durante la preparación de una variedad de vegetales y bebidas fermentadas en China (Guan et al., 2020; Gullo et al., 2003; Wittwer et al., 2022). Sin embargo, se habían reportado pocos registros de la especie en América. Dada su distancia genética con los aislamientos de masa madre, las secuencias del genoma preliminar de cuatro de estas cepas de Kz. humilis de la fermentación de agave sugieren que pueden pertenecer a una población diferente, aunque se necesita analizar más aislados para validar esta hipótesis. Al comparar con aislados de otros entornos, las cepas de levadura en la colección de cultivos YMX-1.0 brindan la oportunidad de comparar las trayectorias evolutivas de especies distintas que han estado expuestas al mismo ambiente.

La mayoría de la atención en el estudio y conservación de la producción de agave se ha centrado en las plantas y prácticas de producción empleadas. Sin embargo, para preservar este sistema integralmente, todavía necesitamos considerar las comunidades microbianas que son esenciales para la fermentación abierta asociada con destilerías pequeñas y medianas. Nuestro estudio proporciona un recurso único para obtener un conocimiento más profundo de la diversidad genómica y funcional de las levaduras asociadas con este sistema de fermentación excepcional en una de las regiones megadiversas del mundo. La colección YMX-1.0 permite directamente la conservación de los microorganismos esenciales para las comunidades rurales en México. La investigación futura basada en nuestro trabajo contribuirá a una mejor comprensión no solo de cómo las comunidades microbianas contribuyen a la producción de destilados de agave, sino también a la biología y la historia natural de los microorganismos que viven en la interfaz de los entornos naturales y las destilerías tradicionales.

DISPONIBILIDAD DE DATOS

Las secuencias ITS y D1/D2 de los aislados de levadura correspondientes a las dos nuevas especies candidatas, *Pichia sp.* 1 y *Pichia sp.* 2, están disponibles en la base de datos de GenBank bajo los números de acceso OR034128, OR225219 y OR034113, OR034133. Las tres nuevas secuencias del genoma *de Kazachstania humilis* han sido depositadas en NCBI bajo los números de acceso JAIWYT000000000, JAIWYS000000000 y JAIWYU000000000.

AGRADECIMIENTOS

Estamos en deuda con más de cien productores de agave en los estados de Colima, Durango, Estado de México, Guanajuato, Guerrero, Jalisco, Michoacán, Oaxaca, Puebla, San Luis Potosí, Sonora, Tamaulipas y Zacatecas, México, que amablemente participaron en esta investigación compartiendo su conocimiento y proporcionando acceso a las muestras. Agradecemos la ayuda en el trabajo de campo y el muestreo a: Karina Abad, Evelia Acedo[†], Maritza Álvarez, Edgar Ángeles, Graciela Ángeles, Vanessa Arellano, Margareta Boege, Aida Carbajal, Jorge Carbajal, Carolina Castañeda, Bertha Castillo Ortega, Jason Paul Cox, Jorge Cuamatzi, Jilberto Dávila, Aarón De Luna, Marco Chávez, Hendrik Giersiepen, Alejandro González Anaya, Mathieu Harcot, Gerardo Hernández, Alejandro Juárez, Claudio López, Xavier López Medellín, Artemiza Martínez, Alberto Martínez López, José Martínez Rodríguez, Rodrigo Medellín, Alejandro Olmedo, Guillermo Ortiz, Anibal Reyes, Benito Salcedo, Ivan Sedeño, Nelly Selem, José Antonio Urban, Carina Uribe, y Karen Vaca. Agradecemos a Luis Aguilar, René Delgado, y Jair García Sotelo por su asistencia técnica y a Raúl Ortiz-Merino por su lectura crítica del manuscrito. Agradecemos a Iván Saldaña sus valiosas sugerencias para el trabajo de campo y el muestreo piloto en Oaxaca, a Cei Abreu-Goodger y Adrián Turjanski por su valiosa ayuda para obtener financiamiento y a Anne Gschaedler por su generoso asesoramiento durante las fases iniciales del proyecto.

