Interspecific hybridization and aneuploidy as adaptive mechanisms in *Saccharomyces* yeasts

PhD Thesis

Miguel Morard Pedrouzo

October, 2019

Supervisors :

Dr. Eladio Barrio Esparducer and **Dr. Christina Toft**

VNIVERSITAT DE VALÈNCIA

Dpto. de Bioquímica y Biología Molecular

Doctorado en Biomedicina y Biotecnología

VNIVERSITAT DE VALÈNCIA Departament de Genètica

I2SysBio INSTITUTE FOR INTEGRATIVE SYSTEMS BIOLOGY

El Dr. Eladio Barrio Esparducer, Profesor Titular del Departamento de Genética de la Universitat de València y la Dra. Christina Toft, Investigadora SEJI del Instituto de Biología Integrativa de Sistemas (I2SysBio), Universitat de València.

CERTIFICAN

Que Don Miguel Morard Pedrouzo, Licenciado en Biología por la Universidad de Santiago de Compostela, ha realizado bajo su dirección el trabajo titulado: "Interspecific hybridization and aneuploidy as adaptive mechanisms in Saccharomyces yeasts.", que presenta para optar al grado de Doctor en el programa de Biomedicina y Biotecnología por la Universitat de València. Asimismo, certifican haber dirigido y supervisado tanto los distintos aspectos del trabajo como su redacción.

Y para que conste a los efectos oportunos, en cumplimiento de la legislación vigente, firman el presente certificado en

Valencia, a 10 de octubre de 2019

Los directores:

Fdo. Eladio Barrio

Fdo. Christina Toft

This PhD Thesis work was funded by a predoctoral grant from the Consellería de Educación, Investigación, Cultura y Deportes from the Generalitat Valenciana, reference ACIF/2015/194. I also was awarded a research stay grant by the same entity to realize a three months stay at the Bijvoet Center for Biomolecular Research at the Utrecht University, reference BEFPI/2018/014. The work was funded by grants AGL2015-67504-C3-R and RTI2018-093744-B-C32 from the Spanish Government and European Union ERDF-FEDER.

Acknowledgements

Caminante, son tus huellas
el camino y nada más;
Caminante, no hay camino,
se hace camino al andar.
Al andar se hace el camino,
y al volver la vista atrás
se ve la senda que nunca
se ha de volver a pisar.
Caminante no hay camino
sino estelas en la mar.

Antonio Machado

En noviembre del año 2007 presenté mi travail de maturité. Había trabajado durante un año sobre la cultura medieval del Camino de Santiago (literatura, arquitectura, etc). Estos trabajos son muy pequeños aunque entonces escribirlo se nos hacía un mundo. Había decidido hacer el trabajo en español porque entonces tenía una cosa muy clara para mi futuro: estudiaría en Santiago. En la presentación del trabajo usé la letra de la canción de Serrat Cantares porque el tema me quedaba bien para ilustrarlo (como no), aunque no recuerdo si entonces supiera que era de Machado. Pasaron años y mucho camino, pero me sigue costando tanto escribir, y mucho más estos agradecimientos. Lo que ya cobra otro sentido es este poema. Me acompaña desde entonces. Reaparece en momentos claves. El camino que recorrí desde 2007 y desde que cogí el primer billete de avión (sin vuelta) en el 2008 estuvo lleno de experiencias y sobresaltos, pero sobre todo lleno de gente. El camino que me llevó hasta esta tesis está repleto de ayuda y de cariño sin lo que habría sido imposible

desbrozarlo. Al andar se hace el camino, pero el caminante nunca camina solo.

Los primeros agradecimientos van a mis directores, Eladio Barrio y Christina Toft. Gracias Eladio por haberme dado esa oportunidad y haberme dejado vía libre para explorar mis ideas. Gracias por todas nuestras conversaciones y nuestros debates inspiradores y todo lo que me enseñaste. Christina, thank you for being there even in the worst moments. Thanks for trusting me and helping with all the hard coding and making sense of what had no apparent sense.

Esta tesis no podría haber existido si no fuera por una gran suerte para conseguir financiarla. Tengo que agradecerles a todos los que luchamos por las VALID 2015. Gracias a FEBiotec por el apoyo desde el principio y a Jesús por haberme dicho : “Sí, Miguel, intentémoslo”. Maria-José, gracias por todos estos mensajes y esta ayuda tanto en esos meses de 2015 como a lo largo de los siguientes años. En el 2015, todos los que estábamos de reserva de las ayudas predoctorales VALID tuvimos un gran golpe de suerte. Dos personas llegaron a la Consellería de educación con oídos y corazón. Gracias Vicent Marzá y Josefina Bueno porque sin vuestra valentía, vuestro trabajo y vuestra voluntad esta tesis no estaría aquí, así como otra treintena que también se habrían quedado en el cajón de los deseos o del trabajo no remunerado. La foto de nuestra reunión del 23 de julio 2015 estuvo acompañándome todo este tiempo para recordarme que nada se obtiene si no es luchando por ello.

En el IATA nuestro “pequeño” despacho estaba integrado en un grupo cuyos responsables merecen muchos agradecimientos. Amparo, gracias por apoyarnos y resolver todos los problemas con la eficiencia que te caracteriza. Ya sabes, eres también un poco directora de este trabajo. Agradezco también a José-Manuel, a Sergi y a Maite por todos los seminarios, las reuniones y tantas ideas y proyectos. Al llegar al IATA, y al SBYBI en particular, uno se encuentra con una familia nueva. Este capítulo de los agradecimientos es bastante difícil de escribir. No quisiera olvidar a nadie aunque ya sabéis cómo es mi memoria… Empezaré esa ruta por el fondo del pasillo. Ceci, por ser siempre tan enérgica y atenta. Anto, siempre ahí, siempre de gran ayuda y de sabio consejo. Lucía por ser más reivindicativa que yo mismo y siempre dispuesta a buscar planes. Ric por confundir a todo el mundo en nuestras fotos y por tu sentido del humor. Sara, gracias por ser tan cariñosa y empática. Ying por ser tan divertida y risueña. Fani, gracias por darme nociones de teoría del color y darme muchos papers. Aurora por cuidar de todos. Alba por tu cariño y tu energía. Javi por más proyectos e ideas conjuntos. Romain pour mettre une touche française au labo. Laura (Pérez) porque eres la única que ha conseguido hacerme hacer trabajo de laboratorio. David porque siempre estás para todos y se te echa mucho de menos, compañero de cumple. Alba por hacernos descubrir las maravillas culinarias de Requena y por esa sonrisa imborrable. Sebastián por esa energía y compartir la vida de bicicleta. Dolores por mantener la sonrisa siempre. Yaiza también eres un poco de la tercera y gracias por ser una maravilla de persona y de amiga. Ana-Cris, gracias por todo lo

que haces y por tu eficiencia y tu dedicación. Roberto por tu optimismo y demostrar que se puede. Peris, gracias por toda tu ayuda, tus ánimos y por ser un ejemplo para mí. Gracias a Jiry, Clara, Eli, Jesús, Víctor, Carla, Sonia, Gabriel, Andrea, Ceci, Tania, Raquel y a todos los que pasaron por esta tercera planta del IATA que es como nuestra segunda casa.

Hay en esa tercera planta un despachito peculiar. Algunos dicen que está habitado por seres extraños… Es el 313. En este cubículo acristalado al que pocos osan asomarse viven dos bioinformáticas cuya amistad es lo mejor que puede llevarse uno de la tesis. María, empezaste con Javi pero pronto te diste cuenta que en bioinformática se vive mejor. Gracias por todas nuestras risas, toda tu ayuda y tu apoyo. Siempre serás nuestra fallera mayor. Laura, de TFM te dimos secuencias de SOLiD y aún así te quedaste. Luchas como nadie para conseguir lo que quieres y me impresionarás siempre por ello. Trabajar contigo es un lujo que no tiene mucha gente. Gracias a las mejores compañeras; esta tesis es lo que es gracias a vosotras dos.

La familia sin embargo lleva mucho más tiempo aguantando mis caídas en el camino. No hay mejor ayuda que sentir el apoyo y orgullo de quienes te han visto y hecho crecer. E coma son unha mezcla de raclette e pulpo, la prochaine section sera écrite telle que je suis.

Je voulais commencer par le Valais pour remercier tata Lucette pour m'appeler tous les dimanches et être là même depuis la distance. Grand-maman, bien sûr, toutes ces années à mettre une bougie au cimetière pour m'aider à chaque examen, merci de toujours penser à moi. Aussi à cette famille à qui je tiens, merci Olga, Régis, Tania et bien sûr mon amour de filleul Mateo.

A miña familia galega tamén estivo a apoiarme e facer que a carreira fora algo especial. Abuela, sempre axudando, sempre querendo o mellor. Manolo, porque a "nai da ciencia é a experiencia". Rubén que sempre estiveches para dar bos consellos. Xoel, que eres unha pila de enerxía e cariño. Beti, que che poido decir a parte de grazas por ser a mellor das madriñas, e pola nosa complicidade. Miguel, por escoitar e axudar a escapar a un café cando as cousas empezan a cansar. Mateo, por poñerme esa sonrisa que sempre é capaz de emocionarme.

David, hermanito, no me olvido de ti no. Nos dicen que somos como el sol y la luna. Yo creo que somos como dos copos de nieve. Estos son distintos porque no hay dos iguales, pero sigue siendo nieve. Gracias por estar cuando te necesito y por ser mi apoyo y cómplice de juego.

Esto está acabando. Esta tesis está dedicada a vosotros. Papa, Mamá… Vous m'avez donné tout pour réussir à realiser mes rêves. Me habéis dado independencia, seguridad, medios, idiomas, consejos, cariño y apoyo… Tanto que es imposible agradecerlo todo si no es dedicándoos este trabajo porque es tan vuestro como mío. Sois quienes han dado más para que sea feliz y que pueda hacer siempre lo que más me apetece. Je suis très fier de venir d'oú je viens. Je suis qui je suis

parce que vous avez posé l’échaffaudage sur lequel me construire et m’avez donné les instruments pour que je puisse travailler par moi-même. Je vous aime plus que tout y no cabe todo para poder deciros gracias.

En el último tramo del camino, apareciste. Eres especial porque contigo comparto las botas. Contigo comparto las pisadas desde entonces. Has revolucionado este camino, le has dado luz y le has puesto indicadores con la palabra felicidad. Gracias Blanca porque sin ti esto no tendría ya ningún sentido. Te amo.

A mis padres

Contents

List of Figures

List of Tables

Resumen amplio en español

Las levaduras del género *Saccharomyces* son de reconocida importancia tanto por ser un modelo genético básico como por su uso biotecnológico. Tienen un genoma muy compacto. En sus aproximadas 12Mb tienen codificados 6000 genes. Además, muy pocos de estos genes contienen intrones. Esto convierte a estos organismos en unos modelos simples para el estudio genético y genómico. El género tiene 9 especies. La divergencia nucleotídica entre especies es interesante ya que puede cubrir la divergencia entre humanos y aves (*S. cerevisiae* vs *S. kudriavzevii*) como entre humanos y primates (*S. cerevisiae* vs *S. paradoxus*). De estas especies, *S. cerevisiae* es la más estudiada y usada biotecnológicamente. Sin embargo las otras especies tienen distintas propiedades fenotípicas que están levantando interés para la industria. Una de ellas es la capacidad de crecimiento a bajas temperaturas que se observa en *Saccharomyces* no cerevisiae. Estas diferencias fenotípicas y la divergencia entre especie convierten al género en un buen modelo para estudios genéticos y evolutivos.

Los mecanismos que generan diversidad genómica son muy diversos. Si un genoma es heterocigoto tiene distintos alelos en el genoma y puede usar mecanismos como la pérdida de heterocigosidad para fijar uno u otro alelo en respuesta a estreses. Con el avance de las tecnologías de secuenciación del ADN y la explosión de estudios genómicos, la importancia de mecanismos de diversificación genómica como las aneuploidías, las polyploidías y la hibridación está quedando más clara. En *Saccharomyces* múltiples estudios están encontrando que variar el número de copias de ciertos cromosomas es un mecanismo muy frecuente en ambientes variables como son los ambientes industriales. Un hecho interesante es que en estos ambientes también se encontraron distintos híbridos interespecíficos del género. El ejemplo más famosos es el híbrido entre *S. cerevisiae* y *S. eubayanus* que es el responsable de la fermentación de cervezas de tipo lager. Pero se aislaron híbridos entre otras especies también como entre *S. cerevisiae* y *S. kudriavzevii* que se aislaron de fermentaciones de vino, cerveza o sidra. La hibridación entre estas especies se piensa que es el producto de la selección de cepas con alta capacidad fermentativa (que heredaría de *S. cerevisiae*) pero que además sean capaces de fermentar a temperaturas más bajas (capacidad que heredarían de la otra especie). Esto supondría que la hibridación es otro mecanismo de adaptación a ambientes cambiantes.

Teniendo en cuenta todo lo expuesto, en esta tesis nos propusimos hacer uso de las nuevas tecnologías de secuenciación para estudiar los mecanismos involucrados en la adaptación de las levaduras a ambientes industriales. Los mecanismos a los que prestamos más relevancia son las aneuploidías y la hibridación ya que son mecanismos de creciente interés que, gracias al avance de la secuenciación masiva, ahora son posibles de estudiar en profundidad. Por ello, los objetivos

de esta tesis son : 1) Analizar las diferencias genómicas entre cepas de *S. cerevisiae* que muestran distinta tolerancia al etanol para determinar la contribución de la variación en número de copias de los cromosomas en su diferencia fisiológica. 2) Realizar un análisis de genómica comparativa de híbridos naturales entre *S. cerevisiae* y *S. kudriavzevii* para averiguar el mecanismo de conjugación que usaron para hibridar y cómo la hibridación influye en sus características fisiológicas. 3) Analizar los factores que influyen en la inestabilidad genómica de híbridos entre *S. cerevisiae* y *S. kudriavzevii*. 4) El estudio de la evolución temprana de híbridos artificiales entre *S. cerevisiae* y *S. kudriavzevii* para entender su respuesta, tanto a nivel genómico como transcriptómico, a presiones selectivas presentes en ambientes fermentativos, la baja temperatura y el estrés por etanol.

En el primer capítulo de esta tesis nos centramos en el estudio genómico de cepas de *S. cerevisiae* con distintas tolerancia al etanol. El etanol es uno de los mayores estreses que sufren las levaduras en ambientes industriales. Esta pequeña molécula puede pasar a través de la membrana plasmática, incrementando su permeabilidad y fluidez. El etanol también puede interferir en el plegamiento de proteínas, el balance redox de la célula, además de ser mutagénico. Las levaduras producen etanol al consumir azúcares, por lo que han desarrollado mecanismos para protegerse de los efectos adversos que tiene. Sin embargo, las distintas especies del género no toleran los mismos niveles de etanol. La más tolerante de las especies es *S. cerevisiae*, pero existen diferencias intra-específicas. La tolerancia al etanol es una característica muy relevante a nivel industrial y, por ello, existen muchos estudios sobre sus mecanismos. Pero, al ser una característica compleja, que puede involucrar a más de 200 genes, estos mecanismos siguen poco conocidos. En este capítulo secuenciamos el genoma de cepas cuya tolerancia al etanol se había mostrado que eran distintas.

Una de muy baja tolerancia al etanol, la cepa Temohaya-MI26, aislada de la fermentación de mezcal en Méjico. Dos de muy alta tolerancia, las cepas CECT10094 y GBFlor-C, ambas cepas de flor, responsables del envejecimiento biológico de vinos de jeréz. Secuenciamos también una cepa de tolerancia intermedia la T73, aislada de vino. Además usamos datos de secuenciación extraídos de bases de datos de otra cepa de tipo flor pero con tolerancia intermedia, la EC1118, aislada de champagne en Francia.

El ensamblaje y la anotación de los genomas de las cepas nos permitió extraer unas 6000 secuencias codificantes por genomas. Alineamos 2115 secuencias codificantes presentes en las cepas de este estudio y 38 cepas representantes de distintos clados reconocidos en *S. cerevisiae* y concatenamos dichos alineamientos para realizar una filogenia, y así determinar el origen de las cepas de estudio y determinar si las diferencias en tolerancia al etanol se explican por su origen. Determinamos que la cepas Temohaya-MI26 no pertenecía a ninguno de los grupos considerados. Las otras cepas se agruparon con las cepas de vinos y europeas. Sin embargo las cepas de flor forman un subgrupo diferenciado en el que se encuentras las cepas GBFlor-C, CECT10094 y EC1118. Como las cepas más tolerantes y la EC1118, con una tolerancia intermedia, se agrupan juntas, concluimos que aunque puede existir un efecto del origen, consistente en una historia de presiones selectivas diferentes en ambientes distintos, existe variación dentro de estos grupos.

Estudiamos, mediante mapeo de las lecturas y búsqueda de variantes, los niveles de heterocigosis y el contenido genómico de las cepas de interés. Observamos que los niveles de heterocigosis diferían entre los distintos aislados. La cepa Temohaya-MI26 era completamente homo-

cigota. la T73 y la GBFlor-C tenían niveles de heterocigosis intermedios mientras que la EC1118 y la CECT10094 tenían niveles muy altos. Estas diferencias sin embargo no se podían relacionar con diferencias en la tolerancia al etanol. Gracias a las frecuencias de las variantes y a análisis por citometría de flujo determinamos la ploidía y las posibles aneuploidías de las cepas. A excepción de la GBFlor-C, que era triploide, todas las cepas eran diploides. Curiosamente, encontramos aneuploidías exclusivamente en las dos cepas más tolerantes al etanol. La cepa CECT10094 contenía copias suplementarias del cromosoma III y del cromosoma XII, y la cepa GBFlor-C del cromosoma III.

La coincidencia en las dos cepas más tolerantes al etanol de una polisomía en el cromosoma III nos hizo investigar si pudiese tener un efecto a nivel de transcripción bajo el estrés por etanol. Usamos datos de transcriptómica disponibles para la cepa más tolerante CECT10094 y la menos tolerante Temohaya-MI26 bajo estrés por etanol y sin estrés para descifrar la estructura de la transcripción a nivel cromosómico. Cuando la cepa menos tolerante no está estresada por etanol, el nivel de transcripción de los genes del cromosoma III es normal. Sin embargo en la fase temprana después de que se le someta a estrés por etanol la transcripción del cromosoma III se dispara. La cepa más tolerante, ella, tenía los niveles de expresión de este cromosoma ya más altos de lo normal sin necesidad de estrés, debido a que contenía más copias de este. Esto apunta a que la presencia de una aneuploidía permite tener más expresión de genes relevantes para la tolerancia al etanol. Quisimos determinar si el cambio de copias era dependiente del fondo genético. Para ello usamos datos de ploidía, aneuploidía y el fenotipado en estrés por etanol extraídos de un estudio de 1000 cepas distintas de *S. cerevisiae*. En las cepas diploides vimos una tendencia clara. Aunque

la variación entre cepas se deba a muchos factores, y por ello las cepas euploides tenían un rango extenso de tolerancia, las cepas aneuploides para el cromosoma III con dos copias suplementarias tenían mayor tolerancia que las que tenían una única copia extra. Las que peor toleraban el etanol eran las cepas a las que les faltaba una copia del cromosoma. En conjunto estos resultados indican que las aneuploidías en el cromosoma III, aunque no sean el único factor, contribuyen significativamente a la tolerancia al etanol, posiblemente por la mayor expresión de algunos genes presentes en el, y que es una característica independiente del fondo genético.

Aunque las evidencias de que existía una relación entre aneuploidía y tolerancia al etanol eran varias, era necesario determinar si era un cambio adaptativo. Para ello era necesario tener la misma cepa con y sin aneuploidía para determinar si variaba el fenotipo. Usamos una cepa de laboratorio evolucionada bajo condiciones de estrés por etanol que había adquirido una aneuploidía en el cromosoma III. Integramos un cassette en una copia del cromosoma III que nos permitió seleccionar aquellas células que perdiesen una única copia del cromosoma, retornando a la euploidía. Fenotipamos la cepa original con aneuploidía y la misma cepa a la que le habíamos quitado esta copia suplementaria y observamos que la tolerancia al etanol bajaba. Este experimento nos permitió determinar que el cambio no era sólo circunstancial sino que tenía un valor adaptativo en el mismo fondo genético.

En el segundo capítulo usamos genómica comparada para estudiar los procesos de conjugación para la formación de híbridos industriales entre *S. cerevisiae* y *S. kudriavzevii* y los mecanismos de evolución genómica que ocurren en estos. La hibridación es un fenómeno muy común en la

naturaleza. Es ampliamente reconocido en plantas, pero también existe en animales y hongos. En el género *Saccharomyces* el ejemplo más conocido es el híbrido entre *S. cerevisiae* y *S. eubayanus*, también llamado *S. pastorianus* o *S. carlsbergensis*, responsable de la fermentación de la cerveza tipo lager. Los híbridos entre *S. cerevisiae* y *S. kudriavzevii* se encontraron asociados a distintas fermentaciones, como vino o cerveza, e incluso asociados a ambientes clínicos.

Secuenciamos el genoma de siete híbridos entre *S. cerevisiae* y *S. kudriavzevii* de distintos orígenes. Realizamos el ensamblaje y anotación de los genomas y con las secuencias codificantes, y la secuencias de representantes de distintos grupos de *S. cerevisiae,* y todas las disponibles de *S. kudriavzevii* reconstruimos la filogenia de ambos subgenomas de los híbridos por separado. Primero, analizamos el subgenoma de *S. cerevisiae*. Las cepas de vinos se agruparon con las cepas de *S. cerevisiae* vínicas en dos subgrupos distintos, tres con las de flor y dos con las de vinos. Otra cepa, MR25, aislada de tracto respiratorio no se agrupaba claramente con ningún clado. La cepa CECT11002, aislada de cerveza trapista, se agrupaba con el grupo beer2 que contiene cepas de cervezas. El subgenoma de *S. kudriavzevii* también muestra que los híbridos forman distintos grupos pero no se asocian con cepas de la especie que estén disponibles. En conjunto este análisis demuestra que la hibridación no ocurrió una única vez. Esto es, varios eventos de hibridación independientes, al menos tres, en ambientes distintos, dieron lugar a estos híbridos. Esto confirma que la hibridación es un fenómeno frecuente en ambientes industriales.

Otro aspecto a estudiar es el contenido genómico de los híbridos y su ploidía que permiten esclarecer los mecanismos de conjugación usados para su formación. Usamos para ello mapeo de

lecturas y búsqueda de variantes junto con datos previos de análisis de ploidía por citometría de flujo. Observamos que la mayoría de los híbridos *S. cerevisiae* x *S. kudriavzevii* eran triploides. En estos, la contribución de *S. cerevisiae* era diploide mientras que la de *S. kudriavzevii* haploide. Analizando los niveles heterocigosis de la parte *S. cerevisiae* vimos que estos eran parecidos a los observados en cepas de *S. cerevisiae* de su mismo origen. Esto indica que los cruces para formar estos híbridos ocurrieron entre una *S. cerevisiae* vínica o cervecera diploide y heterocigota con una espora haploide. Sería imposible obtener un resultado como éste por duplicación selectiva del subgenoma de *S. cerevisiae* después de un cruce espora-espora y una posterior acumulación de mutación. Por ello propusimos que estos híbridos, la mayoría de los que tenemos, se formaron por rare-mating. Aunque este tipo de híbridos fuera el predominante, encontramos dos híbridos tetraploides. AMH es un tetraploides con muy poquita contribución de *S. kudriavzevii*. La mayoría del genoma contenía cuatro copias de *S. cerevisiae* y una alta heterocigosis. Esto nos hizo plantear la hipótesis de que este híbrido proviniese de un plan de mejora de cepas comerciales en las que hicieron esporular un híbrido triploide, lo que genera esporas con contenido muy diversos, con otra *S. cerevisiae* comercial. La otra excepción es la cepa PB7. Esta es tetraploide con una contribución diploide de ambos subgenomas. Curiosamente ambos subgenomas son completamente homocigotos. Otra curiosidad es que es fértil. Todo ello indica que el cruce para formar este híbrido ocurrió entre dos esporas haploides y que posteriormente se dulicó todo el genoma. Este mecanismo es conocido y es el responsable de la recuperación de la fertilidad en híbridos, y esta es la primera vez que se observa en una cepa aislada de ambientes industriales.