FINANCIAMIENTO

Este trabajo fue financiado por el Consejo Nacional de Ciencia y Tecnología de México (Conacyt) a través de los proyectos FORDECYT-PRONACES/103000/2020 y CB-2016-01/284992, el Fondo de Investigación y Desarrollo Tecnológico del Cinvestav (SEP-CINVESTAV/023), el Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica DGAPA-UNAM (IN209021, IN230420) y el BBSRC del Reino Unido bajo el Global Challenges Research Fund (GCRF) Growing Research Capability call a través del CABANA Innovation Fund (BB/P027849/1). PG-C es estudiante de doctorado del Posgrado de Biotecnología de Plantas en el Cinvestav con una beca de Conacyt (862595); LFG-O es un investigador postdoctoral financiado por Conacyt (4133922); AMG-A fue financiada por una pasantía de investigación de CABANA. Los financiadores no tuvieron ningún papel en

el diseño del estudio, la recopilación y análisis de datos, la decisión de publicar ni la preparación del manuscrito.

CONTRIBUCIÓN DE LAS AUTORÍAS

Conceptualización: LM, EM, AD. Metodología: PG-C, AE-C, LFG-O, JAA-R, XA-D. Trabajo de campo: PG-C, AE-C, JAA-R, AC-R, FG-Z, AD-L, FM-F, CJ-H, AH-L, XA-D, LM, EM, AD. Investigación, trabajo de laboratorio: PG-C, AE-C, CGT-L, AC-R, MF-B, AC, MRK. Investigación, análisis formal: LFG-O, JAA-R, AMG-A. Visualización: PG-C, LFG-O, JAA-R, AMG-A, AD. Recursos: FG-Z, AD-L, FM-F, CJ-H, AH-L, LD, XA-D, MRK, LM, EM, AD. Administración del proyecto: SR-C, MF-B, AC. Obtención de fondos: AH-L, LD, LM, EM, AD. Supervisión: LD, LM, EM, AD. Escritura, preparación del borrador original: PG-C, LFG-O, AD. Escritura, revisión y edición: AE-C, LFG-O, JAA-R, XA-D, MRK, LM, EM, AD. Todos los autores leyeron y aprobaron la versión final del manuscrito.

CONFLICTO DE INTERÉS

Los autores declaran que la investigación se llevó a cabo sin la presencia de ninguna relación comercial o financiera que pudiera interpretarse como un conflicto de intereses potencial.

INFORMACIÓN SUPLEMENTARIA

Table S1. Confirmación de la asignación de especies mediante secuenciación de ITS (archivo PDF)

Table S2. Estadísticas de ensamblaje de cuatro cepas de *Kazachstania humilis* de la fermentación de agave y la cepa tipo CLIB 1323 (archivo PDF)

Table S3. Resumen de la anotación estructural SyRI de los genomas de Kazachstania humilis (archivo PDF)

Dataset S1. Colección YMX-1.0 y metadatos asociados (archivo XLSX)

REFERENCIAS CITADAS

- Aguirre, X., C. Illsley, y J. Larson. 2006. Dulce semblanza de los mezcales del Altiplano y del Balsas. México Desconocido 352:36-45.
- Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., Wang, X., Lippman, Z. B., Schatz, M. C., & Soyk, S. (2022). Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biology, 23(1), 1–19. https://doi.org/10.1186/s13059-022-02823-7
- Arellano-Plaza, M., Paez-Lerma, J. B., Soto-Cruz, N. O., Kirchmayr, M. R., & Gschaedler Mathis, A. (2022). Mezcal Production in Mexico: Between Tradition and Commercial Exploitation. Frontiers in Sustainable Food Systems, 6(March), 1–16. https://doi.org/10.3389/fsufs.2022.832532
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology, 19(5), 455–477. https://doi.org/10.1089/cmb.2012.0021
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1). https://doi.org/10.18637/jss.v067.i01
- Burini, J. A., Eizaguirre, J. I., Loviso, C., & Libkind, D. (2021). Non-conventional yeasts as tools for innovation and differentiation in brewing. Revista Argentina de Microbiologia, 53(4), 359–377. https://doi.org/10.1016/j.ram.2021.01.003
- Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A. S., & Yandell, M. (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Research, 18(1), 188–196. https://doi.org/10.1101/gr.6743907
- Carbonetto, B., Nidelet, T., Guezenec, S., Perez, M., Segond, D., & Sicard, D. (2020). Interactions between Kazachstania humilis yeast species and lactic acid bacteria in Sourdough. Microorganisms, 8(2). https://doi.org/10.3390/microorganisms8020240
- Chao, K. H., Barton, K., Palmer, S., & Lanfear, R. (2021). SangeranalyseR: Simple and Interactive Processing of Sanger Sequencing Data in R. Genome Biology and Evolution, 13(3), 1–7. https://doi.org/10.1093/gbe/evab028
- 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

- Ciani, M., & Comitini, F. (2011). Non-Saccharomyces wine yeasts have a promising role in biotechnological approaches to winemaking. Annals of Microbiology, 61(1), 25–32. https://doi.org/10.1007/s13213-010-0069-5
- Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. Australian Journal of Ecology, 18(1), 117–143. https://doi.org/10.1111/j.1442-9993.1993.tb00438.x
- Colunga-GarcíaMarín, P., Zizumbo-Villarreal, D., Torres, I., Casas, A., Figueredo Urbina, C. J., Rangel-Landa, Selene Delgado-Lemus, A., Vargas, O., Cabrera-Toledo, D., Aguirre-Dugua, X., Eguiarte, L. E., & Carrillo-Galván, G. (2017). Los agaves y las prácticas mesoamericanas de aprovechamiento, manejo y domesticación. Domesticación En El Continente Americano Vol. 2, Investigación Para El Manejo Sustentable de Recursos Genéticos En El Nuevo Mundo, May, 273–309.
- CONABIO (2006). Magueyes. Biodiversidad. Retrieved September 23, 2021, from https://www.biodiversidad.gob.mx/diversidad/alimentos/magueyes
- Cubillos, F. A. (2016). Exploiting budding yeast natural variation for industrial processes. Current Genetics, 62(4), 745–751. https://doi.org/10.1007/s00294-016-0602-6
- Cubillos, F. A., Gibson, B., Grijalva-Vallejos, N., Krogerus, K., & Nikulin, J. (2019). Bioprospecting for brewers: Exploiting natural diversity for naturally diverse beers. Yeast, 36(6), 383–398. https://doi.org/10.1002/yea.3380
- Dashko, S., Liu, P., Volk, H., Butinar, L., Piškur, J., & Fay, J. C. (2016). Changes in the relative abundance of two Saccharomyces species from oak forests to wine fermentations. Frontiers in Microbiology, 7(FEB), 1–12. https://doi.org/10.3389/fmicb.2016.00215
- De Cáceres, M., Legendre, P., & Moretti, M. (2010). Improving indicator species analysis by combining groups of sites. Oikos, 119(10), 1674–1684. https://doi.org/10.1111/j.1600-0706.2010.18334.x
- De Chiara, M., Barré, B. P., Persson, K., Irizar, A., Vischioni, C., Khaiwal, S., Stenberg, S., Amadi, O. C., Žun, G., Doberšek, K., Taccioli, C., Schacherer, J., Petrovič, U., Warringer, J., & Liti, G. (2022). Domestication reprogrammed the budding yeast life cycle. Nature Ecology and Evolution, 6(4), 448–460. https://doi.org/10.1038/s41559-022-01671-9
- Delcher, A. L., Salzberg, S. L., & Phillippy, A. M. (2003). Using MUMmer to Identify Similar Regions in Large Sequence Sets . Current Protocols in Bioinformatics, 00(1). https://doi.org/10.1002/0471250953.bi1003s00
- Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. Genome Biology, 20(1), 1–14. https://doi.org/10.1186/s13059-019-1832-y

- Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Saels, V., Herrera-Malaver, B., Merlevede, A., Roncoroni, M., Voordeckers, K., Miraglia, L., Teiling, C., Steffy, B., Taylor, M., Schwartz, A., Richardson, T., White, C., Baele, G., Maere, S., & Verstrepen, K. J. (2016). Domestication and Divergence of Saccharomyces cerevisiae Beer Yeasts. Cell, 166(6), 1397-1410.e16. https://doi.org/10.1016/j.cell.2016.08.020
- García, E. (1998). Climas, escala 1:1,000,000 (clasificación de Koppen, modificado por García). Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO), Mexico.
- García-Gamboa, R., Kirchmayr, M. R., Gradilla-Hernández, M. S., Pérez-Brocal, V., Moya, A., & González-Avila, M. (2021). The intestinal mycobiota and its relationship with overweight, obesity and nutritional aspects. Journal of Human Nutrition and Dietetics, 34(4), 645–655. https://doi.org/10.1111/jhn.12864
- García-Ortega, L. F., Colón-González, M., Sedeño, I., Santiago-Garduño, E., Avelar-Rivas, J. A., Kirchmayr, M. R., DeLuna, A., Delaye, L., Morales, L., & Mancera, E. (2022). Draft Genome Sequence of a Kazachstania humilis Strain Isolated from Agave Fermentation. Microbiology Resource Announcements, 11(3), 20–21. https://doi.org/10.1128/mra.01154-21
- Gencturk, E., & Ulgen, K. O. (2022). Understanding HMF inhibition on yeast growth coupled with ethanol production for the improvement of bio-based industrial processes. Process Biochemistry, 121(June), 425–438. https://doi.org/10.1016/j.procbio.2022.07.015
- Giannakou, K., Cotterrell, M., & Delneri, D. (2020). Genomic Adaptation of Saccharomyces Species to Industrial Environments. Frontiers in Genetics, 11(August), 1–10. https://doi.org/10.3389/fgene.2020.00916
- Giannakou, K., Visinoni, F., Zhang, P., Nathoo, N., Jones, P., Cotterrell, M., Vrhovsek, U., & Delneri, D. (2021). Biotechnological exploitation of Saccharomyces jurei and its hybrids in craft beer fermentation uncovers new aroma combinations. Food Microbiology, 100, 103838. https://doi.org/10.1016/j.fm.2021.103838
- Godoy, L., Silva-Moreno, E., Mardones, W., Guzman, D., Cubillos, F. A., & Ganga, A. (2017). Genomics Perspectives on Metabolism, Survival Strategies, and Biotechnological Applications of Brettanomyces bruxellensis LAMAP2480. Journal of Molecular Microbiology and Biotechnology, 27(3), 147–158. https://doi.org/10.1159/000471924
- Goel, M., Sun, H., Jiao, W. B., & Schneeberger, K. (2019). SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biology, 20(1), 1–13. https://doi.org/10.1186/s13059-019-1911-0
- Gorsich, S. W., Dien, B. S., Nichols, N. N., Slininger, P. J., Liu, Z. L., & Skory, C. D. (2006). Tolerance to furfural-induced stress is associated with pentose phosphate

pathway genes ZWF1, GND1, RPE1, and TKL1 in Saccharomyces cerevisiae. Applied Microbiology and Biotechnology, 71(3), 339–349. https://doi.org/10.1007/s00253-005-0142-3

- Gostinčar, C. (2020). Towards genomic criteria for delineating fungal species. Journal of Fungi, 6(4), 1–18. https://doi.org/10.3390/jof6040246
- Guan, Q., Zheng, W., Huang, T., Xiao, Y., Liu, Z., Peng, Z., Gong, D., Xie, M., & Xiong, T. (2020). Comparison of microbial communities and physiochemical characteristics of two traditionally fermented vegetables. Food Research International, 128(July 2019), 108755. https://doi.org/10.1016/j.foodres.2019.108755
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. Bioinformatics, 29(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086
- Gullo, M., Romano, A. D., Pulvirenti, A., & Giudici, P. (2003). Candida humilis -Dominant species in sourdoughs for the production of durum wheat bran flour bread. International Journal of Food Microbiology, 80(1), 55–59. https://doi.org/10.1016/S0168-1605(02)00121-6
- Jacques, N., Sarilar, V., Urien, C., Lopes, M. R., Morais, C. G., Uetanabaro, A. P. T., Tinsley, C. R., Rosa, C. A., Sicard, D., & Casaregola, S. (2016). Three novel ascomycetous yeast species of the Kazachstania clade, Kazachstania saulgeensis sp. Nov., Kazachstania serrabonitensis sp. nov. and Kazachstania australis sp. nov. Reassignment of Candida humilis to Kazachstania humilis f.a. comb. nov. and Cand. International Journal of Systematic and Evolutionary Microbiology, 66(12), 5192– 5200. https://doi.org/10.1099/ijsem.0.001495
- Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nature Communications, 9(1), 1–8. https://doi.org/10.1038/s41467-018-07641-9
- Jolly, N. P., Varela, C., & Pretorius, I. S. (2014). Not your ordinary yeast: Non-Saccharomyces yeasts in wine production uncovered. FEMS Yeast Research, 14(2), 215–237. https://doi.org/10.1111/1567-1364.12111
- Johnson, E. A., & Echavarri-Erasun, C. (2011). Yeast Biotechnology. C. P. Kurtzman, J.W. Fell, & T. Boekhout, The Yeast A Taxonomy Study (pag. 21-45). ElServier.
- 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