Uno de las modos de generación de diversidad genómica que pudimos observar en los híbridos

era la recombinación homeóloga. Aqui desarrollamos una técnica para medir la proporción de cada subgenoma para cada gen basada en la profundidad de secuencación de cada parte. Calculando esta ratio pudimos observar que la mayoría de los cambios observados implicaba una disminución de la contribución de *S. kudriavzevii*. El aumento en proporción de la parte *S. kudriavzevii* es entonces muy rara. Estos eventos de recombinación, sin embargo, no son compartidos entre las distintas cepas, lo que hace pensar que son fenómenos aleatorios que luego son seleccionados de forma distinta. Con estos datos pudimos observar una correlación curiosa. Observamos que cuánto más cantidad de genoma estaba afectada por recombinaciones homeólogas, menos fertilidad tenía el híbrido. Sería necesario usar más cepas para poder confirmarlo, pero esto parecía indicar que la recombinación homeóloga refuerza la barrera reproductiva postzigótica al mismo tiempo que incrementa la diversidad fenotípica de la población.

Al observar que la mayoría de los híbridos naturales eran triploides y se formaron por rare-mating, quisimos ver si la estructura genómica que estos tenían se reproducía al realizar una hibridación en condiciones de laboratorio. Incubamos una cepa de *S. cerevisiae* vínica (T73) junto con una cepa de *S. kudriavzevii* europea (CR85), cada una con un plásmido de resistencia a antibióticos distintos, en un medio de rare-mating, y aislamos el híbrido resultante por doble resistencia. Secuenciamos los genomas tanto de las cepas parentales como del híbrido resultante. La cepa parental *S. cerevisiae* era diploide y heterocigota y la cepa de *S. kudriavzevii* diploide homocigota, lo que era lo esperable. El híbrido resultante era triploide con dos copias de *S. cerevisiae* y una de *S. kudriavzevii*, y heterocigosis en el subgenoma *S. cerevisiae*. La estructura era igual que los hibridos naturales, a excepción de que el híbrido artificial había perdido una copia del cromosoma III de *S.*

cerevisiae. El proceso de hibridación en laboratorio involucró la conjugación de una espora de *S. kudriavzevii*, que esporula incluso en medio rico, y una célula de *S. cerevisiae* diploide que había perdido una copia del cromosoma III, lo que la convirtió en competente para ya que solo tenía un alelo del locus MAT. Por lo tanto la estructura genómica de híbridos artificiales obtenidos por rare-mating se corresponde con la que más se observa en híbridos naturales, confirmando que ese mecanismo de reproducción es el que se usó en la naturaleza. En el tercer capítulo, estudiamos un caso particular de híbrido entre *S. cerevisiae* y *S. kudriavzevii* : la cepa VIN7. El genoma de esta cepa se reportó en varios estudios distintos. Un primer estudio alegaba que la cepa era completamente allotriploide con dos copias del subgenoma de *S. cerevisiae* y una copia de *S. kudriavzevii*. Un estudio de aCGH, aunque menos resolutivo que la secuenciación, reportaba que existían algunas recombinaciones homeólogas que generaban incremento del contenido de *S. cerevisiae* y también la pérdida de algunos cromosomas de *S. kudriavzevii*, sin reemplazamiento. La secuenciación del genoma que nosotros realizamos, a partir del mismo material que el del estudio de aCGH, mostraron que los cromosomas III, VI y XI del subgenoma de *S. kudriavzevii* tenían muy baja cobertura, sólo detectable por la alta profundidad de secuenciación que usamos. En el cromosoma VII también detectamos una recombinación homeóloga que se observaba con el aCGH. Lo que nos indicaban los experimentos era que dependiendo del material de partida que se usaba, la estructura genómica era diferente. Confirmamos este hecho usando PCR cuantitativa para detectar los cromosomas III, VI y XI de *S. kudriavzevii* en los distintos tipos de poblaciones de la cepa VIN7 y observamos que el genoma era efectivamente inestable, perdiendo ciertos cromosomas en ciertas condiciones.

Quisimos saber si la inestabilidad genómica que mostraba la cepa tenía una influencia sobre sus

características fisiológicas. Estudiamos el crecimiento de las distintas poblaciones de VIN7 a baja temperatura y bajo estrés por etanol. A baja temperatura no se observaron diferencias. Sin embargo en etanol se pudo ver que la población que había perdido los cromosomas de *S. kudriavzevii* crecía mejor que las poblaciones que los conservaban. Esto nos indicó que la inestabilidad genómica podía tener una ventaja para la adaptación en ambientes cambiantes, al permitir mayor diversidad genómica en pocas generaciones.

La inestabilidad genómica observada en VIN7 daba lugar a unas pérdidas de ciertos cromosomas. Investigamos si estos cromosomas podían tener un contenido génico que pudiese conferirles alguna característica especial y por ello esto en particular eran los afectados. Primero usamos los genomas de 1000 cepas de *S. cerevisiae* secuenciados para determinar para cada cromosoma cual era su probabilidad de encontrarse en aneuploidía, definida como el número de cepas aneuplóides para un cromosoma particular en la especie. Luego correlacionamos esta probabilidad con distintas caracterísiticas de interés de los genes que contienen los cromosomas. Estudiamos la cantidad de interacciones genéticas, interacciones físicas tanto entre genes del mismo cromosoma como entre cromosomas, la pertenencia a complejos multiprotéicos o la longitud de los cromosomas. También estudiamos si los genes de los cromosomas se sobreexpresaban en determinadas condiciones de estrés. La principal correlación que encontramos fue con la longitud del cromosoma. El resto de factores no mostraban un patrón claro que explicase las aneuploidías, menos las interacciones entre cromosomas. Lo que observamos fue que los cromosomas con mayor frecuencia de aneuploidía también eran los que menos interacciones tenían con otros cromosomas. En resumen, parece que los cromosomas más pequeños y con menos interacciones son más frecuentemente aneuploides ya

que el efecto negativo sobre la célula podría ser más neutro y por ello se observarían más.

El cuarto y último capítulo de esta tesis estudia los efectos de la evolución experimental en distintas condiciones de híbridos artificiales de *S. cerevisiae* y *S. kudriavzevii* a niveles de genoma y de transcriptoma. Estudiamos las condiciones de evolución en etanol ya que se piensa que favorecería la fijación de la parte de *S. cerevisiae*, y la evolución a baja temperatura que fijaría la parte *S. kudriavzevii*. Muchos estudios de evolución experimental de híbridos artificiales usan híbridos obtenido por el cruce de dos esporas de distintas especies, y por lo tanto estudian híbridos diploides con una copia de cada subgenoma. Como observamos en el capítulo dos, en los híbridos naturales entre *S. cerevisiae* y *S. kudriavzevii*, éste no es el mecanismo más común de hibridación, sino que es el rare-mating, y por lo tanto híbridos triploides con dos copias de *S. cerevisiae* y una de *S. kudriavzevii*. En este capítulo usamos el híbrido obtenido en el capítulo dos y lo evolucionamos durante unas 400 generaciones, repartidas entre crecimiento en medio líquido y medio sólido para aumentar la fijación de mutaciones.

Secuenciamos el genoma de los híbridos evolucionados en las distintas condiciones en los pases 4, 9 y 16 de la evolución experimental para entender qué tipo de mutaciones aparecían y en qué momento se producían. Una de las principales fuentes de variabilidad genómica que habíamos observado en híbridos naturales eran recombinaciones homeólogas. Sorprendentemente, en la evolución de híbridos artificiales no observamos cambios de este tipo. El contenido relativo de *S. cerevisiae* y *S. kudriavzevii* era estable a lo largo del experimento en todas las condiciones. Los grandes cambios que se observan es una polisomía en el cromosoma XII del subgenoma de *S. cerevisiae* que

aparece en el pase 9 de la evolución a baja temperatura. A nivel de variantes nucleotídica vimos que la mayoría de lo aparecía correspondía a pérdidas de heterocigosidad en el subgenoma de *S. cerevisiae*. Estas pérdidas de heterocigosidad afectaban a las evoluciones en etanol y la evolución control a 25ºC, pero no a la evolución a baja temperatura. La aparición de pérdidas de heterocigosidad en el subgenoma de *S. cerevisiae* y no de recombinaciones homeólogas, que se originan de la misma forma, indicarían que éstas son menos frecuentes entre las partes más divergentes del genoma. Esto es que la divergencia que existe entre los subgenomas *S. cerevisiae* y *S. kudriavzevii* haría más difícil la recombinación entre ellos, y por ello serían menos frecuentes que entre las dos copias, heterocigotas, de *S. cerevisiae*, en el tiempo del experimento.

El análisis de variantes es una herramienta ampliamente usada en bioinformática genómica. Sin embargo el trabajar con genomas complejos como son los de estos híbridos que contienen dos subgenomas divergentes pero además copias heterocigotas de uno de los subgenomas dificulta mucho el proceso y hace que sea difícil diferenciar cambios de variantes de errores de secuenciación. Por ello en este trabajo desarrollamos una aproximación que, haciendo uso de la secuenciación de los disitintos pasos de evolución, nos permitió ser conservadores en el descubrimiento de nuevas variantes para evitar falsos positivos. El procedimiento se basa en transformar la frecuencia de las variantes en “estados”. Cada variante puede estar en heterocigosis, en homocigosis presente (frecuencia cercana a 1) o bien en homocigosis ausente (frecuencia cercana a 0). Se transforman las frecuencia de cada variante en cada muestra de esta forma. Luego, las muestras se ordenan según el tiempo, es decir que se agrupan por condición de evolución y se ordenan por el momento de la evolución. De esta forma, se obtiene una matriz de estado que representa el estado de cada variante

a lo largo de la evolución en cada condición. Esta matriz se usa luego para realizar un clustering que agrupa las distintas variantes según su evolución a lo largo del experimento. Algunos tipos de comportamientos de las variantes no son plausible, lo que permite eliminarlas y depurar los datos. También permite depurar las variantes comparando el comportamiento en las distintas condiciones y de tal forma evitar una gran cantidad de falsos positivos. No solamente nos interesaba la evolución del genoma sino también cómo afectaba la evolución y las distintas condiciones de evolución al transcriptoma de los híbridos. Para ello realizamos un experimento de RNA-seq con los híbridos evolucionados a 25ºC y a baja temperatura, en los pases 4 y 16 de evolución, crecidos a baja temperatura, bajo estrés por etanol y en la condición control. La primera pregunta que nos hicimos era si algún subgenoma se sobreexpresaba globalmente más en alguna condición de crecimiento. No observamos que una condición de crecimiento ni que la evolución favoreciese la expresión de uno u otro subgenoma. Aunque *S. kudriavzevii* sea una especie mejor adaptada a las bajas temperatura, no se expresaba más esta parte del genoma en frio. Tampoco se expresaba más *S. cerevisiae* en etanol, a pesar de estar mejor adaptada a este estrés. Lo que observamos es que la condición de crecimiento era el principal factor que influía en la transcripción pero no parecía favorecer algún subgenoma diferencialmente. También observamos que los genes homólogos se expresaban según el número de copias de cada subgenoma y que las aneuploidías y los cambios en números de copia de los cromosomas era un factor primordial para entender la transcripción global.

La segunda pregunta que nos hicimos era si la evolución cambiaba el transcriptoma en la misma condición. Vimos que únicamente la evolución a baja temperatura cambiaba el transcriptoma de los híbridos. El transcriptoma de los híbridos evolucionados a 12ºC era distinto de los que no

estaban evolucionado, en el experimento de crecimiento a 12ºC. Realizamos un análisis de expresión diferencial entre estos dos híbridos en esta condición y encontramos diferencias significativas. Observamos que aquí tampoco se veía favorecido un subgenoma sobre el otro. Un número aproximadamente igual de genes estaban expresados diferencialmente en ambos subgenomas. Sorprendentemente, de ellos, pocos eran compartidos por ambos subgenomas. Hicimos un análisis de enriquecimiento de términos GO con los genes y vimos que en los híbridos evolucionados se sobreexpresaban genes relacionados con la traducción y se reprimían genes relacionados con la respiración. Estos procesos ya se habían relacionados con dianas para la mejora del crecimiento a bajas temperaturas en ambas especies pero nunca en híbridos. Clasificamos los genes según el subgenoma del que provenían y volvimos a hacer el análisis de enriquecimiento de términos GO con los que eran específicos de un único subgenoma. Curiosamente las categorías funcionales enriquecidas eran las mismas aunque los genes fueran distintos. Los análisis realizados apuntan a que en estos híbridos lo que se seleccionó en la evolución fueron los procesos y no genes concretos. En cada subgenoma genes distintos se vieron implicados pero el sistema que mejoraba el fenotipo y que se seleccionó durante la evolución era el mismo.

De esta investigación pudimos extraer varias conclusiones. La ploidía y las aneuploidías son factores importantes en la evolución y la adaptación de *S. cerevisiae*. Confirmamos que la aneuploidía en el cromosoma III era un mecanismo adaptativo contra el estrés por etanol. Los análisis filogenómicos de híbridos entre *S. cerevisiae* y *S. kudriavzevii* determinaron que se originaron por varios eventos de hibridación independientes. La ploidía de estos híbridos era el resultado directo del mecanismo usado para la conjugación. La alotriploidía era la composición genómica más

común de los híbridos y esta sólo se podía obtener por rare-mating. PB7 era un híbrido tetraploide, completamente homocigoto y fértil. Esta composición se explica por conjugación de dos esporas y una posterior duplicación del genoma. La recombinación homeóloga era un importante mecanismo para generar diversidad genómica y fenotípica en estos híbridos pero al coste de aumentar la infertilidad. La inestabilidad genómica apareció como un mecanismo importante de generación de diversidad genómica en un híbrido de *S. cerevisiae S. kudriavzevii*. La evolución experimental de híbridos artificiales reveló la importancia de la ploidía para el acceso a mecanismos de evolución genómica. La transcripción también se vio afectada por la ploidía de los híbridos enfatizando que esta característica es primordial en el funcionamiento de los híbridos. La baja temperatura es la única condición en la que se observó un cambio adaptativo a lo largo de la evolución experimental de los híbridos. Este cambio adaptativo afectó a ambos subgenomas de forma diferente a nivel de genes pero afectando a los mismos procesos. Esto evidenció que la selección y la adaptación ocurre a nivel de sistema y no a nivel de genes específicos.

Introduction

On yeast ecology and evolution

Fungi is an eukaryote kingdom that was historically included in plants. The reason was that they are immobile and the presence of cell wall and vacuole. An important difference is that fungi are heterotrophic, they need the intake of carbon from other sources to produce energy, a character they share with animals. Phylogenetic analysis confirmed that fungi were more similar to animals than to plants (Baldauf and Palmer, 1993). The fungi clade is highly diverse and is found in all kinds of environments, even if they are overrepresented in terrestrial environments. They are key organisms in the recycling of nutrient (mushrooms), live in symbiosis with other organisms (lichens) and even live as a parasite of plants or animals.

Among Fungi, a particularly interesting and well-studied group is yeasts. It is defined by yeast, unicellular fungi that divide whether by fission or budding and has sexual states but do not form

fruiting bodies (Kurtzman and Sugiyama, 2015). Fungi access nutrients through the secretion of hydrolytic enzymes and make an extensive use secondary metabolisms to live. Nevertheless, yeasts have a reduced secondary metabolism (Krause et al., 2018) making their life-style characteristic among the clade (Naranjo-Ortiz and Gabaldón, 2019). They, therefore, live in competitive environments where the ability to quickly intake easily accessible nutrients is the key to survive. That makes them direct competitors of bacteria.

The best–studied group of yeast is the *Saccharomycotina* clade. This is a highly diverse group with a wide genetic diversity which is comparable to that of plants or animals (Shen et al., 2018) and includes human-relevant species such as the baker yeast *Saccharomyces cerevisiae* or the human pathogen *Candida albicans*. One trait that makes the clade interesting to geneticists is that there is a tendency to have compact genomes. The genome size is usually between 8 to 20 Mb, have high gene density (between 46 and 75% of the genome) and only from 3 to 14 % of the genes contain introns (Naranjo-Ortiz and Gabaldón, 2019; Dujon and Louis, 2017; Dujon, 2015) . This low genomic complexity eases the bioinformatics analysis and converts the clade into an excellent model for genomic studies.

Saccharomyces cerevisiae is among the best known genetic model. It was the first eukaryote genome to be sequenced (Goffeau et al., 1996) and its relevance for biotechnology or even biomedical research is undoubted. Even if *S. cerevisiae* is the most studied species, the interest in the complete *Saccharomyces* genus is increasing. The genus is composed of 8 species : *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. jurei*, *S. kudriavzevii*, *S. arboricola*, *S. eubayanus* and *S. uvarum*

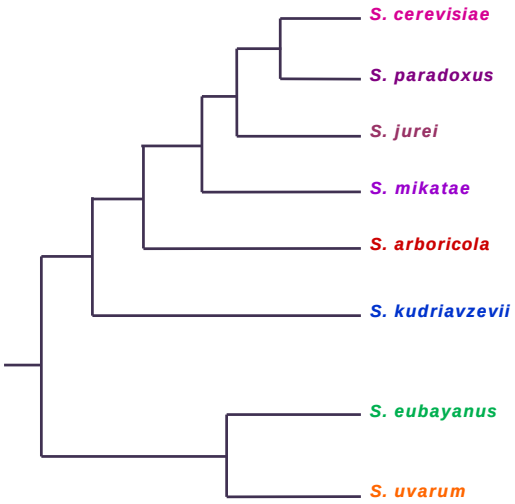

FIGURE 1 Cladogram of the species of the *Saccharomyces* genus.

(Figure 1). Even if the genomes of all of them are around 12Mb and mostly syntenic, the sequence divergence between them is huge. It can be as high as between humans and chicken (between *S. cerevisiae* and *S. kudriavzevii*) or as similar as between human and macaque (between *S. uvarum* and *S. eubayanus*) (Dujon, 2006; Hittinger, 2013) .

The habitat of the different species differ. *Saccharomyces cerevisiae* is the principal species isolated from fermentative processes. Nevertheless, it was isolated from other sources such as fruit, soil or oak trees. However, except in industrial processes, its abundance is normally low, what opened the hypothesis that it is a nomadic species (Goddard and Greig, 2015). The industrial strains of *S. cerevisiae*, though, have marks of domestication and clear non-domesticated linages exist in other habitats (Peter et al., 2018; Duan et al., 2018; Liti et al., 2009; Barbosa et al., 2016; Gallone et al., 2016). The recent population genomics studies on thousands of isolates from different sources marked the origin of the species in Asia and followed by a further expansion and

several independent domestication events (Peter et al., 2018; Duan et al., 2018). Apart from *S. cerevisiae*, only *S. uvarum* was isolated from wine or cider fermentations (Demuyter et al., 2004; Naumova et al., 2000; Rodríguez et al., 2014). The other species inhabit natural habitat. The most commonly found is *S. paradoxus* which is found associated with oak trees (Naumov et al., 1998). Also, *S. uvarum* and *S. kudriavzevii* and *S. mikatate* were isolated in Europe and Asia associated with oak trees. *S. eubayanus* was not isolated in Europe, it was found related to *Nothofagus* trees in Patagonia and also Asia. The last discovered species is *S. jurei* which only was found in Europe nowadays (Naseeb et al., 2017). The species of the genus *Saccharomyces* also show phenotypic differences between them. Some species are used in industry due to the production of different by-products or aromas (Pérez-Torrado et al., 2017; Pérez-Través et al., 2016). Though, the most notable difference between species is their temperature tolerance (Salvadó et al., 2011; Gonçalves et al., 2011). *Saccharomyces kudriavzevii*, *S. uvarum* and *S. eubayanus* grow at low temperature, being *S. kudriavzevii* the more cryotolerant. The highest temperature tolerant in the genus is *S. cerevisiae*.

Saccharomyces yeast live most of the time as diploid cells dividing by mitosis (Figure 2). However, they also have a sexual cycle. Sex is codified in the *MAT* locus. Located on chromosome III. Diploid cells have two alleles of this locus, or two mating-types, named *MAT* a and *MAT*α. When life conditions are harsh, due to low nutrient availability and particularly low nitrogen availability, cells enter meiosis. They then produce an ascus with four spores (two *MATa* and two *MAT*α). To mate and return to the diploid state, spores need to find a spore of their opposite mating-type. From this point, different outcomes are possible. Spores from the same ascus can mate between them to

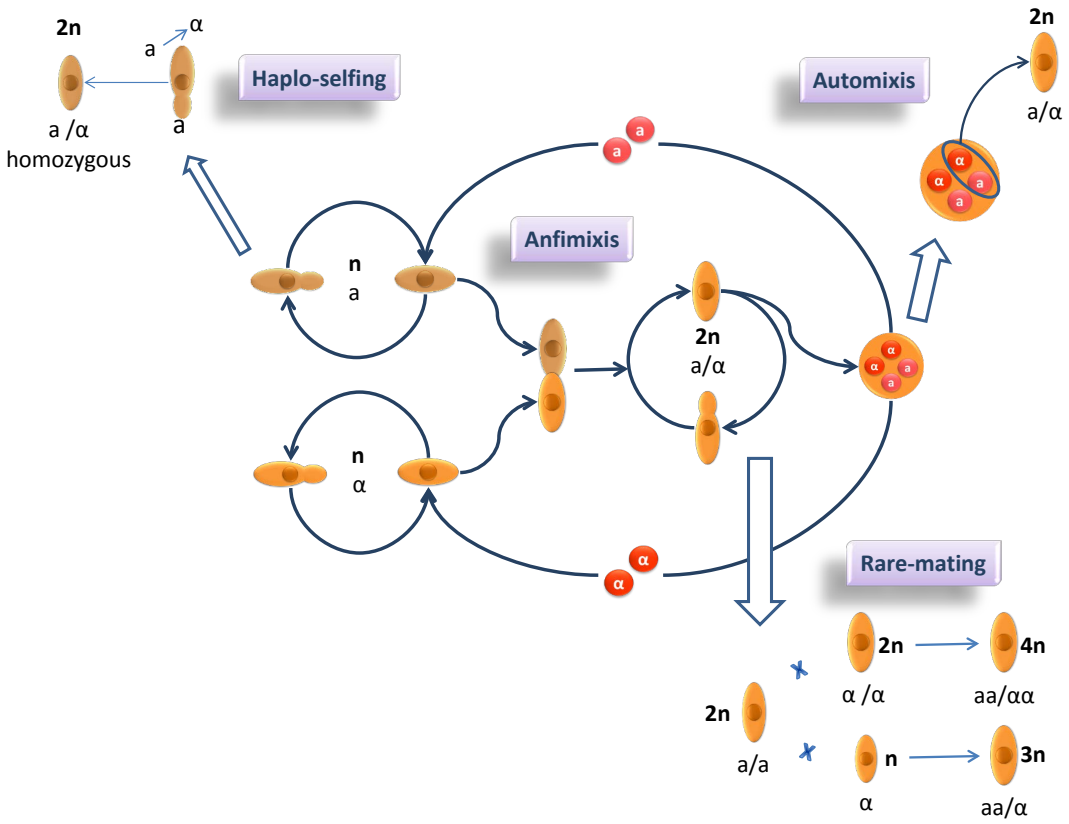

FIGURE 2 Life cycle of *Saccharomyces* yeasts.

form a diploid (automixis). Automixis is calculated to occur in 94% of the cases cells use sexual cycle (Hittinger 2013). In 4% of the cases, they use Haplo-selfing. This makes use of the mating-type switching process. It begins when the endonuclease codified at the *HO* locus cuts a specific site of the *MAT* locus. Then this site is degraded by an exonuclease and the DNA reparation machinery replaces one copy of the *MAT* locus by one of the silenced copies of the locus that are present at both ends of the chromosome, changing the mating type of the cell. All this process occurs while the haploid spore is dividing, and it changes the mating type of the mother cell, while the daughter cell remains from the original mating type. Then the daughter cell and the mother cell mate, forming a completely homozygous diploid cell, except for the heterozygous *MAT* locus. Interestingly the mating between spores, outcrossing or amphimixis, from different origin only occurs in 1% of

the cases. Finally, another way to mate is possible in *Saccharomyces*. If the heterozygosity of *MAT* locus is lost, whether by DNA repair mechanisms after a DNA double-strand break or by the loss of one copy of the chromosome III, then even a diploid cell becomes competent to mate with another cell of opposite mating-type. This is called rare-mating and can generate different resulting cell types, from triploids if a diploid and a haploid mate, to tetraploid if two diploids mate.