- Ko, B. J., Lee, C., Kim, J., Rhie, A., Yoo, D. A., Howe, K., Wood, J., Cho, S., Brown, S., Formenti, G., Jarvis, E. D., & Kim, H. (2022). Widespread false gene gains caused by duplication errors in genome assemblies. Genome Biology, 23(1), 1–26. https://doi.org/10.1186/s13059-022-02764-1
- Kolde R. Pheatmap: pretty heatmaps. R package version. 2012;1(2):726.
- Korlach, J., Gedman, G., Kingan, S. B., Chin, C. S., Howard, J. T., Audet, J. N., Cantin, L., & Jarvis, E. D. (2017). De novo PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads. *GigaScience*, 6(10), 1–16. https://doi.org/10.1093/gigascience/gix085
- Lachance, M. A. (1995). Yeast communities in a natural tequila fermentation. Antonie van Leeuwenhoek, 68(2), 151–160. https://doi.org/10.1007/BF00873100
- 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
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. Bioinformatics, 23(21), 2947–2948. https://doi.org/10.1093/bioinformatics/btm404
- Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., & Carey, V. J. (2013). Software for Computing and Annotating Genomic Ranges. PLoS Computational Biology, 9(8), 1–10. https://doi.org/10.1371/journal.pcbi.1003118
- Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., Davey, R. P., Roberts, I. N., Burt, A., Koufopanou, V., Tsai, I. J., Bergman, C. M., Bensasson, D., O'Kelly, M. J. T., Van Oudenaarden, A., Barton, D. B. H., Bailes, E., Nguyen, A. N., Jones, M., ... Louis, E. J. (2009). Population genomics of domestic and wild yeasts. Nature, 458(7236), 337–341. https://doi.org/10.1038/nature07743
- Liti, G., Warringer, J., & Blomberg, A. (2017). Isolation and laboratory domestication of natural yeast strain. Cold Spring Harbor Protocols, 2017(8), 626–630. https://doi.org/10.1101/pdb.prot089052
- Mancilla-Margalli, N. A., & López, M. G. (2002). Generation of Maillard compounds from inulin during the thermal processing of Agave tequilana Weber var. azul. Journal of Agricultural and Food Chemistry, 50(4), 806–812. https://doi.org/10.1021/jf0110295
- Marsit, S., Leducq, J. B., Durand, É., Marchant, A., Filteau, M., & Landry, C. R. (2017). Evolutionary biology through the lens of budding yeast comparative genomics. Nature Reviews Genetics, 18(10), 581–598. https://doi.org/10.1038/nrg.2017.49