On generating genome diversity

"*One of the most striking features of genome structure is its lability.*" (Otto, 2007) Many kinds of changes can occur in the genome structure. With the advent of genome sequencing technologies, the detection of such changes became easier. Therefore the attention paid to mechanisms such as Loss of Heterozygosity (LOH), translocations, aneuploidy or polyploidy is increasing. In fungi, and *Saccharomyces* many of these are observed.

In heterozygous yeasts, long traits of homozygous regions are observed (Magwene et al., 2011; Gallone et al., 2016; Peter et al., 2018). Loss of heterozygosity can arise by different mechanisms in mitotically dividing cells. Short-range LOH are products of gene conversion or double-chromosome cross over. However most of the observed LOH begin from a specific site of the chromosome until the end and these are produced by break-induced replication (for a review Llorente et al. (2008)). LOH can increase the phenotypic diversity of a population as it fixes heterozygous alleles and therefore can affect recessive alleles. In *Candida albicans*, for example, LOH is more

frequent than point mutations and increases upon stress exposure (Forche et al., 2009). It is also related to antifungal resistance in clinical isolates of this species (Coste et al., 2006). In *Saccharomyces cerevisiae*, it was also recently demonstrated that LOH is an important adaptive mechanism (James et al., 2019).

Another mechanism that generates diversity in populations are chromosomal rearrangements. An example of it in *Saccharomyces cerevisiae* is the translocation increasing sulfite resistance in wine strains (Pérez-Ortín et al., 2002). In this case, the translocation between the chromosome VIII and XVI changed the promotor of the gene *SSU1*, which encodes for a sulfite efflux pump, increasing its expression and therefore increasing sulfite tolerance in wine strains. This is an example of adaptive translocation as wine strains are exposed to sulfite and need it to survive. The detection of such structural variants are harsh with small read sequencing technology as the read length do not cover sufficient length of the break between chromosome to doubtlessly recognize them. Long read sequencing such as PacBio or Oxford Nanopore will promptly shed light on these events. As an example such technology permitted to detect an inversion in chromosome XVI which, as happened with the translocation, enhanced sulfite resistance of the strain (García Ríos et al., 2019).

Horizontal gene transfer is a major source of evolvability in bacteria. In Eukaryotes, however, the topic remains controversial with discordant voices (Martin 2017). Nevertheless, in fungi, plenty of examples of it can be found (for a review (Feurtey and Stukenbrock 2018)). Even genetic material exchange between bacteria and fungi was found (C. Gonçalves et al. 2018; Husnik and McCutcheon 2017). In *Saccharomyces*, the most famous example is the acquisition of a large

genomic region from *Torulaspora microellipsoides* in *S. cerevisiae* wine strains. This region conferred to the strains that maintain it a competitive advantage on grape must fermentation (Novo et al. 2009; Marsit, Mena, et al. 2015; Marsit, Sanchez, et al. 2015).

The number of sets of chromosomes a cell contains is named ploidy. Species tend to have stable and defined ploidy. The ploidy state can change in the course of the life cycle, for example, the spores of *Saccharomyces* are haploid and the cell diploid. However the ploidy state can change and this change has important consequences on fitness and evolvability (Marad et al., 2018; Fisher et al., 2018; Scott et al., 2017; Todd et al., 2017; Berman, 2016). Aneuploidy occurs when a chromosome (or chromosomes) have an unbalanced copy number. This state can be strongly detrimental, for example in Down syndrome in humans. In fungi, however, it is intriguingly tolerated and it can even be advantageous. An extreme example of it is the *Cryptococcus neoformans* "Titan" cells and their progeny. In response to stress, this species produces gigantic cells with a ploidy of up to 200N (normal cells are diploids). These cells then produce haploid offspring that harbor different aneuploidies. These aneuploids ensure the survival of part of the offspring in stress conditions (Gerstein et al., 2015). In *Saccharomyces*, many aneuploids and polyploids were detected in different populations (Peter et al., 2018; Gallone et al., 2016). Modifying the number of copies of a chromosome changes its expression and increasing evidence is accumulating about the relevance of this mechanism to increase stress resistance (Zhang et al., 2016; Yona et al., 2012; Gorter et al., 2017; Gilchrist and Stelkens, 2019).

On hybridization and species concept

Species is a fundamental concept in biology. However, depending on the criteria used, species can de defined in different manners. The biological species concept focuses on reproduction. It defines species as a group of organisms that interbreed and are reproductively isolated from other groups (Mayr, 1970). Therefore, two organisms are considered to belong to different species if the offspring from the cross between them is not fertile. The application of the biological concept of species is then incompatible with asexual species, or when reproduction testing is difficult. Therefore, other ways to define species emerged such as the ecological concept, which considers species as organisms sharing a particular niche that exclude others. A particularly interesting concept is the phylogenetic concept of species. This concept considers a species as an independently evolving lineage, or metapopulation, that has a common ancestor that excludes other species (Wiley, 1978; Mallet, 1995; De Queiroz, 2007). This concept, therefore, focuses the definition of species on that species share a unique evolutionary history, being then less restrictive than the biological concept as sexual reproduction is not needed. In *Saccharomyces* yeasts, both the biological and phylogenetic concepts of species are suitable as they have sexual reproduction.

The biological species concept implies that, if effective reproduction between species is not possible, mechanisms that isolate them and prevent breeding exist. Such mechanisms, reproductive barriers, can be prezygotic, that act making impossible the cross between species, or postzygotic, when breeding is possible then the resulting offspring is not fertile, converting the result of the

cross as a dead end. In *Saccharomyces*, the reproductive isolation is mainly postzygotic (**?**). At the genomic level, several mechanisms were postulated as reducing the hybrid fertility and being responsible for the species isolation. Karyotypic changes between species can result in missegregation of chromosomes and abnormal recombinations during meiosis and therefore the formation of inviable spores. As explained above, the genome of the *Saccharomyces* species is mostly syntenic and therefore this mechanism is thought to have a small contribution to isolation, even if some cases were observed (Leducq et al., 2015; Marsit et al., 2017). Another mechanism proposed to affect hybrid fertility is the action of the mismatch repair system. As sequence divergence between species is high this would activate the mismatch-repair system which reduces recombination rates between homeologues causing aneuploidies during meiosis and reduction of fertility (Liti et al., 2006; Chambers et al., 1996). Genetic interactions are keys to multiple processes. The diverging evolution of these interactions, or genes involved in interactions, can result in strong negative interactions, without the need of higher divergence between the rest of the genome, reducing viability or fertility of hybrids. These are called Bateson-Dobzhansky Muller incompatibilities and, in yeast, it was shown that such incompatibilities can evolve rapidly (Anderson et al., 2010) and also exist between different species of *Saccharomyces* especially affecting mito-nuclear interactions (Chou and Leu, 2010; Špírek et al., 2015). Finally, it was recently proposed that hybridization between species increases genome instability and is fundamental to reproductive isolation (Dion-Côté and Barbash, 2017). Genome instability is complex and encompasses some of the mechanisms cited above, and can be of importance, as mentioned, in the adaptability in asexual populations.

Even if boundaries or barriers exist between species, it is recognized that these are to some

extent permeable. Multiple examples in plants (Goulet et al., 2017; Yakimowski and Rieseberg, 2014), animals (Abbott et al., 2013) or even in humans (Ackermann et al., 2019) show that hybridization between species happen. It not only exists but it can even have unexpected outcomes. Hybridization between species can be a way to improve crops for example due to heterosis (Barton, 2001; Govindaraju, 2019). Heterosis or hybrid vigor is the result of an improved phenotype of the hybrid compare to its parents. Using hybridization to take advantage of heterosis for adaptation to a new environment is then an interesting mechanism (Barton, 2001). In *Saccharomyces*, sexual reproduction is facultative, so hybrids can propagate asexually even if they are not fertile. Examples of hybridization in the genus are recurrent. Introgressions,small fragments of genome replaced by the homologous sequence from a distinct species as a trace of an old hybridization event, were found between different species of the genus (Albertin et al., 2018; Barbosa et al., 2016; Almeida et al., 2017). Also, complete hybrids were found in industrial environments. The most well-known are hybrids between *S. cerevisiae* and *S. eubayanus* in lager beer fermentations (Monerawela and Bond, 2016; Wendland, 2014). But hybrids between *S. cerevisiae* and *S. kudriavzevii* or *S. cerevisiae* and *S. uvarum* and even hybrids formed by three different species (Peris, Pérez-Torrado, Hittinger, Barrio and Querol, 2017; Sipiczki, 2018*a*; Lopandic et al., 2018). Hybrids are formed between *S. cerevisiae* and a cold-tolerant sister species that pointed to hybridization in the genus as a domestication mechanism in industrial environment in colder climates (Boynton and Greig, 2014; Peris, Pérez-Torrado, Hittinger, Barrio and Querol, 2017).

Objectives

Industrial *Saccharomyces* yeasts are highly specialized organisms that have evolved to utilize the different environments provided by human activity. This process can be described as unconscious domestication and is responsible for their particular physiological characteristics.

With the advent of the genome sequencing technology, an important effort is devoted to understanding the molecular mechanisms involved in the adaptation of *Saccharomyces* yeasts to the industrial processes. Among these mechanisms, genome structure variations due to chromosomal rearrangements, aneuploidy, and allopolyploidy due to interspecies hybridization, are of increasing interest due to their adaptive relevance to explain the particular properties of these industrial yeasts. Accordingly, in the present doctoral thesis, I focus on the following objectives:

1) The analysis of genomic differences between *S. cerevisiae* strains, exhibiting different ethanol tolerances, to determine the relevance of chromosomal variations in their physio-

logical differences.

2) The comparative genome analysis of natural *S. cerevisiae* x *S. kudriavzevii* hybrids to unveil the mechanisms of hybridization and how hybridization influences their physiological properties, of biotechnological interest.
3) The analysis of the factors influencing genomic instability in *S. cerevisiae* x *S. kudriavzevii* hybrids.
4) The study of the early evolution of artificial hybrids to understand their responses, at the genome and transcriptome levels, to selective pressures due to cold and ethanol stresses, present in fermentative environments.

The thesis was organized into four chapters :

Chapter 1, *Aneuploidy and ethanol tolerance in Saccharomyces cerevisiae* corresponds to objective 1.

Chapter 2, *Genome structure reveals the diversity of mating mechanisms in Saccharomyces cerevisiae x S. kudriavzevii hybrids, and the genomic instability that promotes phenotypic diversity* aims to disentangle objective 2.

Chapter 3, *VIN7 : a case study of instability in interspecific hybrids between S. cerevisiae and S. kudriavzevii* addresses objective 3.

Chapter 4, *Genomic and transcriptomic analysis of the short-term evolution of artificial hybrids*

between S. cerevisiae and S. kudriavzevii deals with objective 4.

CHAPTER 1

Aneuploidy and ethanol tolerance in *Saccharomyces cerevisiae*

1.1 Introduction

The yeast *Saccharomyces cerevisiae* is among the most beneficial microorganisms for humans, especially industrial strains involved in the production of fermented products, such as bread, beer or wine. *S. cerevisiae*, as well as other *Saccharomyces* species, are characterized by their ability to ferment simple sugars into ethanol, even when oxygen is available for aerobic respiration (Crabtree effect), due to an overflow in the glycolysis pathway (Hagman and Piškur, 2015). Although alcohol fermentation is energetically less efficient than respiration, it provides a selective advantage to these yeasts to out-compete other microorganisms. This way, sugar resources are consumed faster and the ethanol produced during fermentation, as well as high levels of heat and CO2, can be harmful or less

This chapter is published in Morard et al. (2019) *Frontiers in Genetics* .

tolerated by their competitors. Once competitors are overcome, *Saccharomyces* yeasts can use the accumulated ethanol as a substrate for aerobic respiration in the presence of oxygen. This ecological strategy was named (ethanol) 'make-accumulate-consume' (Thomson et al., 2005; Piškur et al., 2006).

With the advent of the human hunter-gatherer societies, *S. cerevisiae*, due to its fermentative capabilities, successfully occupied a new ecological niche in the crushed grape berries, collected by humans to produce the first fermented beverages. With agriculture, Neolithic societies improved fermentations as a way to preserve their foods and beverages. Since then, human-associated *S. cerevisiae* yeasts have been exposed to selective pressures due to fluctuating stresses occurring during fermentations, such as osmotic stress, ethanol toxicity, anaerobic stress, acid stress, nutrient limitation, etc. (Querol et al., 2003). As a result of this passive domestication, human-associated *S. cerevisiae* yeasts exhibit differential adaptive traits and conform genetically separated populations (Gallone et al., 2016; Duan et al., 2018; Legras et al., 2018; Peter et al., 2018), according to their sources of isolation rather than to their geographic origins.

One of the most important selective pressures imposed to *S. cerevisiae* is ethanol tolerance. High ethanol concentrations also have a strong effect on *S. cerevisiae* yeast growth and metabolic efficiency (Ansanay-Galeote et al., 2001). Ethanol is a small amphipathic alcohol that can cross through cell membranes, increasing their fluidity and permeability, interfering in the folding and activity of proteins, and also affecting intracellular redox balance and pH homeostasis (reviewed in Auesukaree, 2017).

It is sometimes hard to differentiate between tolerance and resistance because they are defined in different ways depending on the research field. Most of the literature related to ethanol stress uses both concepts as synonymous to refer to the ability of yeasts to grow and survive in the presence of ethanol, although 'ethanol tolerance' is the most frequently used (Snoek et al., 2016). In an attempt to differentiate these terms in microbiology, Brauner et al. (2016) defined resistance as the ability of a microorganism to grow in the presence of high concentrations of a drug, resulting in a higher minimal inhibitory concentration (MIC), and tolerance as the ability of the cell to survive to the transient presence of a drug above the MIC. As ethanol is the main product of the *Saccharomyces* respire-fermentative metabolism, and, as mentioned, the basis of the 'make-accumulate-consume' strategy, *Saccharomyces* yeasts acquired mechanisms to survive the transient presence of ethanol, and hence, we consider that the term 'ethanol tolerance' would be more appropriate.

Different studies have been devoted to understand the molecular mechanisms responsible of yeast response and tolerance to ethanol (for a review see Snoek et al., 2016). However, ethanol tolerance is a multilocus trait, not well characterized, because genes related to ethanol tolerance are broadly distributed throughout the genome (Giudici et al., 2005). In fact, as many different cellular processes are affected by ethanol, more than 200 genes have been linked to ethanol tolerance. Therefore, although many efforts have been made, mechanisms of ethanol tolerance are not fully understood yet.

In recent years, researchers have looked at adaptation to different stresses (Yona et al., 2012; Voordeckers and Verstrepen, 2015; Adamczyk et al., 2016), including ethanol, in non-tolerant yeast

exposed to gradually increasing stress levels. An interesting outcome of these experiments was the fixation in yeast of different genome rearrangements of adaptive value (Gorter de Vries et al., 2017).

In a previous study, we determined significant differences in ethanol tolerance between natural and fermentative *S. cerevisiae* strains, including strains isolated from different sources, from wine to traditional fermentations of Latin America (Arroyo-López et al., 2010). In the present study, we have sequenced the genomes of the most and least ethanol-tolerant *S. cerevisiae* strains reported in Arroyo-López et al.'s study to determine if they differ in their chromosomal constitution.

1.2 Materials and Methods

1.2.1 Strains and sequencing

The *S. cerevisiae* strains used in this study are those exhibiting extreme differences in their ethanol tolerance (Arroyo-López et al., 2010). Temohaya-MI26 was isolated from the fermentation of Mezcal production in Durango, Mexico, and shows the lowest ethanol tolerance. Wine strain T73 was selected as a commercial dry yeast Querol et al. (1992). It was isolated from a red wine fermentation in Alicante, Spain, and possesses an intermediate ethanol tolerance. Finally, strains CECT10094 and GBFlor-C are flor strains isolated from red Pitarra wine in Extremadura, Spain, and González Byass Sherry wine (Esteve-Zarzoso et al., 2001) in Jérez de la Frontera, Spain, respectively. They both exhibit the highest ethanol tolerances. Yeast cells were grown in an overnight culture of GPY in 5 ml. Cells were pelleted in a microcentrifuge and suspended in 0.5 ml of 1 M

sorbitol-0.1 M EDTA, pH 7.5. Then, they were transferred to a 1.5-ml microcentrifuge tube, with 0.02 ml of a solution of Zymolyase 60 (2.5 mg/ml). A microcentrifuge was used to spin down cells for 1 min, which were suspended in 0.5 ml of 50 mM Tris-HCl-20 mM EDTA, pH 7.4. After suspension, 0.05 ml of 10% sodium dodecyl sulfate was added and the mixture was incubated at 65°C for 30 min. Then, 0.2 ml of 5 M potassium acetate was added and the tubes were placed on ice for 30 min. Then they were centrifuged at maximum speed in a microcentrifuge for 5 min. Supernatant was transferred to a fresh microcentrifuge tube, and the DNA was precipitated by adding one volume of isopropanol. After incubation at room temperature for 5 min, the tubes were centrifuged for 10 min. The DNA was washed with 70% ethanol, vacuum dried, and dissolved in 50 µl of TE (10 mM Tris-HCl, 1 mM EDTA, pH 7.5). T73 was sequenced with a 300-bp paired-end library in an Illumina HiSeq 2500 equipment. Strains CECT10094, GBFlor-C and Temohaya-MI26 were sequenced with paired-end libraries of 100 bp with a mean insert size of 300 bp in an Illumina HiSeq 2500 instrument. EC1118 sequencing data was downloaded from NCBI with identifiers: SRA, ERS484054; BioSample SAMEA2610549.

1.2.2 Genomes assembly and annotation

Reads were trimmed with Sickle v1.2 (Joshi and Fass, 2011) with a minimum quality value per base of 28 at both end and filtered at a minimum read length of 85 bp. A first preassembly step was carried out with Velvet v1.2.03 (Zerbino and Birney, 2008) to determine the best k-mer value for each library. The assembly was done with Sopra v1.4.6 (Dayarian et al., 2010) integrated with Vel-

vet with the k-mer value determined in the previous step. Then, refinement of the results was carried out with SSPACE v2.0 (Boetzer et al., 2011) and GapFiller v1.11(Boetzer and Pirovano, 2012) to improve scaffold length and remove internal gaps. Several rounds of Sopra/SSPACE/GapFiller were performed until the number of scaffolds could not be reduced. At each step of the process, the scaffolds were aligned against the reference genome of *S. cerevisiae* S288C with Mauve v2.3.1 (Darling et al., 2010). These steps can lead to overfitting and the nature of our sequence data mean we cannot verify any new recombination events, so, they were manually corrected. The final scaffolds were then aligned against the S288C reference genome with MUMmer v3.07 (Kurtz et al., 2004) and ordered into chromosomes with an in-house script.

Genomes annotation was done using two different strategies. First, the annotation from S288C genome was used to transfer to the new genomes by sequence homology using RATT (Otto et al., 2011). Second, an *ab initio* gene prediction was performed using Augustus web server (Stanke and Morgenstern, 2005), to complete the annotation of low homology regions where RATT was not able to transfer annotation. Both results were merged and the annotation was then manually corrected using Artemis (Rutherford et al., 2000) to remove false gene discovery and incorrect RATT transfer were either removed or corrected dependent on the nature of the mistake (e.g. wrong placement, lack of intron, etc.).

1.2.3 Variants detection and chromosome copy number analysis

Mappings against the reference *S. cerevisiae* S288C genome (version R64-2-1) were done using bowtie2 v2.3.0 (Langmead and Salzberg, 2012) with default parameters. Read Depth (RD) for each position was computed with bedtools v2.17.0 Quinlan and Hall (2010). To smooth the representation of RD by chromosome, a sliding windows analysis was performed. Mean mapping reads was calculated for 10kb windows moving by 1000nt. Variant calling analysis was performed with breseq v0.27.1 (Deatherage and Barrick, 2014) pipeline with polymorphisms mode to enable heterozygotic variants to be called. Minimum polymorphism frequency was set to 0.15 to avoid low frequency variants calling. Variants annotation and manipulation was done with gdtools v0.27.1 from breseq package. Variants whose frequency was higher than 0.95 were considered homozygotic and they were considered heterozygotic if it was lower. R and ggplot2 package were used for data representation.

1.2.4 Phylogenetic analysis

All gene sequences were extracted from the annotations of the genomes assembled, as well as sequences from 38 strains representatives of different known clades (Supplemental table S1). Orthologous genes among *S. cerevisiae* strains were translated and aligned with MAFFT v7.221 (Katoh and Standley, 2013). Then the alignments were back translated to nucleotides, and concatenated. Maximum Likelihood phylogeny was performed on the concatenated genes alignment

with RAxML v8.1.24(Stamatakis, 2014) with model GTR-Γand 100 bootstrap replicates. The concatenated-gene ML tree was drawn with R and ggtree package (Yu et al., 2017).

1.2.5 Determination of the ploidy by flow cytometry

The total DNA content of the strain of interest was estimated by flow cytometry analysis in a BD FACSVerse cytometer following the SYTOX Green method as described in Haase and Reed (2002). Ploidy levels were scored on the basis of fluorescence intensity compared to the reference haploid S288c and diploid FY1679 *S. cerevisiae* strains. The estimated ploidy of the strains was obtained from three independent measurements.

1.2.6 Expression analysis

The expression data from a previous work on ethanol response (Navarro-Tapia et al., 2016) was used in this study (GEO accession : GSE44863). In brief, transcriptomic analysis come from a microarray analysis after ethanol shock. Temohaya-MI26 and CECT10094 were subjected to a 10% ethanol treatment and RNA was extracted 1 and 10 hours later. As a control, RNA was extracted after 1 and 10 hours of growth without ethanol treatment. Samples were hybridized for each condition against a pool of all the samples from all the conditions in the analysis. Expression data is reported as the log2 of the ratio of signal intensities between each condition and the pool. After combining each replicate, the genes were assigned to chromosomes according to their systematic names. Wilcoxon-Test implemented in ggpubr package v0.1.7 was used for the statistical analysis

of differences between the expression of the chromosomes.

1.2.7 Aneuploidy analysis in *S. cerevisiae* strains

Ploidies, aneuploidy presence, and 15% ethanol tolerances were extracted from the recent study of 1011 *S. cerevisiae* genomes (Peter et al., 2018). The 1011 strains were grouped by ploidy level and the presence of chromosome III copy number variations. Wilcoxon paired test was used to test differences between euploid diploids and the rest of ploidies and aneuploidies. Relative growth rates are represented as the normalization of the ratio between growth on standard YPD at 30ºC and the stress condition (Peter et al., 2018).

1.2.8 Yeast chromosome III removal

A counter selectable marker (Kutyna et al., 2014) was used to remove a single copy of chromosome III in *S. cerevisiae* 2-200-2 strain (Voordeckers et al., 2015) to obtain a derivative strain with one less copy of chromosome III, named as 2-200-2-S4. An integrative cassette targeted to a wide intergenic region (YCR027C-YCR028C) was synthetized from pCORE5 vector using the following primers: CHRIIIdel_F: CTGTAGCCATATTAAATTCCTTTGTCTCTGGACTCTTTC-GAGCCCCCGATTTAGAGCTT and CHRIIIdel_R: TTAACGTTCAAGCAGCGTCAGTGA-GAACTAAAATCATCCAATCTCGAGGTCGACGGTATCGAT. The 2-200-2 strain was transformed and colonies were selected in GPY with G418. Correct integration was corroborated by PCR using the following primers: test-CHRIII del_F: TCGACATCATCTGCCCAGAT and test-

CHRIII del_R: ACTTAGGTGGAGGAGCAAG. After overnight growth in GPY (2% glucose, 0.5%, peptone, 0.5% yeast extract), cells were plated on galactose counter selection media (2% galactose, 0.5% peptone, 0.5% yeast extract), and colonies were used to measure chromosome III copy numbers.

1.2.9 Chromosome III copy number measurements

Genomic DNA was isolated and ethanol precipitated from the GPY liquid cell suspension in five independent culture replicates of 2-200-2 and 2-200-2-S4. DNA purity and concentration were determined in a NanoDrop ND-1000 spectrophotometer (Thermo-Scientific), and the integrity of all samples was checked by electrophoresis in agarose gel (1%). The PCR primers used to study the chromosome III copy number were designed from the available genomic sequence of *S. cerevisiae* strain S288C (R64-2-1, *Saccharomyces* genome database, http://www.yeastgenome.org). The sequences of PCR primer pairs used in this study are: ARE1-F: CCTCGTGTACCAGATCAAC; ARE1-R: AGGAAGATGGTGCCAATGAT; YCL001W-A-F: TGCTACGGTGGTTCTGCAAG; YCL001W-A-R: ACCACTGTGTCATCCGTTCT; POF1-F: TAATGGAGAGCTTCATGTCGGG; POF1-R: CCCTCAAGGATGTCACTGGC; ACT1_F: ATGTTCCCAGGTATTGCCG; ACT1_R: GCCAAAGCGGTGATTTCCT; YFR057W-F: ACACCGCCAAGCTTCCAATA; YFR057W-R: TTGCCACGCAAAGAAAGGAC; ACT1_F: CATGTTCCCAGGTATTGCCG; ACT1_R: GCCAAAGCGGTGATTTCCT; YFR057W-F: ACACCGCCAAGCTTCCAATA and YFR057W-R: TTGCCACGCAAAGAAAGGAC. Primers were

designed to get amplicons of 100-200 bp in size to ensure maximal PCR efficiency, and the accuracy of quantification. PCR amplification was performed in a 10-µL final volume that contained 2.5 µL of the DNA template, 1.5 µL MilliQ water, 0.2 µM of each primer, and 5 µL of LightCycler 480 SYBR Green I Master (Roche). Reactions were performed in 96-well plates in an LightCycler 480 (II) PCR amplification and detection instrument with an initial denaturalization step at 95°C for 5 min, followed by 45 cycles of 95°C for 10 s, either 53 or 54°C for 10 s and 72°C for 4 s.. A melting curve analysis was included at the end of each amplification program to confirm the presence of a single PCR product of all the samples with no primer-dimers. The Advanced Relative Quantification program v.1.5.1, implemented in the LightCycler 480, was used to analyze the results, and the efficiency of all the primer pairs was previously determined and included in the analysis. Normalization of the quantification results of genes ARE1, YCL001W-A and POF1 was performed using the levels of genes ACT1 and YFR057W as reference genes.

1.2.10 Ethanol tolerance assays by drop test experiments

Drop test experiments were carried out to assess strains 2-200-2 and 2-200-2-S4 ethanol tolerances. Rectangular GPY plates supplemented with different ethanol percentages (0, 6, 10, 14, 16 and 18%) were prepared. Yeast cells were grown overnight at 28ºC in GPY media and diluted to an OD600= 0.1 in sterile water. Then, serial dilutions of cells (10-1 to 10-3) were transferred on the plates with replicates and incubated at 28ºC for ten days with the plates wrapped in plastic paraffin film to avoid ethanol evaporation. Each strain was inoculated twice in the same plate but in

different positions, and an exact replicate of the plate was done. With this method, four biological replicates of each strain were generated.

1.3 Results

1.3.1 Phylogenetic position of the strains

We used whole genome sequencing of four strains, exhibiting different levels of ethanol tolerance (Arroyo-López et al., 2010), to investigate the relationship between genomic differences and ethanol tolerance. Our assembly and annotation pipeline allowed us to extract about 6000 coding sequences per genome, of which 2115 concatenated gene sequences in common with other 38 strains, representative of different pure lineages described for *S. cerevisiae* (Liti et al., 2009), were used to reconstruct a multi-locus ML phylogeny (Figure1.1). When looking at the placement of our stains we observed that Temohaya-MI26 does not appear to cluster within any of the groups we selected, and shows a central position in the tree. This strain may be from a different American population not considered here, like the recently described Ecuadorean population (Peter et al., 2018). The wine strains (T73 and the two flor strains) clustered with wine/European strains, but within two sister-clades. More specifically, the flor strains GBFlor-C and CECT10094 group with EC1118 and T73 in the other clade which contains wine strains. The position of EC1118 is consistent with previous results (Coi et al., 2017) describing that strain EC1118 clusters with flor strains which form a subpopulation among wine/European strains. As EC1118 was closely related to high

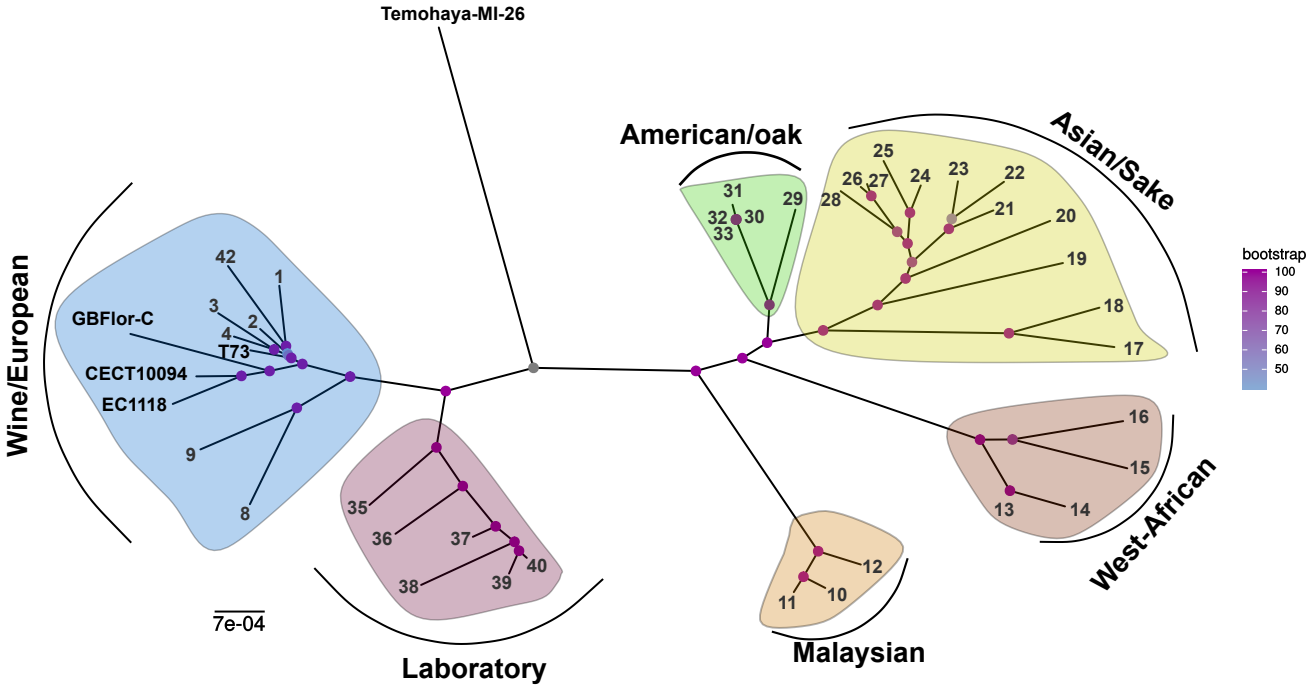

FIGURE 1.1 Phylogeny of *S. cerevisiae* strains from different origins. Multilocus Maximum Likelihood tree of 47 strains representative of different clades. Yellow: Asian/Sake, Green: American/Oak, Brown: West-African, Orange: Malaysian, Blue: Wine/European, Red: Laboratory.

ethanol-tolerant strains, and showed an intermediate tolerance, we used published genomic data of this strain for further analysis.

1.3.2 Heterozygosity levels differ between strains

The frequency of sexual reproduction and outcrossing in *S. cerevisiae* has a high impact on heterozygosity levels, which can indicate differences in life-style between strains. We assessed heterozygosity levels here calculating the number of heterozygous positions in coding regions of the studied strains. High differences are found between strains. Temohaya-MI26 has the lowest heterozygosity with 2433 hetrozygous positions in the genome. T73 and GBFlor-C have 4586 and 3094 heterozygous positons, respectively. However EC1118 and CECT10094 have the highest

levels of heterozygosity with 12983 and 13789 heterozygotic SNPs in the genome, respectively, which represent a mean of two SNPs per gene in their genomes. Interestingly, no relationship is observed between ethanol tolerance and differences in heterozygosity.

In general, heterozygous SNPs were uniformly present along the genome, although several events of loss of heterozygosity (LOH) were observed (Figure1.2). These events affected large chromosome portions, but mostly present at chromosome ends.

To identify possible genes involved in ethanol tolerance, we checked for non-synonymous SNPs fixed only in both highly ethanol-tolerant strains CECT10094 and GBFlor-C. Due to the heterozygosity and the phylogenic relatedness that these strains have with EC1118, only seven amino-acid changes were fixed and exclusive to both strains (Table 1.1). These are located in proteins encoded by six genes: *CUZ1*, *GCY1*, *RPN7*, *KAR3*, *DPB2* and *ATG13*. With the exception of *CUZ1* and *GCY1*, these genes were located on the right arm of chromosome XVI, which was affected by a LOH event exclusive of CECT10094 and GBFlor-C. Interestingly, *CUZ1* and *RPN7* are two genes related with ubiquitin and proteasome pathways, which are important processes in the maintenance of protein homeostasis and the degradation of unfolded proteins. Both processes could be related with the presence of aneuploidies in the studied strains and their ethanol tolerance, as discussed below.

TABLE 1.1 Genes affected by non-synonymous changes exclusive of CECT10094 and GBFlor-C.

Systematic name	Gene name	Description	aa position	Ref aa	Alt aa
YNL155W	*CUZ1*	*CDC48*-associated UBL/Zn-finger protein	80	Y	F
YOR120W	*GCY1*	Galactose-inducible Crystallin-like Yeast prot	86	Q	E
YPR108W	*RPN7*	Regulatory Particle Non-ATPase	6	E	K
YPR141C	*KAR3*	KARyogamy	670	K	N
YPR175W	*DPB2*	DNA Polymerase B subunit 2	565	V	F
			584	E	Q
YPR185W	*ATG13*	AuTophaGy-related	158	T	S

1.3.3 Highly ethanol-tolerant strains share chromosome III aneuploidy

Most strains of *S. cerevisiae* are diploids, but it has been shown that industrial strains, associated with human-related environments, present different ploidy levels (Gallone et al., 2016; Peter et al., 2018). We assessed our strains' ploidy by flow cytometry, and found that T73, CECT10094, EC1118 and Temohaya-MI26 were diploids. Cytometry average of triplicates compared to a known diploid were respectively: 2.117 ± 0.029, 2.200 ± 0.030, 2.196 ± 0.029, 2.218 ± 0.027. Contrastingly, GBFlor-C was found to be a triploid strain (3.510 ± 0.055).

Another method to confirm the ploidy state is to use the heterozygotic SNP frequency distribution along the genome (Figure1.2, left panel). The diploid and heterozygotic strains showed a SNP frequency distribution around 0.5, which confirms their diploid state. In the same way, GBFlor-C, which is triploid, showed a typical SNP frequency distribution around 0.33 and 0.66. As Temohaya-MI26 is completely homozygous, it was not possible to confirm its ploidy state with this method.

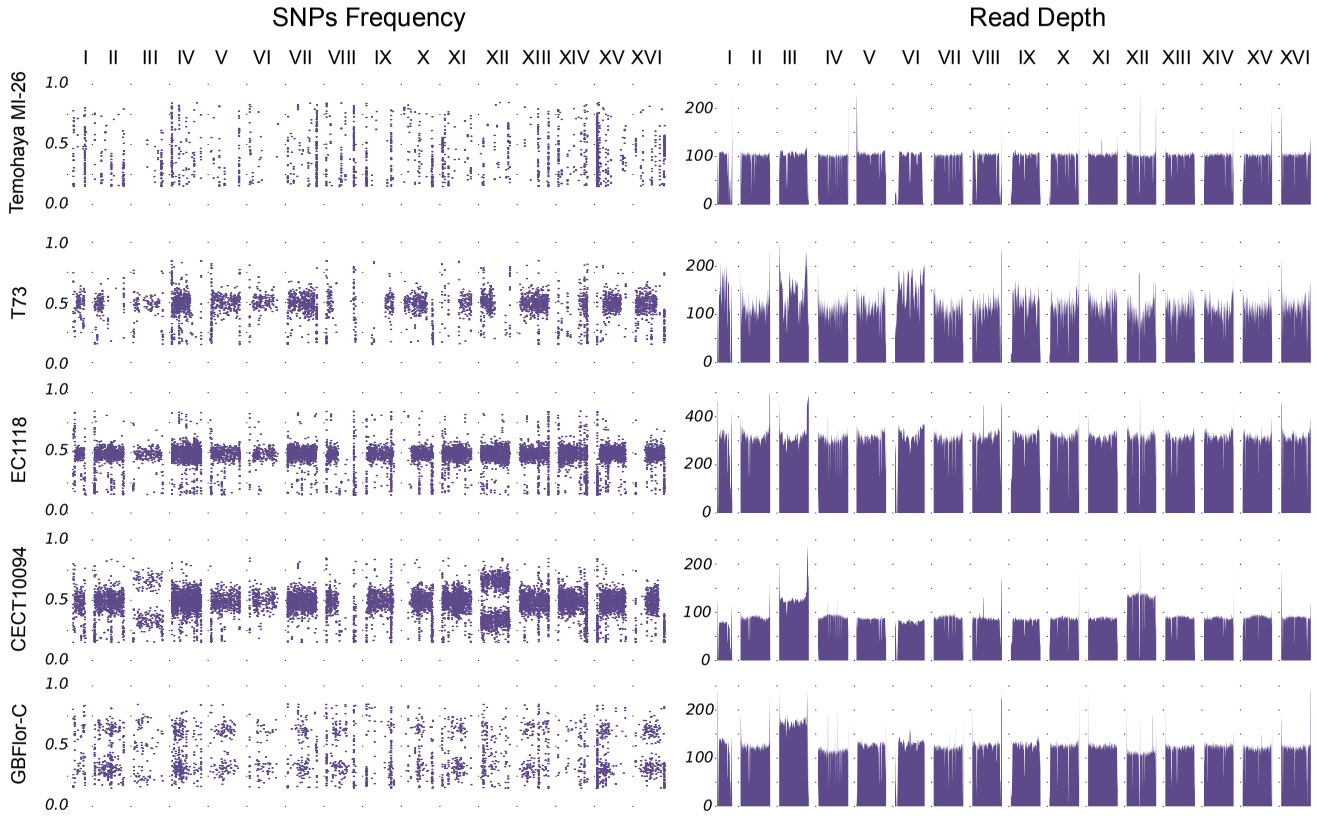

FIGURE 1.2 Genome structure of the strains. Left panel: SNP frequencies along the genome. Frequency distribution shows the different ploidies and aneuploidies present. Right panel: Read Depth by chromosomes, smoothed by the mean in 10kb windows, confirm chromosome III aneuploidy in both GBFlor-C and CECT10094 and the chromosome XII trisomy in CECT10094.

Fast adaptation to a stressful environment can be driven by large-scale genomic rearrangements. Among these, aneuploidies are getting much attention as a potential driver of adaptation of industrial relevance in *S. cerevisiae* (Gorter de Vries et al., 2017). We checked for the presence of aneuploidies in two ways: changes in read-depth between chromosomes and changes in heterozygous SNPs frequency compared to the overall genome frequency distribution (Figure1.2). Interestingly, the highest ethanol-tolerant strains, CECT10094 and GBFlor-C were aneuploids. CECT10094 had an extra copy of chromosomes XII and III, and GBFlor-C also showed an extra copy of chromosome III. As chromosome III polisomy was shared between these two strains in different ploidy backgrounds, we further investigated if this could be of importance to explain their higher ethanol tolerance.

1.3.4 Chromosome III expression increases with ethanol stress

A higher number of copies of a chromosome is related with a higher expression of the genes in this chromosome (Torres et al., 2007). We asked if in this case the higher expression of chromosome III could be related with ethanol tolerance. We used transcriptomic data from a previous study (Navarro-Tapia et al., 2016) to shed light on the importance of chromosomes expression on the ethanol tolerance. In brief, the strain with the lowest and highest ethanol tolerance of a set of strains from diverse isolation sources (Temohaya-MI26 and CECT10094 respectively) were selected and their RNA was extracted after 1 or 10 hours of growth in two conditions: after a 10% ethanol shock or without stress. The genes were grouped by chromosomes to show the global contribu-

tion of these in the expression profile of the different conditions (Figure1.3). Contribution of each chromosome in the complete transcriptome of each strains were different but here we focused on chromosome III due to its aneuploidy in the highly ethanol-tolerant strains. Without ethanol stress, chromosome III showed a significantly higher expression at one hour of growth compared to other chromosomes in CECT10094 but not in Temohaya-MI26. At 10 hours of growth, chromosome III global expression is up-regulated in both strains and in both growth conditions. One hour after ethanol stress, however, chromosome III is significantly up-regulated in both Temohaya-MI26 and CECT10094. In Temohaya-MI26, chromosome III is the most significantly overexpressed chromosome in the genome after short exposure to ethanol. Thus, the expression pattern observed here could be related with a higher expression in CECT10094 due to the aneuploidy. Furthermore, the change in the expression contribution of chromosome III in Temohaya-MI26 after ethanol shock is consistent with the presence of several genes in the chromosome contributing to the ethanol stress response even in the low ethanol-tolerant strain.

1.3.5 Chromosome III aneuploidies affect growth on ethanol in different backgrounds

Several studies showed that aneuploidies could be of importance in certain conditions (Gorter et al., 2017). In particular, chromosome III copy number variation was related with higher heat tolerance (Yona et al., 2012) and was duplicated in ethanol adaptation experiments (Voordeckers and Verstrepen, 2015). In a recent study (Peter et al., 2018), more than 1000 *S. cerevisiae* strains

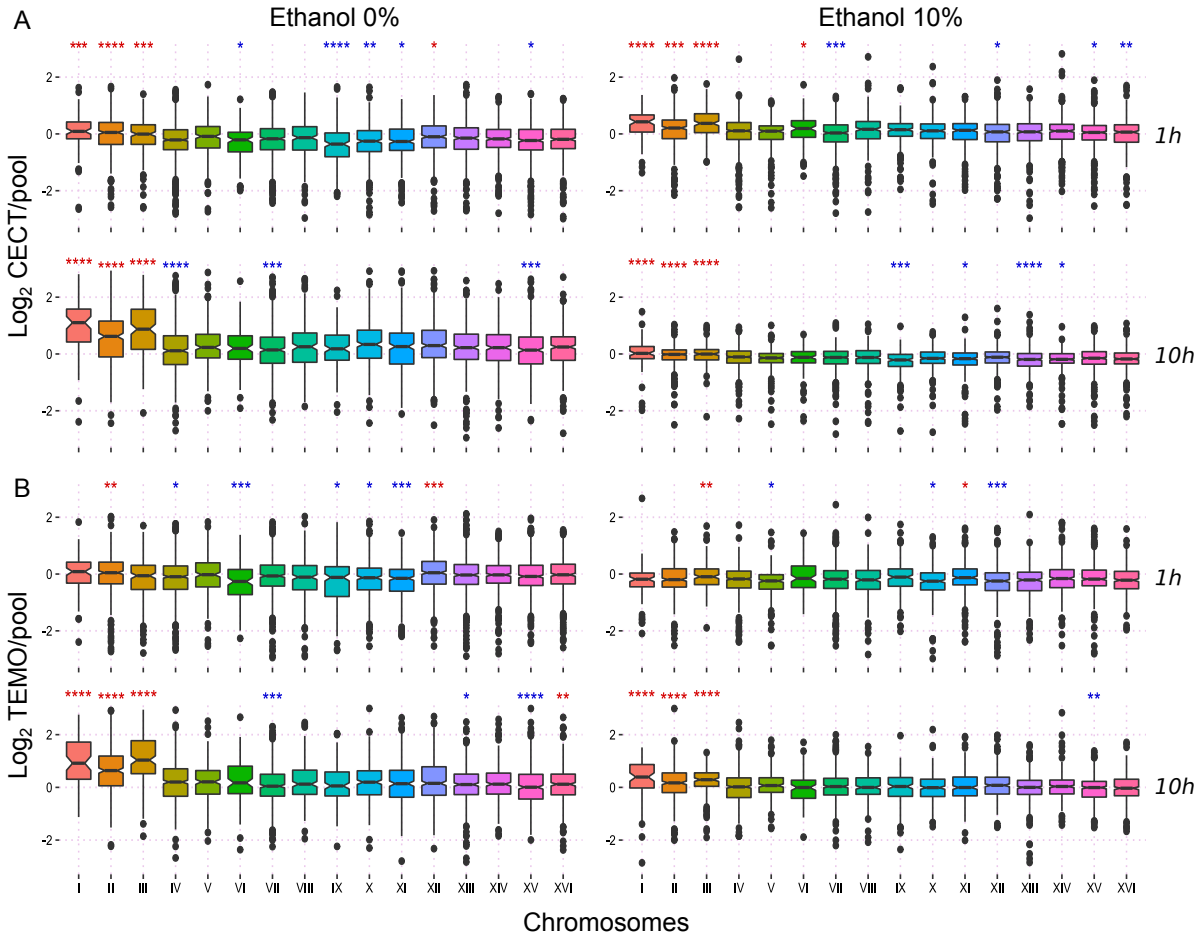

FIGURE 1.3 Transcriptional response to ethanol at a chromosomal level. **A**: expression fold change of CECT10094 compared to the pool at 1 and 10 hours growth with 0 and 10% ethanol stress. **B**: expression fold change for Temohaya-MI26. Statistics Wilcoxon test among all groups: ns : p > 0.05, * p< 0.05, ** p< 0.01, *** : p< 0.001, **** : p< 0.0001. Significance symbols are coloured in red if the chromosome is upregulated and in blue if it is downregulated.