- Molinet, J., & Cubillos, F. A. (2020). Wild Yeast for the Future: Exploring the Use of Wild Strains for Wine and Beer Fermentation. Frontiers in Genetics, 11(November), 1–8. https://doi.org/10.3389/fgene.2020.589350
- Nolasco-Cancino, H., Santiago-Urbina, J. A., Wacher, C., & Ruíz-Terán, F. (2018). Predominant yeasts during artisanal mezcal fermentation and their capacity to ferment maguey juice. Frontiers in Microbiology, 9(DEC), 1–12. https://doi.org/10.3389/fmicb.2018.02900
- Oksanen, r J., Simpson, G. L., Blanchet, F. G., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., Durand, S., Beatriz, H., Evangelista, A., Friendly, M., Hannigan, G., Hill, M. O., ... Weedon, J. (2022). Title Community Ecology Package. Vegan: Community Ecol Package.
- Ortiz-Merino, R. A., Varela, J. A., Coughlan, A. Y., Hoshida, H., da Silveira, W. B., Wilde, C., Kuijpers, N. G. A., Geertman, J. M., Wolfe, K. H., & Morrissey, J. P. (2018).
 Ploidy variation in Kluyveromyces marxianus separates dairy and non-dairy isolates.
 Frontiers in Genetics, 9(MAR), 1–16. https://doi.org/10.3389/fgene.2018.00094
- Pagès H, Aboyoun P, Gentleman R, DebRoy S (2023). _Biostrings: Efficient manipulation of biological strings. doi:10.18129/B9.bioc.Biostrings, R package version 2.68.1,https://bioconductor.org/packages/Biostrings.
- Páez-Lerma, J. B., Arias-García, A., Rutiaga-Quiñones, O. M., Barrio, E., & Soto-Cruz, N. O. (2013). Yeasts Isolated from the Alcoholic Fermentation of Agave duranguensis During Mezcal Production. Food Biotechnology, 27(4), 342–356. https://doi.org/10.1080/08905436.2013.840788
- Park, S. E., Koo, H. M., Park, Y. K., Park, S. M., Park, J. C., Lee, O. K., Park, Y. C., & Seo, J. H. (2011). Expression of aldehyde dehydrogenase 6 reduces inhibitory effect of furan derivatives on cell growth and ethanol production in Saccharomyces cerevisiae. Bioresource Technology, 102(10), 6033–6038. https://doi.org/10.1016/j.biortech.2011.02.101
- Peter, J., Chiara, M. De, Friedrich, A., Yue, J., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., & Liti, G. (2018). Saccharomyces cerevisiae isolates Saccharomyces cerevisiae isolates. Nature, 556, 339–344.
- Sayers, E. W., Beck, J., Brister, J. R., Bolton, E. E., Canese, K., Comeau, D. C., Funk, K., Ketter, A., Kim, S., Kimchi, A., Kitts, P. A., Kuznetsov, A., Lathrop, S., Lu, Z., McGarvey, K., Madden, T. L., Murphy, T. D., O'Leary, N., Phan, L., ... Ostell, J. (2020). Database resources of the National Center for Biotechnology Information. Nucleic Acids Research, 48(D1), D9–D16. https://doi.org/10.1093/nar/gkz899

- Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A.,
 Chen, W., Bolchacova, E., Voigt, K., Crous, P. W., Miller, A. N., Wingfield, M. J.,
 Aime, M. C., An, K. D., Bai, F. Y., Barreto, R. W., Begerow, D., Bergeron, M. J.,
 Blackwell, M., ... Schindel, D. (2012). Nuclear ribosomal internal transcribed spacer
 (ITS) region as a universal DNA barcode marker for Fungi. Proceedings of the
 National Academy of Sciences of the United States of America, 109(16), 6241–6246.
 https://doi.org/10.1073/pnas.1117018109
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and postanalysis of large phylogenies. Bioinformatics, 30(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033
- 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), 1–16. https://doi.org/10.1093/femsyr/fov002
- Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Molecular Biology and Evolution, 38(7), 3022–3027. https://doi.org/10.1093/molbev/msab120
- Turker, M. (2015). Yeast Biotechnology: Diversity and Applications. Pakmaya, Kocaeli, Turkey: R&D, Environment and Quality Coordination Manager.
- Verdugo Valdez, A., Segura Garcia, L., Kirchmayr, M., Ramírez Rodríguez, P., González Esquinca, A., Coria, R., & Gschaedler Mathis, A. (2011). Yeast communities associated with artisanal mezcal fermentations from Agave salmiana. Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology, 100(4), 497–506. https://doi.org/10.1007/s10482-011-9605-y
- Warringer, J., Zörgö, E., Cubillos, F. A., Zia, A., Gjuvsland, A., Simpson, J. T., Forsmark, A., Durbin, R., Omholt, S. W., Louis, E. J., Liti, G., Moses, A., & Blomberg, A. (2011). Trait variation in yeast is defined by population history. PLoS Genetics, 7(6). https://doi.org/10.1371/journal.pgen.1002111
- Wittwer, A. E., Sicard, D., & Howell, K. S. (2022). Kazachstania humilis. Trends in Microbiology, 30(10), 1012–1013. https://doi.org/10.1016/j.tim.2022.05.007