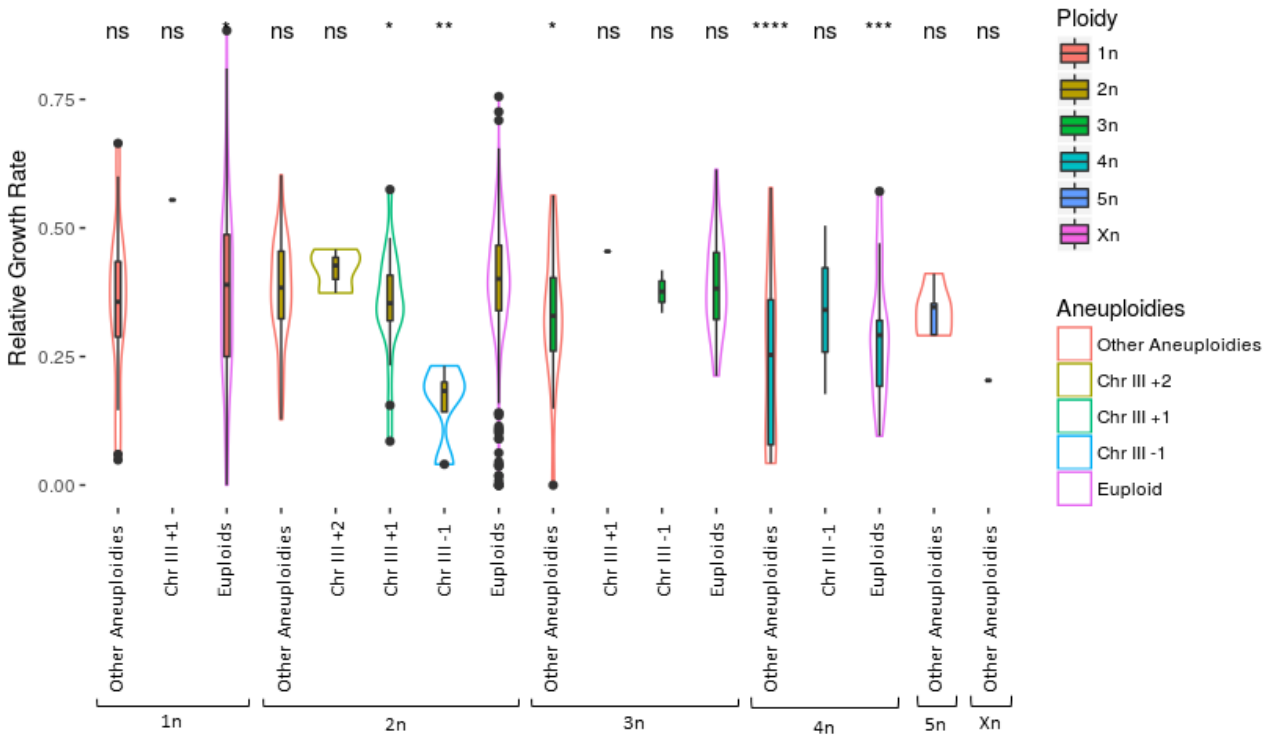

FIGURE 1.4 Relative growth rate of 1011 strains on 15% ethanol. Relative growth rate is the ratio between growth of the strain on YPD at 30ºC and growth on 15% ethanol. Strains are grouped by their ploidy and presence of different aneuploidies. Chr III +1, Chr III +2 and Chr III-1 tagged contain one extra copy, two extra copies and one less copy of chromosome III, respectively, and can exhibit other aneuploidies. 'Other Aneuploidies' group are strains that have different kind of aneuploidies but not on chromosome III, and 'Euploids' have no aneuploidy. Significance is Wilcoxon paired with diploid euploid strains as reference (ns: $p > 0.05$, * $p< 0.05$, ** $p< 0.01$, *** : $p< 0.001$, **** : $p< 0.0001$). Original data were obtained from Peter et al. (2018).

were sequenced and phenotyped in several conditions. Here we used the phenotype on 15% ethanol stress and the ploidy and aneuploidy information, and grouped the strains to search for differences in ethanol stress tolerance in a wider genetic background set (Figure1.4). Most of the ploidies and chromosome copy numbers did not show many differences compared to diploid strains exhibiting a perfect euploidy. As groups are of different sizes and many factors are involved in ethanol tolerance, differences are hard to see. However, in diploids the number of chromosome III copies showed a specific trend (Figure1.4). Strains lacking one of the chromosome copy were significantly worse than diploids growing on 15% ethanol respect to control condition. As the number of copies of

chromosome III increases, higher is the relative growth rate exhibited. Strains with an extra copy were better than the euploids with two copies, and these better than monosomic strains, with one single chromosome III.

1.3.6 Removing the extra copy of chromosome III strongly affects ethanol tolerance

To confirm that that the aneuploidy on chromosome III directly influenced the ethanol tolerance of the strains, we removed the extra copy from the genome (see materials and methods), returning strains to the euploid state. Unfortunately, we could not obtain any modified strain of CECT10094 and GBFlor-C with the experimental approach used. We therefore used a laboratory evolved strain (2-200-2) (Voordeckers and Verstrepen, 2015).

Voordeckers and Verstrepen (2015) evolved different derivative of the S288C FY5 strains in chemostats and increasing ethanol concentration during 200 generations. They found that several strains shared an aneuploidy on chromosome III at the end of the experiment, even if some of them had other ones. The strain 2-200-2 is the result of 200 generations of evolution of an original haploid strain. After 200 generation this strain was diploid and had only one trisomy on chromosome III. As this strain shared this genomic feature with the strains studied here, we used it to remove its extra chromosome III copy to test if its ethanol tolerance was affected.

We first determined the copy number of chromosome III in the strain 2-200-2 and the modified

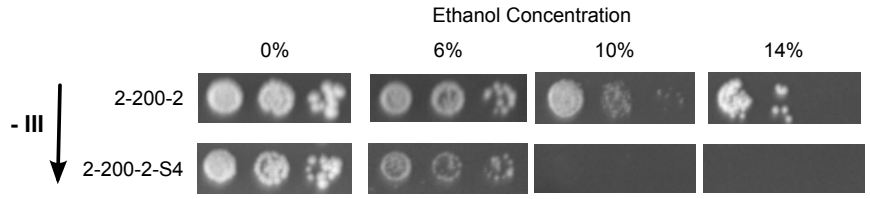

FIGURE 1.5 Removing the aneuploidy on chromosome III decreases ethanol tolerance. Drop test assays of strains 2-200-2 and 2-200-2-S4 in ethanol plates. Plates were incubated 10 days at 28ºC. 2-200-2 had three copies of chromosome III and 2-200-2-S4 had two copies. Removing the aneuploidy affects ethanol tolerance.

strain 2-200-2-S4 by qPCR to confirm that we removed the extra copy. We used primers for three genes spread along the chromosome and compared chromosome dosage using as reference genes ACT1 and YFR057W from chromosome VI. The results showed that 2-200-2-S4 had lost an extra copy of chromosome III for each one of the tested genes resulting in a gene copy number close to 2 (1.99±0.31 for ARE1, 2.14±0.28 for YCL001W-A and 2.06±0.20 for POF1).

After removing chromosome III extra copy, we tested growth of strains 2-200-2 and 2-200-2-S4 on GPY with different ethanol concentrations (Figure1.5). The strain 2-200-2, which has three copies of chromosome III, was able to grow even on 14% ethanol concentration. 2-200-2-S4, which had the extra copy of the chromosome III removed, was not able to grow at 10% ethanol medium.

1.4 Discussion

Ethanol is one of the major stresses suffered by yeasts in industrial environments. Among the different species of its genus, *S. cerevisiae* is the most tolerant to ethanol (Arroyo-López et al., 2010). Even if this characteristic is widely studied due to its importance in biotechnology and industry, it is still unknown what are the key factors that drive adaptation to high ethanol concentrations

(Snoek et al., 2016). Here, we sequenced the genome of *S. cerevisiae* strains especially selected for their differential ethanol tolerance. A previous work showed that Temohaya-MI26 was low ethanol-tolerant, T73 and EC1118 had an intermediate tolerance and GBFlor-C and CECT10094 were high ethanol-tolerant (Arroyo-López et al., 2010).

The phylogenetic analysis performed showed that the wine strains could be divided in two subclades. The first one grouped typical wine strains and contained the T73 strain, and the second grouped flor strains (GBFlor-C and CECT10094) with EC1118. These results confirm that flor strains form a different subpopulation among wine strains, as previously described (Legras et al., 2014, 2016; Eldarov et al., 2018; Coi et al., 2017). Temohaya-MI26, in contrast, was not included in any of the groups considered. The ethanol tolerance was higher in flor strains but not in EC1118 which is in the same group and closely related to CECT10094 and GBFlor-C. This points that this phenotype is variable even within the same population.

Until recently, most of the sequenced *S. cerervisiae* strains were homoploid spore derivatives to improve assembly and analysis. These methods nevertheless shadow interesting parts of genome structure. Heterozygosity levels were related to differences in strains lifestyle (Magwene et al., 2011). In industrial environments, this species reproduces asexually and has higher heterozygosity levels than natural strains (Gallone et al., 2016; Peter et al., 2018). The strains studied also showed similar trend. Temohaya-MI26, which is not related to industrial strains, showed a low heterozygosity. This may mainly be due to the use of haploselfing in its environment. In contrast, wine related strains showed higher heteozygosity, with events of LOH, and were in the range of levels

previously described for wine strains (Gallone et al., 2016; Peter et al., 2018).

We found that the highly ethanol-tolerant strains shared an aneuploidy on chromosome III in different ploidy backgrounds. A high fidelity of genome replication and segregation is vital for the survival of any organism as well as for the production of future generations. Errors in these steps during meiosis, and also during mitosis in unicellular organisms, can lead to a change in ploidy or chromosome numbers. In fact, it has been suggested that the ethanol itself could induce chromosome malsegregation (Crebelli et al., 1989). These severe genome changes can be detrimental, causing a decrease in the fitness of the organism. However, during specific circumstances, such as periods of stress, in which gene dose increase can be beneficial, polyploidy or aneuploidy can provide a higher fitness (Todd et al., 2017). Aneuploidy is gaining attention for its relevance in industrial *S. cerevisiae* strains (Gorter de Vries et al., 2017) and for its possible implications in driving adaptation in general (Chen et al., 2012; Bennett et al., 2014). Consequences of aneuploidy are usually detrimental for strain growth Mangado et al. (2018). However, it was described that specific chromosome copy-number variations could improve resistance to specific stresses. This way, chromosome III aneuploidy was related with improvement of heat tolerance (Yona et al., 2012). Other authors also found that chromosome III aneuploidies were generated as a response to ethanol stress (Gorter de Vries et al., 2017). Moreover, artificial segmental aneuploidies of chromosome III increased ethanol tolerance (Natesuntorn et al., 2015). Evolution on mild ethanol stress showed that different aneuploidies appeared, including chromosome III copy number increases (Adamczyk et al., 2016). Finally, a long-term evolution study showed that chromosome III aneuploidy was a common event (Voordeckers et al., 2015). Here, we found that relationship in non-laboratory

strains, which may indicate that it is a long-term adaptation and its fixation seems to be important for ethanol tolerance. We also showed that the number of copies of the chromosome plays a role in this phenotype in different backgrounds (Figure1.4), and that it is adaptive and affects directly to ethanol tolerance.

We dissected the expression profile by chromosomes of the high and low ethanol-tolerant strains. We found that the low ethanol-tolerant strains up-regulated chromosome III expression after ethanol stress and that the high ethanol-tolerant had its expression increased even in the absence of ethanol. Therefore, aneuploidy can be a way to change dosage of important genes present in chromosome III (Yona et al., 2012). This is consistent with our results, but further investigation is needed to find which genes in this chromosome could be involved in this process. Nevertheless, genes present in aneuploid chromosomes can change expression of other genes in other chromosomes, causing broad expression changes (Selmecki et al., 2008).

Aneuploidy itself affects the cell in different ways. Additional copies of a chromosome increases proteotoxic stress, which affect the protein folding processes in the cell (Torres et al., 2007). Interestingly, yeast were found to induce unfolded protein response under ethanol stress (Navarro-Tapia et al., 2016, 2017). Two out of seven nonsynonymous changes found in the high ethanol-tolerant strains affected genes related to protein homeostasis. We open here the possibility that ethanol and aneuploidy tolerance could involve similar processes, which may involve fixing variants affecting these processes and therefore aneuploidy itself could play a role on improving ethanol tolerance. As chromosome III is one of the smallest chromosome in *S. cerevisiae* genome,

we cannot discard that the aneuploidy tolerance induced could be the cause of the observed phenotype.

In conclusion, in this work we showed that ethanol tolerance was related to an aneuploidy on chromosome III in wine *S. cerevisiae*. Further work will be needed to elucidate the actual mechanism by which this phenomenon happens, but we confirmed that this is an adaptive trait that seems to be a widespread trend.

CHAPTER 2

Genome structure reveals the diversity of mating mechanisms in *Saccharomyces cerevisiae* x *S. kudriavzevii* hybrids, and the genomic instability that promotes phenotypic diversity.

2.1 Introduction

Hybridization, considered as reproduction between individuals belonging to genetically distinct populations or different species (Harrison, 1993), has played an important role in the evolution of

many eukaryotic organisms. This way, the genetic interchange between divergent lineages due to hybridization can generate new phenotypic diversity through the expression of hybrid vigor, allows for adaptation to new environments through the introgression of novel alleles and transgressive segregation (Bell and Travis, 2005), and may contribute to the formation of new hybrid species(Abbott et al., 2013) , either by allopolyploidy, when the ploidy of hybrids increases with respect to that of the parental species, or homoploidy, when ploidy remains unaltered(Arnold, 1997).

Hybridization and its evolutionary consequences in speciation and adaptation have been widely studied in plants (Goulet et al., 2017; Yakimowski and Rieseberg, 2014) and animals (Abbott et al., 2013; Dowling and Secor, 1997; Mavárez and Linares, 2008) including hominins (Ackermann et al., 2019), but not so extensively in fungi. In fungi, attention was mainly focused on hybridization in pathogenic fungi (Depotter et al., 2016; Stukenbrock, 2016) and yeasts of biotechnological interest (Morales and Dujon, n.d.; Peris, Pérez-Torrado, Hittinger, Barrio and Querol, 2017) , being hybrids of the *Saccharomyces* genus the most studied examples (Sipiczki, 2008, 2018*b*) , including the role of hybridization in yeast speciation and adaptation (Eberlein et al., 2019; Greig et al., 2002; Lopandic et al., 2018; Ortiz-Merino et al., 2017; Pérez-Través et al., 2014; Stelkens et al., 2014).

At present, the *Saccharomyces* genus is composed of eight species: S. arboricola, *S. cerevisiae*, *S. eubayanus, S. jurei, S. kudriavzevii, S. mikatae, S. paradoxus,* and *S. uvarum* (Boynton and Greig, 2014; Hittinger, 2013; Naseeb et al., 2017). These species show postzygotic reproductive isolation (**?**), and therefore, mating between them is possible and hybridization is easily achieved in the laboratory (Naumov, 1996; Pérez-Través et al., 2012). Some studies also demonstrated that

interspecific hybridization can also occur in nature, in the insect gut (Pulvirenti, A. and Zambonelli, C. and Todaro, A.and Giudici, P., 2002; Stefanini et al., 2016). These interspecific hybrids are sterile, mainly due to nucleotide divergence that prevents spore viability (**?**Liti et al., 2006). However, they are viable and can reproduce asexually by budding (Naumov, 1996; Sipiczki, 2008).

The first and well-known example of interspecific *Saccharomyces* hybrid is the lager yeasts *S. pastorianus* (syn. *S. carlsbergensis*) (Kodama et al., 2006) , which is a partial allotetraploid hybrid between *S. cerevisiae* and another species, later identified as *S. eubayanus* (Libkind et al., 2011). Most natural hybrids have been isolated from fermentative environments in European regions with Continental and Oceanic climates, and were generated by spontaneous hybridization between *S. cerevisiae* (*Sc*) and a cryophilic species: *S. eubayanus* (*Se*), *S. kudriavzevii* (Sk), and *S. uvarum* (*Su*) (Boynton and Greig, 2014; Peris, Pérez-Torrado, Hittinger, Barrio and Querol, 2017) .

Contrastingly, natural hybrids seem to be almost absent in wild environments, where only a few hybrids between the closely related species *Sc* and *S. paradoxus* (*Sp*) have been isolated (Barbosa et al., 2016). However, the presence of introgressed nuclear genome regions between *Sc* and Sp (Liti et al., 2006; Peter et al., 2018; Pontes et al., 2019; Zhang et al., 2019) , as well as between *Se* and *Su* (Almeida et al., 2014), another pair of closely related species, suggests hybridization occurs in the wild between closely related species but gives rise to unstable hybrids (Guillamón and Barrio, 2017; Sipiczki, 2018*a*). This is also confirmed by the presence of introgressions in the mitochondrial genome of different *Saccharomyces* species (Peris, Arias, Orlić, Belloch, Pérez-Través, Querol and Barrio, 2017).

The physiological characterization of industrial *Saccharomyces* hybrids demonstrated that they inherited the good fermentation performance of the *Sc* parent and the capability to grow at lower temperatures of the non-*Sc* partner, in addition to other properties of biotechnological interest (Belloch et al., 2008; Gibson et al., 2013; Krogerus et al., 2018; Ortiz Tovar, 2018; Pérez-Torrado et al., 2015; Pfliegler et al., 2014). These interesting properties contributed by the non-*Sc* species prompted the development of artificial interspecific hybrids for industrial applications (García-Ríos et al., 2019; Hebly et al., 2015; Krogerus et al., 2015; Pérez-Torrado et al., 2017; Pérez-Través et al., 2012) .

For more than one decade, our laboratory described and characterized, both at the molecular and physiological level, *Sc* x *Sk* hybrids (as reviews see Pérez-Torrado et al., 2018; Peris et al., 2018). By combining the comparative genome characterization of hybrids, deduced from microarray hybridization (Peris, Lopes, Belloch, Querol and Barrio, 2012*a*), with a multilocus phylogenetic analysis (Peris, Lopes, Arias and Barrio, 2012) , seven potential hybridization events were predicted as the origin of *Sc* x *Sk* hybrids (Peris, Lopes, Belloch, Querol and Barrio, 2012*a*), including the two most frequent hybrid lineages. One was predominant in Wädenswill, Switzerland, and was related to Trappist brewing hybrids, and the other was widely distributed from the Rhine valley (Alsace and Germany) to the Danube valley (Pannonian region: Austria, Croatia, and Hungary).

In the present study, we selected different hybrid strains as representatives of the different groups defined according to those previous studies, to obtain whole-genome sequences to unveil the mechanisms involved in the origin and evolution of these hybrids, as well as the ecological and

geographic contexts in which spontaneous hybridization and hybrid persistence take place. The understanding of the mechanisms involved in hybrid formation is also of interest to develop programs of industrial yeast improving based on artificial hybridization.

2.2 Materials and methods

2.2.1 Strains and genome sequencing

In this study, we selected different hybrid strains (Table 2.1) as representatives of the different groups defined according to previous characterizations (Peris, Lopes, Arias, et al., 2012; Peris, Lopes, Belloch, et al., 2012). Total yeast DNA extraction was performed according to the method described by Querol et al. (1992). Natural hybrids were sequenced with paired-end libraries of 100 bp with a mean insert size of 300 bp in an Illumina HiSeq 2500 instrument. The artificial *Sc* x *Sk* hybrid obtained (see below) was sequenced with ABI SOLiD paired-end of 25-75 nt. Genome sequencing reads and assemblies of reference strains *Sc* T73 and *Sk* CR85 were obtained in previous studies (Macías et al., 2019; Morard et al., 2019).

2.2.2 Genome assembly and annotation

The hybrids genomes sequence reads were trimmed with Sickle v1.2 (Joshi and Fass, 2011) using a minimum quality per base of 28, and filtered with a minimum read length of 85nt. Velvet

v1.2.03 (Zerbino and Birney, 2008) was used to determine which k-mer size was optimum for each sequencing library. The assembly step was performed with Sopra v1.4.6 (Dayarian et al., 2010) integrated with Velvet, by using the k-mer size determined previously. SSPACE v2.0 (Boetzer et al., 2011) and GapFiller v1.11 (Boetzer and Pirovano, 2012) were used to improve scaffold length and remove internal gaps. The resulting scaffolds were then aligned to a concatenated *Sc*-*Sk* genome reference with MUMmer v3.07 (Kurtz et al., 2004). The genomes used as references were *Sc* T73 and *Sk* CR85. After the alignment, the scaffolds were organized into chromosomes with an in-house script.

Hybrid genomes were annotated using RATT (Otto et al., 2011) to transfer the annotation by sequence homology using the *Sc* T73 and *Sk* CR85 genome annotations. AUGUSTUS web server (Stanke and Morgenstern, 2005) was used to complete the annotation in regions in which no gene transfer was obtained with RATT. Annotations were manually checked and corrected using Artemis (Rutherford et al., 2000) .

2.2.3 Mappings, variants detection, and ratio analysis

Hybrid genome mappings were performed against a concatenated reference of *Sc* T73 and *Sk* CR85. Illumina sequences were mapped by using bowtie2 v2.3.0 (Langmead and Salzberg, 2012), with default parameters. SOLiD reads of the artificial hybrid were mapped with bfast v0.7.0a (Homer et al., 2009). To analyze the genome content of hybrids, read depths (RD) were computed with bedtools v2.17.0 (Hung and Weng, 2016). Mean RD in 10-kb sliding windows of 1 kb steps

were calculated and plotted with ggplot2 (Wickham, 2009). A RD ratio was calculated for each gene shared between *Sc* and *Sk* based on the mean RDs for each gene in each subgenome, obtained as:

$$R_{gene} = \frac{\overline{RD_{Sc}}}{(\overline{RD_{Sc}} + \overline{RD_{Sk}})}$$

This ratio goes from 0, when the *Sk* gene is the only present, to 1, when the only gene present is from *Sc*. Count histograms were plotted to calculate the average ratio in each hybrid genome, for this purpose large regions with ratios of 0 or 1, indicating that these regions were lost after the hybridization event, were excluded. This average ratio was considered as the expected hybrid ratio, i.e., the ratio between subgenomes in the ancestral hybrid of each strain just after the hybridization event. The expected hybrid ratio was then subtracted to each gene ratio to obtain the deviation from expectation. This deviation is positive or negative, if the *Sc* or the *Sk* subgenomes, respectively, increased after the hybridization event. We considered that a gene conserved its original hybrid state if its deviation was between 0.05 and -0.05, due to the noise observed. Again, due to noise, we used 0.1 and -0.1 as thresholds to consider that a gene effectively increased its *Sc* and *Sk* dosage, respectively.

2.2.4 Phylogenetic analysis

Each gene sequence was extracted from the annotation of the natural hybrids and classified as belonging to the *Sc* and *Sk* hybrid subgenomes. For each species, alignments were obtained with MAFFT v7.221 (Katoh and Standley, 2013), for the translated amino acid sequences of orthologous genes from hybrids and 4 reference *Sk* strains or 75 *Sc* genomes representative of different clades and origins. The aligned amino acid sequences were back-translated to nucleotides and the whole set of alignments concatenated. RAxML v8.1.24 (Stamatakis, 2014) was used to construct a Maximum Likelihood (ML) phylogeny based on the concatenated alignment with model GTR-Γ and 100 bootstrap replicates. The concatenated alignment was also used to obtain a Neighbor-Net phylogenetic network of the GTR-Γ corrected nucleotide distances with SplitsTree v4.14.6 (Huson and Bryant, 2006). Trees were drawn using iTOL v3 (Letunic and Bork, 2016).

2.2.5 Sporulation assays

Yeast cells were incubated on acetate medium (1% CH3COONa, 0.1% Glucose, 0.125% yeast extract, and 2% agar) for 5–7 days at 28 °C to induce sporulation. Sixteen asci were collected for each strain when they were present. Ascus wall was digested with 1,3-glucuronidase (Sigma) adjusted to 2 mg mL−1, and spores were then dissected in GPY agar plates with a Singer MSM manual micromanipulator. Spores were incubated at 28ºC for 3-5 days, and then, viability was tested.

2.2.6 Artificial *S. cerevisiae* x *S. kudriavzevii* hybrids obtained by rare mating

Artificial hybrids were generated by rare mating between the diploid wine *Sc* T73 strain and the diploid wild European *Sk* CR85 strain. As mentioned, genome sequences of these strains are available from previous studies (Macías et al., 2019; Morard et al., 2019). Antibiotic resistances were used as hybrid selection markers. For this purpose, strains T73 and CR85 were transformed with geneticin G418-resistance pGREG526 (Jansen et al., 2005) and hygromycin B-resistance pRS41H (Taxis & Knop, 2006) centromeric plasmids, respectively, by using the LiAc/SS carrier DNA/PEG method (Gietz, 2014). Rare mating was performed according to Spencer and Spencer (1996), with slight modifications (Pérez-Través et al., 2012). Strains carrying the resistance plasmids were grown separately in 25 mL GPY broth with 200 µg mL-1 of the corresponding antibiotic for 48 h at 25ºC. Cells were recovered by centrifugation (3,000 ×g for 5 min at room temperature), and the pairs of yeast cultures to be hybridized were placed together in the same tube. Aliquots of these mixed strains were inoculated in 20 mL of fresh GPY medium. After 5–10 days of static incubation in the slanted position at 25ºC, cells were recovered by centrifugation (3,000 ×g for 5 min at room temperature), washed in sterile water, re-suspended in 1 mL of PBS and incubated for 2 h. A heavy suspension of the mixed culture was spread on GPY plates supplemented with 200 µg mL-1 of each antibiotic and incubated at 25ºC for 48h. Colonies, resistant to both antibiotics, were isolated and purified by re-streaking on the same medium (GPY with both antibiotics). The hybrid nature of these colonies was confirmed by PCR amplification of the BRE5 and PPR1 protein-encoding nuclear genes, and the subsequent RFLP analysis with restriction enzyme HaeIII (Takara Bio Inc)

as described elsewhere (González, Barrio, & Querol, 2008).

2.3 Results

2.3.1 Phylogeny reveals several independent hybridization events

In this study, we sequenced the genome of seven *Sc* x *Sk* natural hybrids (Table 2.1) to decipher their origins and mating process. One of the first questions is whether these natural hybrids are the result of one single hybridization event followed by diversification or are derived from independent hybridization events. We assembled and annotated the genome of the hybrids and extracted the coding sequence of each gene to reconstruct a multi-locus phylogeny for each of the subgenomes.

In the reconstruction of the phylogenetic history of the *Sk* genome fraction of the hybrids, we only used 647 genes common to all hybrids and the four *Sk* genomes currently available. This low number of genes is due to the extreme reduction of the *Sk* subgenome in the AMH strain

TABLE 2.1 Ploidy and spore viability of *S. cerevisiae* x *S. kudriavzevii* hybrids. Spore Viability is expressed in percentage followed by the number of viable spores/total number of spores tested between brackets. Nd, No data available because the number of asci was very small or absent.[a]Peris, Lopes, Belloch, Querol and Barrio (2012*a*);[b]Peris, Lopes, Arias and Barrio (2012)

Hybrids	Isolation Source	Origin	Ploidy	Spore Viability
VIN7	Wine	Alsace, France	3.07 ± 0.08[a]	7.81% (5/64)
W27	Wine	Wädenswil, Switzerland	3.18 ± 0.08[a]	ND
IF6	Dietary complement	Barcelona, Spain	3.25[b]	0% (0/64)
CECT11002	Trappist beer	Louvaine-la-Neuve, Belgium	3.02 ± 0.14[a]	ND
MR25	Respiratory tract	Barcelona, Spain	2.92[b]	10.94% (7/64)
AMH	Wine	Geisenheim, Germany	3.85 ± 0.18[a]	ND
PB7	Wine	León, Spain	3.96 ± 0.08[a]	95.30% (61/64)

(see Figure 2.2). In the *Sk* subgenome ML phylogeny (Figure 2.1A), hybrids cluster together as a sister group closely related to the reference European *Sk* strains from Spain (CA111 and CR85) and Portugal (ZP591). Despite the low number of strains available, we can observe that hybrids form 3 different subgroups: one including IF6, VIN7 and AMH, a second comprising MR25 and CECT11002, and a third possible subgroup formed by PB7 and W27, although not significant according to its bootstrap value (68%). These results indicate that hybridizations involved several European *Sk* strains, closely related but different from the Iberian strains.

To investigate the origin of the *Sc* subgenome, we used 75 *Sc* strains, representative of different groups described in previous studies on the *Sc* population genomic diversity (Gallone et al., 2016; Legras et al., 2018; Liti et al., 2009; Peter et al., 2018). In this case, 538 orthologous gene alignments were concatenated to obtain an ML phylogeny (Figure 2.1B) and a Neighbor-Net phylogenetic network (Figure 2.1C), both reproducing the different populations described for *Sc*.

Wine hybrids clearly cluster within the wine population in two separated subgroups: the typical wine strains, which includes PB7 and W27, and the biofilm-forming flor strains, comprising hybrids VIN7, IF6, and AMH. Interestingly, IF6, which was isolated from a dietetic complement, belongs to the flor clade and is closely related to VIN7. The clinical isolate MR25 appears in an independent lineage, not included in any of the groups considered. This isolate could be an admixed strain as it appears in the Neighbor-net phylogenetic network in an intermediate position between wine and Beer2 populations (Figure 2.1C). Finally, the brewing hybrid CECT11002 clusters within the Beer2 group, together with the admixed brewing *Sc* strains. In general, hybrid *Sc* subgenomes

clustered according to their isolation sources, a result supporting independent hybridization events in different locations and environments.

2.3.2 Hybridizations mainly involved, but not only, rare-mating as the main conjugation mechanism

Yeasts from the *Saccharomyces* genus usually conjugate by 'canonical' mating (Cross, Hartwell, Jackson, & Konopka, 1988) between haploid cells/spores of opposite mating types, a and α either from the same tetrad (automixis), from different asci (amphimixis), or derived from a mating-type switch (haploselfing), with different genetic consequences (Magwene, 2014; Magwene et al., 2011). In all cases, the resulting cells of 'canonical' crosses are diploid and heterozygous for mating type, which lack the mating ability (non-maters). However, these diploid non-maters can also conjugate by 'rare' mating (Pomper, Daniels, & McKee, 1954), when they become mating competent by a mating-type conversion to a homozygous genotype (Gunge & Nakatomi, 1972). Consequently, hybrid genomic architectures will differ depending on which conjugation type was involved in the hybridization events. To unveil the mating mechanisms involved in the generation of these hybrids, we genetically characterized our hybrids by using read mapping and flow cytometry, to determine their genome compositions and ploidies, as well as variant calling analysis, to measure their levels of heterozygosity. In addition, we also measured sporulation capability and spore viability in most hybrids (Table 2.1).

The ploidy data shows that most of the hybrids are allotriploids, with the exception of AMH

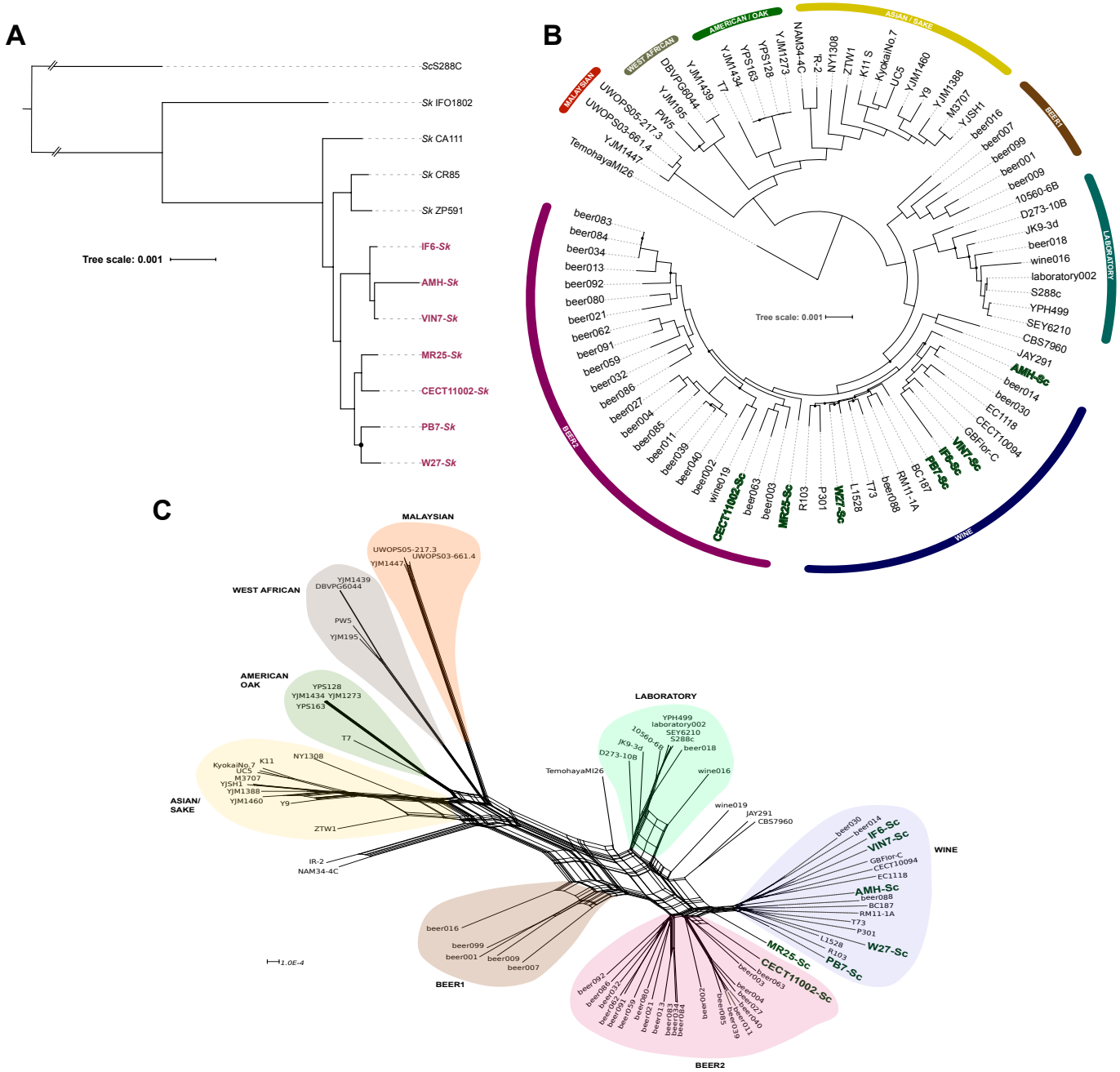

FIGURE 2.1 Phylogenetic analysis of the natural hybrids genomes. A. ML tree of 647 concatenated genes alignment with four *Sk* strains available and *Sc* S288C as outgroup. **B**. 75 *Sc* strains representatives of different groups and 538 genes were used to reconstruct an ML phylogeny of the *Sc* strains and *Sc* subgenome of the hybrids. Black dots represent nodes with bootstrap values < 0.70. **C**. The same alignment used in B was also used to perform a Neighbor-Net analysis.

and PB7, which are allotetraploids (Table 2.1). The contribution of each parental species to the hybrid genomes is shown in Figure 2.2. In the allotriploid hybrids, we observe that the *Sc* content is twice that of *Sk* in most parts of the genome. Hybrids VIN7, W27, IF6, CECT11002, and MR25 are triploid with a diploid contribution of *Sc* and haploid contribution of *Sk*. In addition, they also present different aneuploidies (polysomies and monosomies), chimeric chromosomes due to recombination between *Sc* and *Sk* homeologous chromosomes, or loss of certain non-centromeric chromosomal regions (see next section). Low spore viabilities shown by these hybrids, ranging from 0 to 11% (Table 2.1), are also in accordance with their allotriploid nature.

As the hybrid *Sc* subgenomes are diploid, we assessed heterozygosity levels by constructing a density plot of the SNP frequency for each chromosome. Interestingly, the *Sc* subgenome of the triploid hybrids showed different levels of heterozygosity according to the strain. Thus, wine strains VIN7, W27 an IF6 exhibit between 5000 and 2800 heterozygous SNPs, which is an intermediate heterozygosity level, typical of wine strains (Gallone et al., 2016; Peter et al., 2018). In contrast, the brewing CECT11002 and the clinical MR25 hybrids show higher levels of heterozygosity, 10969 and 16468 heterozygous SNPs, respectively. These higher levels are typical among beer *Sc* strains (Gallone et al., 2016), in accordance with the putative brewing origin of their *Sc* parents. The genome architecture of the allotriploid hybrids and their levels of heterozygosity can only be explained if hybrids were originated by rare-mating crosses between haploid *Sk* cells or spores and mating-competent heterozygous diploid *Sc* strains of wine or beer origins, depending on the hybrid.

As mentioned before, two exceptions to triploidy are hybrids AMH and PB7, which differ from others, and among them, in their genome compositions (Figure 2.2). On one hand, AMH is a tetraploid that shows an extreme reduction of the *Sk* subgenome contribution. Most AMH genome consists of four copies of *Sc* chromosomes, except chromosomes VI, VII, IX, XIII, and XV, with three *Sc* copies and one *Sk*, and chromosome IV, with three *Sc* copies and a chimerical copy with a tiny part of the left arm of *Sk*. This AMH hybrid appears as sterile, unable to sporulate (Table 2.1), which is likely due to the wrong segregation of the *Sc* tetravalent during meiosis.

Interestingly, the AMH *Sc* subgenome is highly heterozygous, and its SNP frequencies are compatible with the tri- or tetrasomy of its *Sc* chromosomes. The fact that AMH tetraploid genome is mainly coming from *Sc*, together with the extreme reduction of the *Sk* haploid contribution, could be explained by two consecutive hybridization events. One possibility is that an allotriploid hybrid, originated by a similar mechanism than the other wine allotriploids, later conjugated by rare mating with a haploid *Sc* to form this tetraploid, with a subsequent drastic *Sk* subgenome reduction due to genome instability. An alternative explanation is based on the observation that hybrids can generate viable spores (<1-5%) because they contain most of their chromosomes coming from the same parent. Therefore, one of these rarely viable spores, carrying two copies of most *Sc* chromosomes and one copy of few *Sk* chromosomes, conjugated with a mating-competent diploid *S. cerevisiae* cell to generate the AMH hybrid.

On the other hand, PB7 is a perfect allotetraeuploid with two complete homozygous copies of each parental subgenome. As its *Sc* subgenome is coming from a parental wine strain, moderate het-

erozygosity would be expected, however, the few heterozygous SNPs are located in subtelomeric repetitive regions. Therefore, the most probable mechanism that originated this allotetraeuploid hybrid, with two homozygous copies of each subgenome, is a spore-to-spore conjugation followed by a subsequent whole genome duplication to become an amphidiploid with spore viability of 95.3% (Table 2.1).

2.3.3 Homeologous recombination drives genome evolution and reduces fertility

An interesting and open question is to understand how the hybrid genome content evolved after hybridization. Here we used a simple ratio between *Sc* and *Sk* gene contents and its deviation from the expected ratio to untangle which changes occurred in the different hybrid genomes. We calculated for each gene of the hybrid genome a read-depth ratio (see Material and Methods section). This ratio goes from 0, when only the *Sk* allele is present, to 1, when only the *Sc* allele is present. We used histograms of the frequencies of read-depth ratios to determine the average read-depth ratio for each strain, which was then subtracted to each gene read-depth ratio to calculate the deviation from the hybrid state expectation.

For this purpose, we analyzed a total of 5,449 genes, that were present and annotated in both reference strains *Sc* T73 and *Sk* CR85. The read-depth ratio clearly shows that the *Sc* subgenome content is higher in all hybrids except PB7 (Figure 2.3A). We established the expected ratio for each strain from the most frequent read-depth ratio (Figure 2.3C), excluding ratios of 0 or 1, indicative of the loss of the *Sc* or *Sk* allele, respectively. Triploid strains MR25, W27, VIN7 and CECT11002

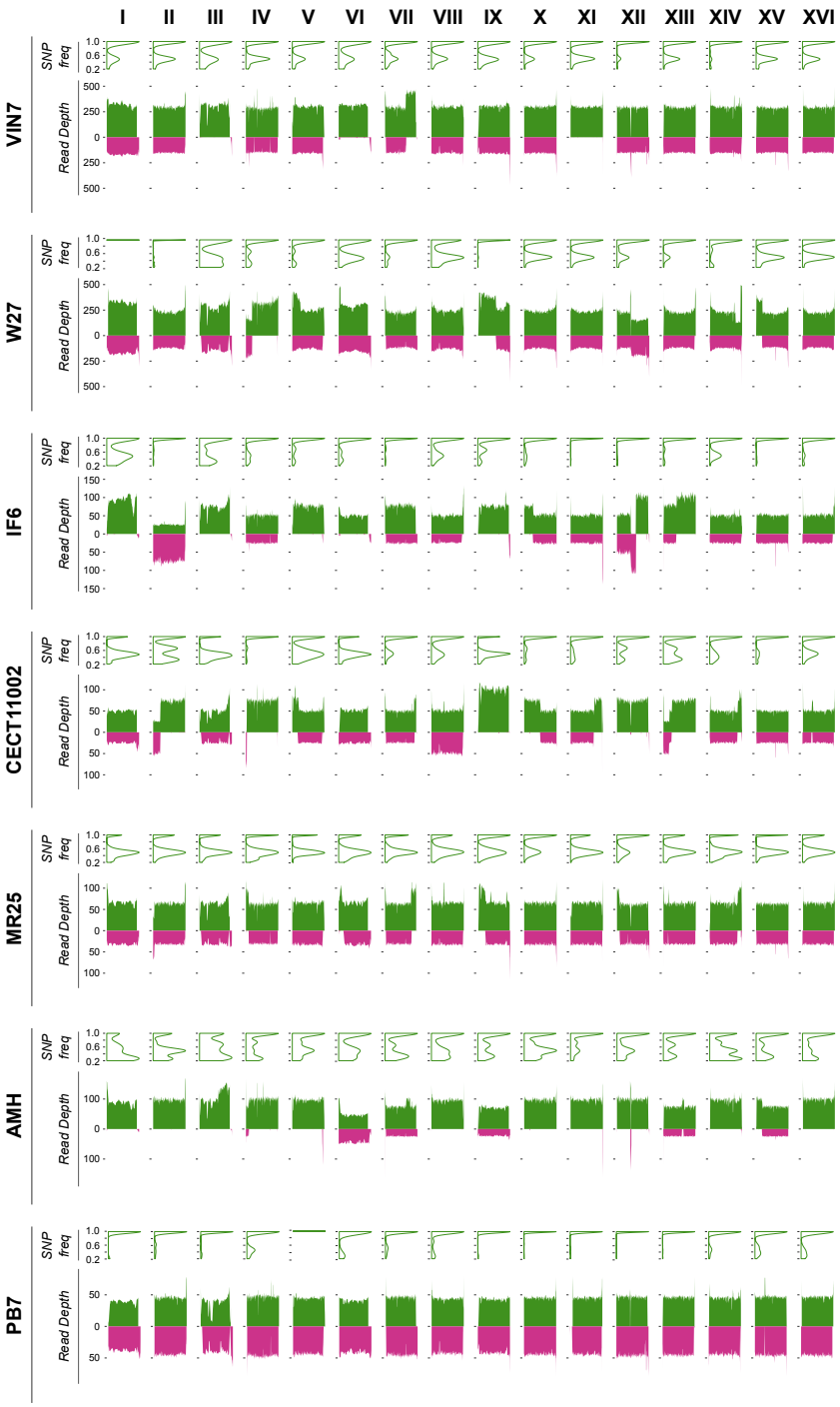

FIGURE 2.2 Genome composition of the strains. For each strain, we represent (up) the SNPs frequency density distribution in *Sc* subgenome and (down) the Read Depth for each chromosome. The SNPs frequencies are represented in the y-axis and the density is represented for the whole chromosome in the *Sc* subgenome (SNPs distribution along the chromosome is shown in Supplemental Figure S1). Most of the chromosomes show two peaks, one around 1 that are homozygous SNPs, and a second one around 0.5, heterozygous SNPs in the strains that have two *Sc* copies. This distribution changes according to the ploidy or presence of aneuploidies. Read depth is represented for 10kb windows moving by 1000nt. The *Sc* subgenome is colored in green and *Sk* subgenome in pink.

had a most frequent ratio of 0.66, congruent with their origins from hybridization events between diploid *Sc* and haploid *Sk* cells. Although the triploid IF6 hybrid shows a wider range of read-depth ratios due to aneuploidies, the expected hybrid ratio is 0.66, indicating a similar origin than the other triploid hybrids. In the tetraploid AMH, the most frequent hybrid ratio is 0.75, which confirms its 3n *Sc* and n *Sk* contributions. Finally, PB7 showed the less diverse ratio distribution with an expected hybrid ratio of 0.5, indicating an equal contribution of *Sc* and *Sk* in its origin.

The deviation from the expected ratio (Figure 2.3D) shows that most changes imply replacements of the *Sk* alleles by *Sc*. Thus, IF6 shows the most important number of changes, which could indicate a higher genomic instability. On the other hand, PB7 shows few changes being an almost perfect hybrid.

It is interesting to remark that only 290 genes maintain their original hybrid ratio in all strains, and most of them, 284, are located on chromosome XV, indicating that this chromosome is the only that preserved its original *Sc*/*Sk* proportion in all hybrids. In general, ratio deviations observed are grouped in blocks of genes located in the same chromosome regions, which indicates that chromosome loss and the generation of chromosome chimeras due to recombination between homeologous chromosomes are the main mechanisms involved in changes of the genome composition of hybrids, usually biased towards a reduction of the *Sk* genome fractions. However, most chromosome rearrangements are specific of each hybrid, and very few are shared between the different strains. In fact, only six genes show a replacement in all hybrids, including PB7, of the *Sk* allele by the *Sc* one. These are BUD5, which is overlapping the MAT locus on the opposite strand of chromosome III;

EFT2, located on chromosome IV and encoding a translation elongation paralogous to EFT1; and a subtelomeric region encompassing genes FZF1, ZRT1, ADH4 and MNT2, which is located on the left arm of chromosome VII. This region corresponds to the only homeologous recombination shared by all hybrid strains.

Finally, there seems to be a correspondence between the number of genome rearrangements and hybrids fertility. Thus, MR25 and VIN7 are clustered together with PB7 in the ratio deviation analysis (Figure 2.3B) because they exhibit the smallest number of chromosomal rearrangements, and interestingly are the allotriploids with highest fertilities, with spore viabilities of 10.9% and 7.8%, respectively (Table 2.1). The other allotriploids exhibit more genome rearrangements and their spore viabilities are 0%, or no asci were detected. Therefore, the higher the number of homeologous chromosome rearrangements, the lower the spore viability.

2.3.4 Artificial hybridization by rare-mating reproduces the genome architecture of natural hybrids

Artificial hybridization is increasingly used to improve industrial *Saccharomyces* yeasts (Krogerus et al., 2015; Pérez-Través et al., 2012). One of the methods used to generate artificial hybrids is rare-mating because they acquire the whole genome of both parents to combine most of their physiological properties.

As seen, our study indicates that most natural hybrids were also generated by rare-mating.

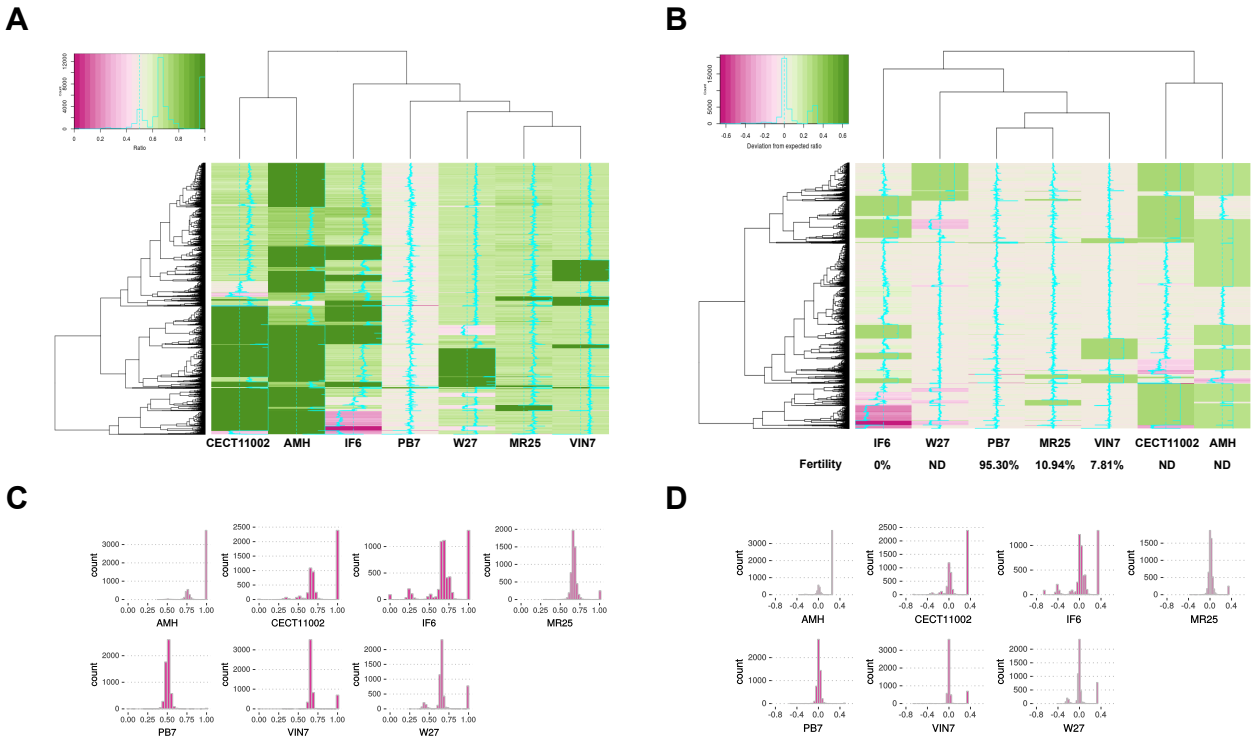

FIGURE 2.3 ***Sc/Sk* genome ratio analysis.** For each gene, we calculated a ratio of the genomic content of *Sc* vs *Sk*. **A**. Heatmap of the ratio of *Sc*/*Sk*. The values go from 0 if only *Sk* alleles are present (pink) to 1 if only *Sc* is present (green). **C**. Histogram of the count of each different ratio used to calculate the most common hybrid constitution. Most hybrids have 0.66 as the most common peak except AMH (0.75) and PB7 (0.5). **B**. Heatmap of the deviation from the expected hybrid ratio (observed in C). Values are negatives if the change is to *Sk* (pink) and positive if the change is to *Sc* (green). Fertility of the different strains are shown under each strain (Table 2.1). **D**. Histogram of the count of the deviation from the expected ratio showing the directional replacement to *Sc* in most of the hybrids.

Therefore, we decided to replicate a natural hybridization in the lab by a 'rate-mate' crossing of the diploid wine *Sc* strain T73, closely related to the *Sc* parental, and the diploid *Sk* CR85, isolated from an oak tree in Spain. For hybrid selection, parental strains were transformed with antibiotic-resistance plasmids (see Materials and Methods section), which are easily removed after hybridization, to avoid the effect of the use of auxotrophic mutations. Once an artificial hybrid was obtained, its genome was sequenced, assembled, and compared with the genomes of the parental strains (Figure 2.4), whose sequences were available (Macías et al., 2019; Morard et al., 2019).

The parental *Sc* T73 possesses a diploid genome with a moderate heterozygosity level, as expected for a wine strain, and several homozygous regions due to Loss of Heterozygosity events. The parental *Sk* CR85 genome is diploid and completely homozygous, which is common in natural *Saccharomyces* strains, due to regular haploselfing events. The resulting hybrid is an allotriploid (3,18 ± 0.01) with two copies of the *Sc* genome fraction, confirmed by the heterozygous SNP frequencies, and one copy of the *Sk* subgenome. Interestingly, the only exception is chromosome III, in which the MAT locus is located. Read depth in the hybrid genome (Figure 2.4) shows that it contains one single chromosome III copy from *Sc* and another from *Sk*. This is confirmed by the analysis of heterozygosity in the *Sc* subgenome. In the other *Sc* chromosomes, levels of heterozygosity are identical to those found in the parental *Sc* chromosomes. However, in the artificial hybrid, the levels of heterozygosity of Sc chromosome III drop to 0, in accordance with the presence of one single copy. This result indicates that the parental diploid Sc T73 cell, involved in the hybridization event, acquired mating-competence not by becoming homozygous for the MAT locus due to gene conversion, but by becoming hemizygous for the MAT locus due to a chromo-

some III copy loss (monosomy). This is congruent with the fact that chromosome III is one of the smallest chromosomes and shows the highest loss frequency in *S. cerevisiae* (Kumaran, Yang, & Leu, 2013).

As the original *Sk* parent was diploid and the hybrid only contains one copy of the *Sk* genome fraction, two hypotheses could explain how this genome composition was generated in the hybrid. In the first hypothesis, the hybrid originated by a 'rare-mating' between two competent diploid cells and an immediate loss of one complete copy of the *Sk* subgenome after mating. The second involves a rare-mating between a competent diploid Sc cell and a haploid *Sk* spore/cell, which would require that *Sk* sporulation occurred in the rare-mating medium. To test whether *Sk* sporulation is possible under these conditions, we performed a 'rare-mating' experiment but only with *Sk* CR85, and after 5 days this culture was completely sporulated, which confirmed that the second hypothesis is the most probable.

To sum up, we confirm that artificial hybridization in laboratory conditions reproduces the most frequent genome architecture observed in natural hybrids, although the mechanisms to generate mating-competent Sc diploid cells were different.

2.4 Discussion

Hybridization between species has for years been an intriguing phenomenon for biologists. Due to the improvement in genome sequencing technologies, its importance in plants (Goulet et al.,

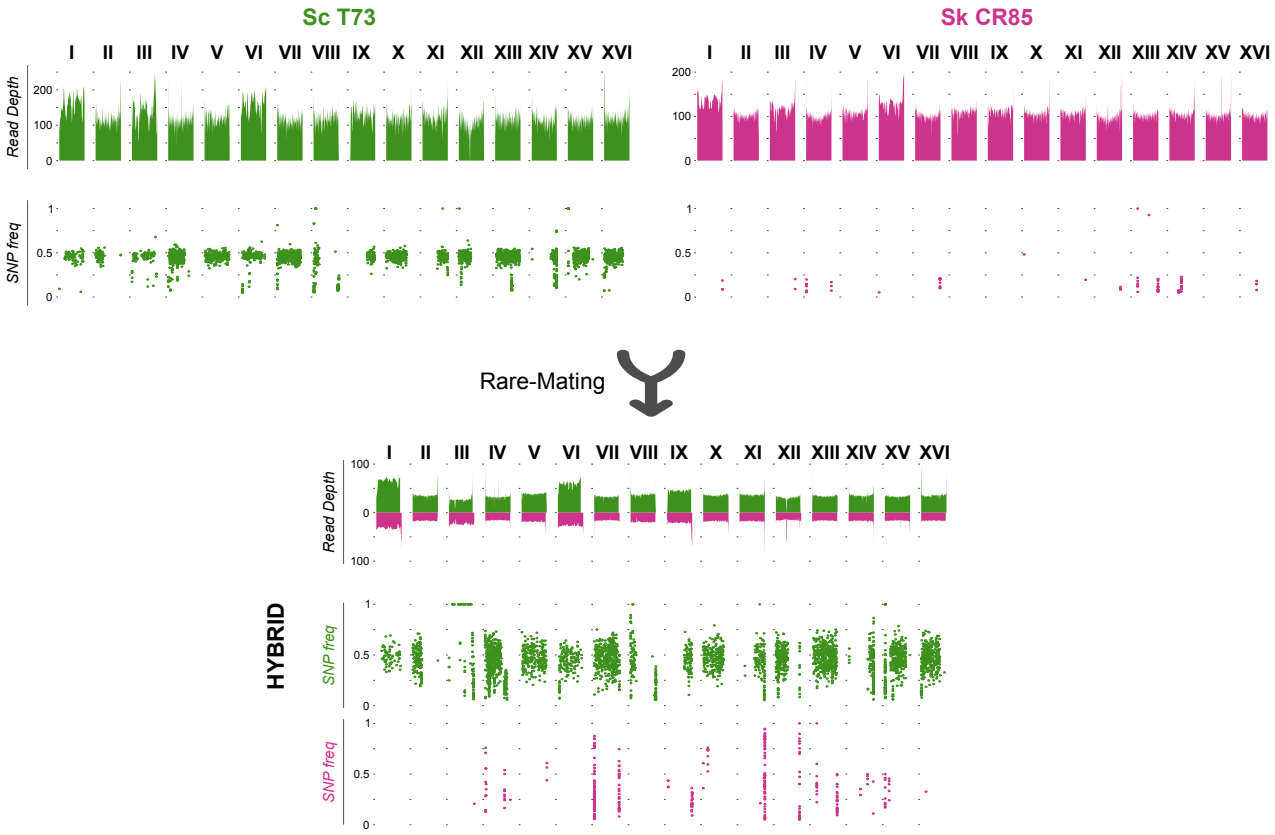

FIGURE 2.4 Genomes of the parental strains and the artificial hybrid obtained by rare mating. *Sc* subgenome of the hybrids is represented in green, and *Sk*, in pink. The Read Depth and SNP frequencies are represented along the genome. The *Sc* T73 is a diploid heterozygous wine strain with several LOH events in different chromosomes. *Sk* CR85 is also diploid but homozygous. The resulting hybrid has two copies of the *Sc* subgenome and one of the *Sk* subgenome and retains the heterozygosity with the LOH events of the *Sc* parental. Chromosome III only has one copy of *Sc* and one of *Sk* as confirms the LOH in the whole chromosome.

2017), animals and fungi (Feurtey and Stukenbrock, 2018) diversity and evolution are becoming clearer (Payseur and Rieseberg, 2016) . Hybridization between different species of the *Saccharomyces* genus was first suspected and confirmed in the lager beer *S. pastorianus*, also known as *S. carlsbergensis* (Nilsson-Tillgren et al., 1981; Vaughan Martini and Kurtzman, 1985), a hybrid between *S. cerevisiae* and *S. eubayanus* (*Se*). Since then, multiple hybrids between different *Saccharomyces* species were found principally in human-related environments, but also introgressions from different species were found in strains from natural habitats (Morales and Dujon, n.d.; Sipiczki, 2008). The clear ability of *Saccharomyces* yeasts to form viable hybrids makes them an interesting model for hybridization studies. Here we investigated the genome of different *Sc* x *Sk* hybrids from different isolation sources to decipher the mechanisms used to mate and the evolution of their genomes.

A first question to ask is if the different hybrids came from a unique hybridization event or multiple events have occurred. Previous work on *Sc* x *Sk* hybrids using six genes pointed out that there were different events that gave birth to these hybrids (Peris, Lopes, Belloch, Querol and Barrio, 2012*b*). Here we have used the whole genome of representatives of the different groups described by Peris et al. (2012) and we have clearly seen that the hybrids came from different hybridization events. The *Sk* subgenome showed that three groups could exist. This result is consistent with those obtained in the *Sc* subgenome phylogenetic analysis. Thus, VIN7, IF6, and AMH clustered together with the flor *Sc* strains, but PB7 and W27 are more related to other wine strain lineages, which clearly indicate that they derived from different hybridization events. This is also supported by the fact that the mating mechanisms involved in their origins differ (see below). Interestingly

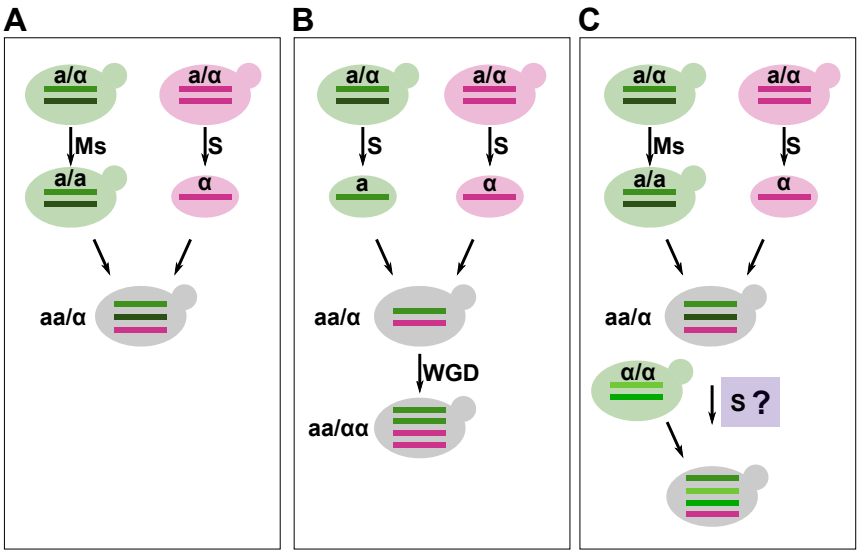

FIGURE 2.5 Models of the different mating mechanisms observed in the hybrids. In green are represented the *Sc* parentals or spores and in pink the *Sk*. Different shades of colors represent heterozygosity. *MAT* loci are assigned randomly and only as an example. Ms, *MAT* switch/loss of heterozygosity. S, sporulation. WGD, Whole Genome Duplication. **A**. Rare-mating. A diploid *Sc* becomes competent to conjugate by the conversion of the *MAT* locus. The *Sk* parental sporulates. The competent diploid Sc mates with an *Sk* spore to form a triploid *Sc* x *Sk*. This is the mechanism used by most of the strains in the study. **B**. Spore to spore cross and a subsequent whole-genome duplication, as observed in PB7. Both *Sc* and *Sk* sporulate, and spores mate to form a diploid hybrid. At some point, a whole-genome duplication occurs forming a tetraploid. **C**. Model for the formation of the AMH strain. A first hybridization event by rare mating occurred, as in A. Subsequently, the hybrid could have been forced to sporulate and mate with another diploid **Sc** to form a tetraploid with an extremely low **Sk** contribution.

the beer strain CECT11002 is clustered within the beer2 or mosaic beer group. The more common hybrids in lager beer are *Sc* x *Se* strains and their *Sc* subgenome clusters ale strains of the beer1 group (Gonçalves et al., 2016). The *Sc* x *Se* strains could come from a single *Sc* ancestor as they cluster together which differ from the *Sc* x *Sk* hybrids (Gonçalves et al., 2016; Monerawela and Bond, n.d.).

The *Sc* x *Sk* hybrids isolated and analyzed so far are mostly triploids with a diploid contribution of *Sc* and an haploid contribution of *Sk* (Borneman et al., 2012, 2016; Erny et al., 2012; Peris, Lopes, Belloch, Querol and Barrio, 2012*b*; Peris, Lopes, Arias and Barrio, 2012). This genome composition opened up the hypothesis that the mating mechanism used to hybridize was rare-mating but could not be completely confirmed (Erny et al., 2012; Peris, Pérez-Torrado, Hittinger, Barrio and Querol, 2017). The heterozygosity levels that these strains show on the *Sc* subgenome is, therefore, clarifying this question. We observe that, in the triploid strains, the heterozygosity levels are related to the isolation source and with the cluster of strains it belongs to. It was previously described that different populations of *Sc* have different heterozygosity levels (Gallone et al., 2016; Peter et al., 2018; Tilakaratna and Bensasson, 2017). CECT11002 and MR25 strains, which are related to beer strains, have similar heterozygosity levels as *Sc* beer strains. Wine hybrids have similar heterozygotic positions as wine *Sc* strains. Conserving these levels in the *Sc* subgenome supports that rare-mating was the hybridization mechanism used by these strains, as summarized in Figure 2.5A. This genome constitution is the most abundant in the strains sequenced here but also in the strains previously studied (Borneman et al., 2016; Erny et al., 2012) which could also be evidence that rare-mating is the most common mating mechanism in natural *Sc* x *Sk* hybrids.

The artificial hybrid generated by rare-mating ended up with a genome constitution similar to the most typical spontaneous hybrids. We used plasmids with antibiotic resistance to avoid the use of auxotrophies that could select for different genome composition or chromosomal aneuploidies. It was astonishing to find out that the mechanism used to become competent to conjugation in *Sc* was the complete loss of one copy of the chromosome III, which does not seem to happen in natural hybrids. Chromosome III has a high loss frequency in the laboratory (Kumaran et al., 2013). Chromosome loss is probably more frequent in a population in laboratory conditions than the loss of heterozygosity in the MAT locus. In natural conditions, aneuploidies can be detrimental and therefore less frequent in the population than MAT homozygous cells. This could explain why the same phenomenon is not observed in natural hybrids. What clearly stands out is that heterozygosity from a 2n wine *Sc* is conserved in the hybrid and that sporulation in *Sk* is possible even in high nutrient concentration media. These two processes together make hybridization between diploid *Sc* and haploid *Sk* more frequent than spore to spore in both laboratory and natural conditions.

Despite triploidy is the most usual ploidy in natural hybrids, some tetraploid strains have been found. Here we reported two tetraploid strains with two different stories. PB7 is tetraploid with a perfect diploid homozygotic *Sc* and *Sk* subgenomes. Moreover, it is a fertile hybrid. The extreme homozygosity in the *Sc* subgenome is not compatible with its wine origin. Such trait can only be explained if the genome content was doubled after a spore to spore hybridization, as recapitulated in Figure 2.5B. Another tetraploid and fertile hybrid was found between *Sc* and *Su*, the so-called S6U strain (Naumov et al., 2000). The strain EL1D4, an *Sc* x *Sk* hybrid, was also found to be tetraploid and homozygous in the *Sc* subgenome but its sporulation was not tested (Erny et al.,

2012). In both cases, it was postulated that the fertility and the homozygosity, respectively, were the result of autopolyploidization after the mating event (Naumov et al., 2000; Erny et al., 2012). Genome doubling to restore fertility is common in plants hybrid speciation (Alix et al., 2017). This mechanism was also responsible for fertility restoration in *Zygosaccharomyces parabailii*, a recently discovered fertile hybrid (Ortiz-Merino et al., 2017). A recent study in the laboratory found that becoming allopolyploid and recovering fertility in *S. paradoxus* hybrids could happen in less than 400 cell divisions (Charron et al., 2019). Here we show that, even if these are rare events, this phenomenon can happen in industrial environments in *Saccharomyces* genus.

The other tetraploid strain, AMH, have a completely different genome composition. In this case, the *Sc* subgenome is heterozygous, more than the rest of the wine strains and have three to four copies depending on the chromosome. The *Sk* subgenome is highly eroded, resulting in few regions of *Sk* remaining in the genome. AMH strain is a commercial strain. It is known that hybridization between different *Sc* strains was used to improve different traits. Some of these strains were not recognized as hybrids and some of them can even sporulate with low viability, as we see here with the strains MR25 and VIN7. Other commercial strains have a reduced *Sk* genome remaining in their genome, this is the case of the Maurivin EP2 (Borneman et al., 2011). We hypothesize that this can be the result of such an improvement program were a spore from a triploid hybrid was crossed with a diploid commercial *Sc* strain (summarized in Figure 2.5C), which could explain slightly higher heterozygosity in the *Sc* subgenome. If a hybrid was sporulated, meiotic recombination could have drastically reduced the *Sk* subgenome and the resulting spores could have only some chromosomes from *Sk* explaining such an important reduction in the *Sk* content of AMH.

Hybridization between *Sc* and other *Saccharomyces* species in wines or beer is thought to be adaptive to low-temperature environments (Belloch et al., 2008; Gibson et al., 2013; Krogerus et al., 2018; Ortiz Tovar, 2018; Pérez-Torrado et al., 2015; Pfliegler et al., 2014). Nor *Sk* or Se, which are more cryotolerant species than *Sc*, were isolated from industrial environments except in the form of hybrids with *Sc* (Pérez-Torrado et al., 2017; Peris et al., 2014). The *Sc* x *Sk* hybrids show a wide range of ability to grow at low temperature, but it was shown that the higher *Sk* content, the better the growth at low temperature (Ortiz-Tovar et al., 2018). In the hybrids studied here, we see that homeologous recombination is an important contributor to genomic diversity. Most of the changes observed are replacements of the *Sk* part by its homeologous *Sc* region that have arisen by mitotic recombination. Few of these regions are shared between strains what points out to a stochastic phenomenon and/or that selective pressures were different between them. This mechanism introduces phenotypic variation which could explain the differences observed between the strains (Ortiz-Tovar et al., 2018). Interestingly, our results also suggest that mitotic recombination reinforce the post-zygotic barrier, as the strains with a higher number of recombination events have no spore viability or cannot even sporulate. Genomic instability is one of the proposed mechanism of post-zygotic isolation (Dion-Côté and Barbash, 2017). Yeast does not need sexual reproduction in industrial environments. What here seems interesting is that genomic instability could be important to improve phenotypic variability and adapt to fluctuant environments at expenses of sexual reproduction. Hybridization could, therefore, be an interesting strategy to improve adaptability in two ways: heterosis, due to the differences between the two subgenomes, proteomes, metabolomes, and interactomes, as well as the increasing of the genomic instability.

Therefore, hybridization between *S. cerevisiae* and *S. kudriavzevii* is a recurrent strategy in industrial environments involving the fusion between the metabolic capabilities of the two species. *Saccharomyces* species can mate using different mating mechanisms, but rare-mating is the most commonly used. The mechanism used to mate determines the genomic structure of the hybrid and its evolutionary outcomes. The evolution of hybrid genomes is triggered by genomic instability and results in a wide diversity in genomic rearrangements. The stressful environmental conditions in industrial fermentations could make hybrid genomes to preserve those chromosome rearrangements of adaptive value (Dunn et al., 2013). Therefore, interactions between both parental genomes, proteomes and metabolomes, together with the harsh environmental conditions present during fermentation, determine the final composition of hybrid genomes. In the case of *S. cerevisiae* × *S. kudriavzevii* hybrids, their genomes are mainly characterized by the preservation of the *S. cerevisiae* subgenome and a progressive reduction of the *S. kudriavzevii* fraction.

CHAPTER 3

VIN7 : a case study of instability in interspecific hybrids between *S. cerevisiae* and *S. kudriavzevii*

3.1 Introduction

The main species used in industrial fermentative processes is *Saccharomyces cerevisiae*. The *Saccharomyces* genus is formed by *S. cerevisiae* and eight other species. *S. cerevisiae* and in some cases *S. uvarum*, are the only ones involved in wine, beer or cider production. Intriguingly, interspecific hybrids between *S. cerevisiae* and other *Saccharomyces* species are found in industrial environments. The most famous case of these is the lager beer yeast *S. pastorianus*, which is a hybrid between *S. cerevisiae* and *S. eubayanus*. In wine, hybrids are less commonly isolated than

in beer but *S. cerevisae* x *S. uvarum* and *S. cerevisiae* x *S. kudriavzevii* hybrids have been observed.

Hybrids have been shown to inhered properties from both parents. When looking at the different traits of the *Saccharomyces* genus, then Sc is the best fermenter but also the most sensitive species to cold temperature. As you move down throughout the phylogeny away from Sc and towards the Most recent Common Ancestor (MCA), the species become less tolerant to ethanol but more tolerant to cold temperature. It has been described that the hybridization of *S. cerevisiae* with one of its natural sister species could be the result of the selection to cold temperature. This has been observed in Northern Europe, where cold temperature could have favoured the hybridization between the good fermenter but the cold sensitive *S. cerevisiae* and the more cold tolerant *S. eubayanus* (in the case of beer) or *S. kudriavzevii* (in other fermentations) to get the best of both worlds in an interesting domestication event. In plants it was shown that polyploidy, both allo- and autopoliploidy, is an important phenomenon in adaptation to harsh environments and in the domestication process. *Saccharomyces* has also been observed with different ploidies. In natural conditions it seems that diploidy is the most stable ploidy in *Saccharomyces* but in industrial environments poliploidy is observed which also is thought as a mark of domestication of yeast by humans. Hybridization in *Saccharomyces* generally results in allopoliploidy.

An interesting fact of polyploids in *Saccharomyces* is that they are more prone to aneuploidy than diploids or haploids. In fact, in the strains sequenced to the date, aneuploidy is found in industrial strains and more in the polyploid strains. Also, many aneuploid hybrids were isolated. The recurrent effort in the last years to sequence industrial strains of *S. cerevisiae* and *Saccha-*

romyces hybrids is revealing that genome structure is much less stable than what was classically thought with an increasing number of evidence that it could be an important aspect of evolution and domestication.

In this study we sequenced the genome of a commercial *S. cerevisiae* x *S. kudriavzevii* hybrid strain : VIN7. This strain genome was analysed first by comparative genome hybridization and then by sequencing in different studies. Intriguingly the genome content reported was not matching between the studies as the origin of the strain was the dry yeast in one case and the mother strain in the other. We show here that the genome of VIN7 is not stable and that the instability change the phenotype of the strain.

3.2 Materials and methods

3.2.1 Yeast strains and culture media

The hybrid yeast *S. cerevisiae* x *S. kudriavzevii* VIN7 used in this study was isolated from a commercial dry yeast sample provided by Anchor Yeast (Anchor, South Africa). The strain was rehydrated and grown in GPY plates (4% glucose, 0.5% peptone, 0.5% yeast extract and 2% agar) at 25ºC. Glycerinated stock of this strain store at −70ºC was used also as starting material for culture onto a GPY plate, and the colonies obtained were re-plating in other GPY plates for 3-4 generations.

3.2.2 DNA isolation

The yeast strain isolates were cultivated in GPY medium (5 g/L yeast extract, 5 g/L peptone, 20 g/L glucose), at 25 °C for 24 h and DNA was isolated according to standard procedures (Querol et al., 1992).

3.2.3 Quantitative phenotypic analysis of stress response in wine yeast hybrid VIN7

The purpose of this study was to investigate the genetic basis for ethanol and temperature sensitivity and genetic instability of *S. cerevisiae* x *S. kudriavzevii* VIN7. For this purpose, we quantitatively investigated the phenotypic variation in stress response, by growth at 12 and 28 °C and under 10 %ethanol stress. Analysis of physiologically relevant growth variables results in a quantitative description of the hybrid strain response variation, which may be the starting platform to understand how the diverse genetic properties of VIN7 engender its phenotype.

3.2.4 Spot plate analysis for determination of strain temperature sensitivity

The spot plate technique was used to examine the effect of temperature on the growth of the wine yeast VIN7 on GPY agar. Starter yeast cultures were obtained by inoculating yeast strains into 10mL of GPY medium and incubating overnight at 25ºC. Tubes were centrifuged at 4000 rpm

for 5 minutes. The supernatant was discarded and pellets were resuspended in sterilised water. Cell suspensions were then prepared based on optical density (OD) determined using a spectrophotometer. Samples were diluted until to obtain an OD of 1 at a wavelength of 600nm. Subsequent 1:10 dilutions were carried out to prepare serially diluted samples. A 10µl volume of each dilution was spotted onto GPY agar plates in triplicate. The plates were then incubated in a static incubator at 12 and 28ºC for 7 days. Data was recorded by photographing spot plates.

3.2.5 Ethanol tolerance growth analysis

Ethanol tolerance of the different VIN7 was evaluated by performing growth essays in GPY (peptone 0.5%, yeast extract 0.5%, glucose 2%) medium with 10% (v/v) ethanol. Growth was monitored at 600 nm in a SPECTROstar Omega instrument (BMG Labtech, Offenburg, Germany). Measurements were taken every 30 min for 43 hours after a 20-second pre-shaking for all the experiments. All the experiments were carried out in sextuplicate. Growth parameters like starting OD, maximum OD, growth rate, and area under the curve were calculated using the R packages growthcurver. Visualisation of the replicates was done with growthcurver (Sprouffske and Wagner, 2016).

3.2.6 Sequencing, Assembly and annotation

The *S. cerevisiae* x *S. kudriavzevii* VIN7 genome was sequenced using 454 (shotgun and 8kb paired end) in combination with 300bp paired end Illumina technology. A de novo assembly

was carried out using MIRA v 3.4.1.1 (https://sourceforge.net/projects/mira-assembler/) and GS De Novo Assembler (Roche/454 Life Sciences, Brandford, CT, USA). The resulting assembly was corrected and manually edited using consed (Gordon and Green, 2013). The scaffolds were aligned to *S. cerevisiae* reference S288C and *S. kudriavzevii* IFO 1802 with the all genome aligner MUMer(Kurtz et al., 2004) and, using an in-house script, were ordered into chromosome structure.

Illumina Reads mappings were done on *S. cerevisiae* reference S288C and *S. kudriavzevii* IFO 1802 using bowtie2 (Langmead and Salzberg, 2012) . The annotation was carried out in three steps: 1) Annotation from the already published VIN7 genome (pubVIN7) (Borneman et al., 2012) was transferred to the new assembly. 2) *ab initio* gene prediction was performed with Augustus (Stanke and Morgenstern, 2005) to detect possible unannotated genes in pubVIN7. 3) The result of the two first steps was carefully and manually checked using artemis (Rutherford et al., 2000) and pubVIN7 re annotated as we detected two principal problems in the previous annotation : a) The presence of indels (an usual problem with 454 sequencing technology) derived in many CDS being removed. b) Genes containing introns were not annotated previously or only one of the exons was present.

3.2.7 Real time quantitative PCR detection

DNA from all yeast strains (hybrid VIN7, *S. cerevisiae* S288C and *S. kudriavzevii* CR85 was extracted following the procedure described by Querol et al. (1992). DNA concentration and purity was determined with a NanoDrop ND1000 spectrophotometer (Thermo Fisher Scientific Inc.) and genomic DNA integrity was checked by electrophoresis in 0.8% agarose gel.

Oligonucleotide primers for quantitative real-time PCR (ISOGEN Life Science, The Netherlands) were species specific and located in different chromosomes (III, VI, XI). We checked each chromosome with two pairs of primers designed (see Table 3.1) from comparison between sequences from S288C and CR85 and pubVIN7 (Borneman et al., 2012).

Real time PCR reactions were performed in triplicate for each sample, using LightCycler® 480 SYBR Green I Master kit (Roche Applied Science, Germany) in a LightCycler 480 System, following manufacturer instructions. Final primer concentration was 300 nM in a 10 µl reaction mixture. The amount of DNA used as template was 20 ng, and the amplification program had 40 cycles. All the amplicons had a size between 100 and 200 base pairs to ensure maximal PCR efficiency. We included positive and negative controls of amplifications in all plates and previously tested all primers to discard crossed-specificity. Data were analyzed using the AbsQuant/2ndDerivative-Max function of the LightCycler 480 software. Cp values found for the sequences from the two strains used as positive controls were among 18 and 20. Cp values higher than 35 were considered no amplification in our 40 cycles program.

3.2.8 Interactions and expression data analysis

Data from physical interaction were extracted from the SGD interaction dataset (https://downloads.yeastgenome.org/curation/literature/interaction_data.tab) All raw labeled as physical interaction were used to perform the analysis. Genetic interactions were downloaded from (http://thecellmap.org/costanzo2016/ ,Costanzo et al. (2016)). Expression data was extracted

TABLE 3.1 Primers used for qPCR These primers were used for the detection of the presence/absence of chromosomes III, VI and XI

Sp	Chr	SEQ	Forward primer 5'-3'	Length	Reverse primer 5'-3'	Length	Size bp
Sk	III	VIN7_5911	ACCTCTTCTCAAGGCTTGGC	20	AGGGGTGCATTTAGAATGGGA	21	119
		VIN7_5888	TGCGTCTTGTGCCAGTTGTA	20	GGAGAGCAGGTCAGGGTAGA	20	151
	VI	VIN7_6908	AAACGACGTATGCCGCAATG	20	CGCCTGATGAGAACCCTGTT	20	101
		VIN7_6891	ACGAAGAGAGAAAGCGTCAGG	21	AGCTGGCTCGACTTCTTCAC	20	102
	XI	VIN7_8307	ATTGGATATCCCTCCGGCAC	20	CTGCAACCATCACAATGCCC	20	101
		VIN7_8382	GATCCTATCACTGGCGCGAT	20	TCTGGCAAGTCCTTGGTGTG	20	150
Sc	III	VIN7_0475	TGCTACGGTGGTTCTGCAAG	20	ACCACTGTGTCATCCGTTCT	20	152
		VIN7_0509	TAATGGAGAGCTTCATGTCGGG	22	CCCTCAAGGATGTCACTAGCA	21	106
	VI	VIN7_1475	AGCATCCAGGGATTCTCACG	20	TCCAGTATCTTGGCCGATGTG	21	199
		VIN7_1573	ACACCGCCAAGCTTCCAATA	20	TTGCCACGCAAAGAAAGGAC	20	143
	XI	VIN7_2927	CTCTAGAACAGGCTGAGGCG	20	GGTCGTATGCCTGTAGACGG	20	129
		VIN7_2979	GCAACGGGCAAAAGCAAGAT	20	ACCACCTTCCCATTTCGGTC	20	124

from suppelemental material of(Keane et al., 2014; Mattenberger et al., 2017) . Each gene was given a 1 value if it was upregulated in the stress condition compared to the control condition, and 0 if it was either not differentially expressed or downregulated. To calculate the number of upregulated genes per chromosome each gene value in it was sum.

Statistical analysis on interactions and expression per chromosome was performed using a bootstrap analysis strategy. First, the number of genes per chromosome was calculated from the *S. cerevisiae* S288C genome. To create a random normal distribution, random genes from the genome were picked creating a set of the same gene number as the chromosome gene number avoiding picking twice the same gene. The mean number of interactions or upregulated genes was calculated for this random chromosome. This operation was repeated 10,000 times. Then the actual mean interactions or upregulated genes was calculated and the *pnorm* function in R was used to calculate the p-value or cumulative density function between the actual mean and the random normal distribution created.

A genetic interaction network between chromosomes was created with the **?** data. The interactions between genes of different chromosomes were summed. We considered that positive interactions would compensate negative ones. A network with the chromosomes as nodes was then drawn. An arrow would go from one chromosome to the other if the sum of interactions between them was the nearest to 0.

3.3 Results

3.3.1 The genome of VIN7

The wine yeast VIN7 obtained from a commercial dry yeast sample provided by Anchor was sequenced by paired-end and shotgun 454 and also a deep shotgun Illumina sequencing. The reads assembly yielded 206 contigs with an N50 of 295,173 bp and N90 of 82,169 bp, which were distributed in 106 scaffolds. The genome was an allotriploid between *S. cerevisiae* and S. kudravzevii but with certain recombination events and chromosome losses.

We detected six homeologous recombination events between the *S. kudriavzevii* and *S. cerevisiae* sub-genomes located on two different chromosomes, three on chromosome IV and three others on chromosome VII. Only one of these events involves the substitution of *Sc* genes by their *Sk* orthologues. More specifically, it is a 13-kb region located on chromosome IV between the genes YDL186W and DLD2/YDL178W (event a in Figure 3.1). The rest of the events are all substitutions of the *Sk* part by *Sc* sequence. On chromosome IV two events are present: first, a 15-kb between the genes NAT1/YDL040C and DBP10/YDL031W (event b inFigure 3.1) and on the other arm of the chromosome a 12kb between the genes FCF1/YDR339V and HXT3/YDR345C (event c in Figure 3.1). The first recombination on chromosome VII involves 26-kb from the beginning of the left arm of the chromosome until the gene HXK2/YGL253W (event d in Figure 3.1). In the middle part of the gene PMA1/YGL008C a tiny replacement of 1788-bp is observed (event e in

Figure 3.1). The biggest recombination seen is located on the right arm of chromosome VII where about 350-kb, from the gene SPT6/YGR116W to the end of the chromosome, *Sk* genes are replaced by their Sc homologues (event f in Figure 3.1).

In most of the genome we found contigs from both *S. cerevisiae* and *S. kudriavzevii* sub-genomes which is consistent with an allotriploid hybrid. Nevertheless, in the assembly no contig from chromosomes III, VI and XI of the *S. kudriavzevii* sub-genome were obtained. Mapping the Illumina reads on these three chromosomes from the *S. kudriavzevii* reference genome IFO1802 gave a mean coverage of 5.6x, 10.4x and 2.5x for chromosomes III, VI and XI respectively. The Illumina library contains 116,062,916 pairs of 100 bp reads. Considering a complete triploid genome containing both species genome, the expected mean coverage would be of 967x by haploid sub-genome. The coverage obtained for these chromosomes is clearly lower than expected but deep sequencing of the genome could find sequences from them indicating that these chromosomes are present at a low frequency in the population.

3.3.2 The genome is different from what was previously described

Borneman et al. (2012) published the sequence of the genome of VIN7 from the original mother culture using 454 technology. In this assembly the homologous recombination in the right arm of chromosome VII is not detected. Interestingly, the rest of the recombination are present. Moreover, the chromosomal losses described above are not observed in this assembly. In another work by the same group (Borneman et al., 2016), VIN7 genome is sequenced with Illumina. In this study, they

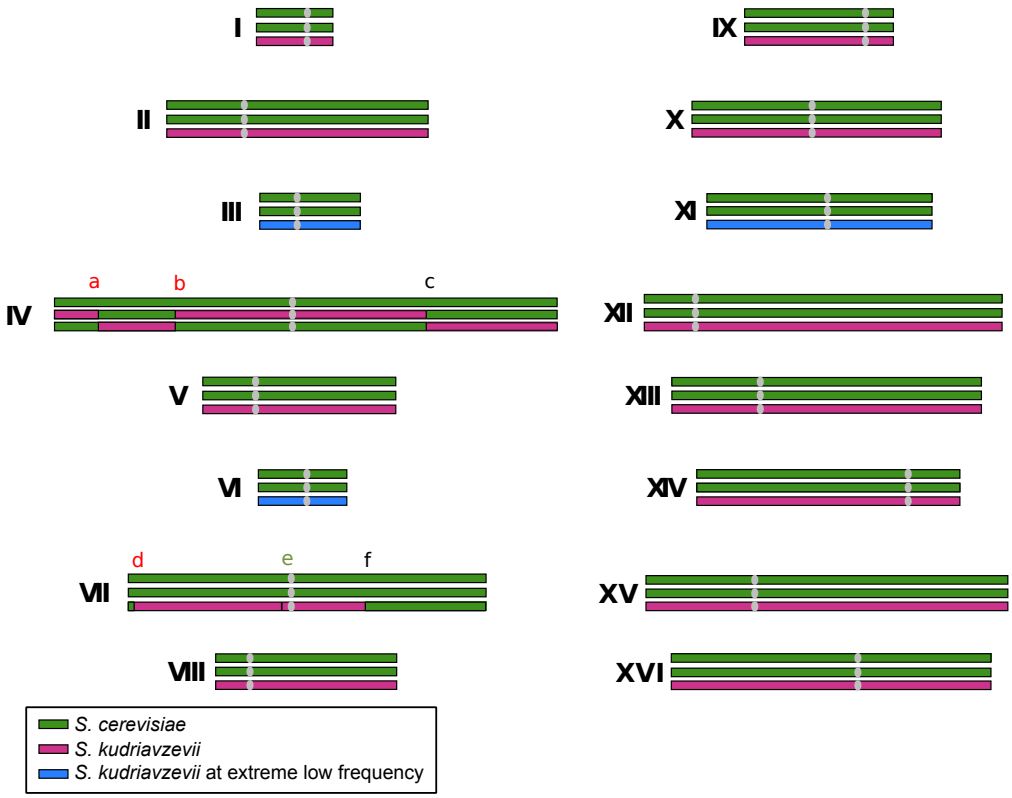

FIGURE 3.1 The genome structure of VIN7. In green Sc subgenome is represented and Sk in pink, in blue, the chromosomes that are lost. Homeologous recombination points are marked with letters.

confirm the genome structure described before. VIN7 is perfectly allotriploid and the chromosome VII recombination is not present.

An aCGH experiment reported the genome structure of VIN7 (Peris, Lopes, Belloch, Querol and Barrio, 2012*b*), in this instance the VIN7 population was obtained from a glycerol stock produced from a dried yeast stock provided by Anchor, as is the case of our sub-population. Coinciding with our results, the right arm of chromosome VII also seemed to be formed by three *Sc* copies. The loss of *Sk* chromosome III is also reported, however, *Sk* chomosome VI and XI were detected in the aCGH experiment. Despite both the sequencing and the CGH experiment were based on the same glycerol stock. The different genome structure in the different populations, points out that the genome of VIN7 is particularly dynamic.

3.3.3 Vin7 genome content is not stable and affects its phenotype

As different genome content was observed in the VIN7 strain depending on the source used to sequence it, the next question to answer was whether we could observe the changes in chromosome loss in growing colonies. We performed real-time quantitative PCR analysis for chromosomes III, VI and XI of *Sc* and Sk. Two genes from each chromosome were chosen in order to detect its presence. Two different growth conditions were tested. First, the rehydrated commercial dry yeast grown on a GPY plate at 25ºC and a glycerinated stock of this stored at -70ºC and plated at 25ºC for three-four generations.

The sequences used to detect the chromosomes were species specific and we could detect them in the *Sc* and *Sk* controls, except for one of the *Sc* sequence of chromosome XI that did not amplify in any populations (Table 3.2). The dry yeast showed amplification of all three chromosomes but *Sk* chromosome XI. This one seemed harder to detect and was probably lost in most of the cells. For the sub-population from the glycerol stock grown for a couple of generation all the *Sk* chromosomes were difficult to detect. *Sk* chromosome XI was not detected at all, indicating that it was definitely lost in the overall population. The other two *Sk* chromosomes were not completely lost but not detectable as in the dry yeast population. Altogether this points out that the genome content of VIN7 is dynamic and that loss occurred in a few generations.

Genomic instability could have a potential impact on the phenotype of the strain. Here we tested two relevant growth conditions for *Saccharomyces* hybrid to look for potential adaptive value of

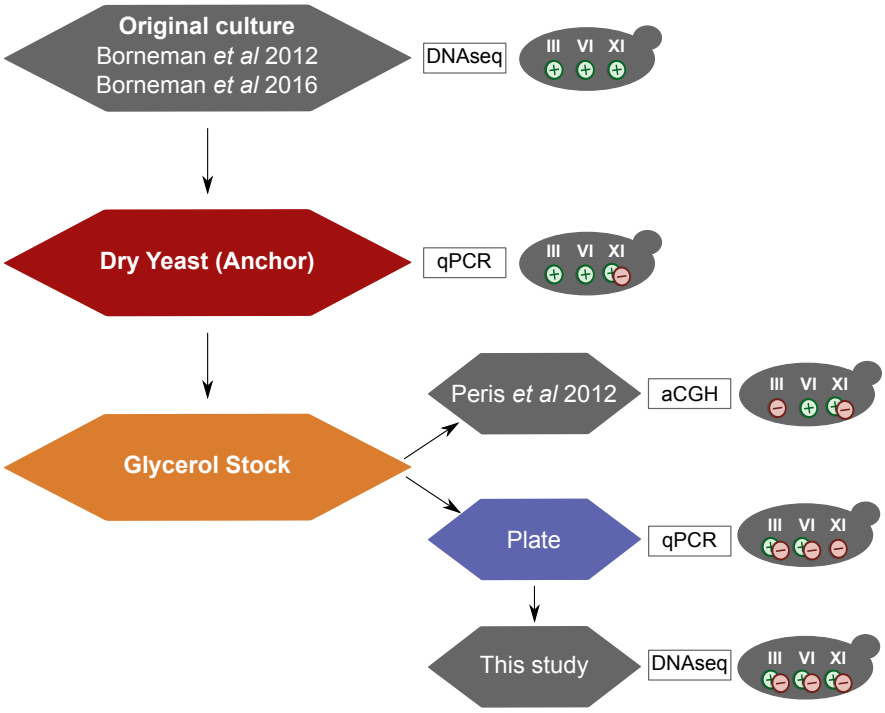

FIGURE 3.2 Different sub-populations of VIN7 have different chromosomal losses. Representation of the different observed presence absence of the chromosomes III, VI and XI in the different publications of VIN7. (+) : present; (-): absent; (+/-): low frequency or dubious.

TABLE 3.2 qPCR, including the results of the GPY plate with negative controls.

CHR	seq	S288c	CR 85	VIN7 Dry yeast	VIN7 Plate population
Chr III - Sc	0475	+	-	+	+
	0509	-	-	+	+
Chr III - Sk	5888	-	+	+	+/-
	5911	-	+	+	+/-
Chr VI - Sc	1475	-	-	+	+
	1573	+	-	+	+
Chr VI - Sk	6891	-	+	+	+/-
	6908	-	+	+	+/-
Chr XI - Sc	2927	-	-	-	-
	2979	+	-	+	+
Chr XI - Sk	8307	-	+	-	-
	8382	-	+	+/-	-

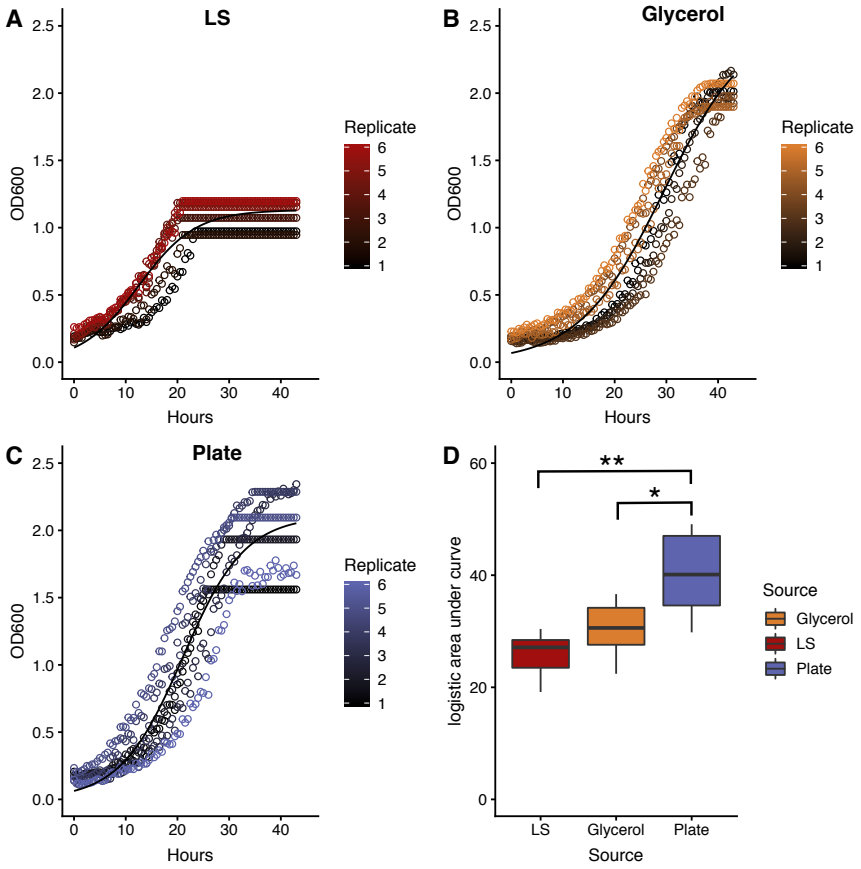

FIGURE 3.3 Phenotypic characterisation of the different sub-populations of VIN7 on ethanol. **A**. Growth of the dry yeast stock (LS). **B**. Growth of the Glycerol stock sub-population. **C**. Growth of the Plate sub-population. **D**. Area under the curve of the three sub-populations.

the chromosomal loss. Firstly, growth at low temperature is thought to be one of the advantages that the *Sk* sub-genome confers to the hybrids. Using spot-assay we did not see any differences in the growth of the VIN7 sub-populations (data unshown). The three different population show similar growth at 28ºC and also similar phenotype at 12ºC. Secondly, growth in 10% ethanol stress was tested. Looking at an overall comparison between the three VIN7 populations, we observed that the plate/3-4 generation grows significantly better than the other two, when using the area under the curve as a measure of growth (Figure 3.3). This indicates that the population with the lowest polymorphism of the three chromosomes is more tolerant to ethanol than the other two.

3.3.4 Chromosome aneuploidy pattern analysis

Many industrial *Sc* strains show aneuploidies with adaptive potential. Here we observed a more complex genotype. We asked if the chromosomes in themselves could have some special traits that could explain why these were lacking in VIN7. The first question is if certain chromosomes are more frequently aneuploids than others. As no sufficient *Sc* x Sk hybrids genomes are available, we used *Sc* as a model. 1011 genomes of *Sc* were recently sequenced (Peter et al., 2018). In this dataset 217 *Sc* strains were aneuploids for at least one chromosome. We counted the number of copies of each chromosome in these strains to then calculate the aneuploid frequency of each chromosome, defined as the frequency of appearance of abnormal copies of the chromosome in the aneuploid strains (Figure 3.4). As previously described, having an extra copy of a chromosome is more frequent than losing a copy of it. Interestingly, all chromosomes could be, at some extent, aneuploid. We could determine that the more frequently aneuploid chromosomes were I ($f = 0.33$), IX ($f = 0.27$), III ($f = 0.15$), VIII ($f = 0.14$) and XI ($f = 0.13$). The least frequently aneuploid chromosome was chromosome IV ($f = 0.05$). The chromosomes lost in VIN7 are among the frequently aneuploid ones, except chromosome VI, that has an $f = 0.06$.

Chromosomes have different chances of being present in abnormal copy number. This could be due to different characteristics of the genes present in those that could influence the possible detrimental or beneficial effect of aneuploidy for certain chromosomes. The change in copy number of a chromosomes unbalances the content of the proteins from the genes in this chromosome. Several

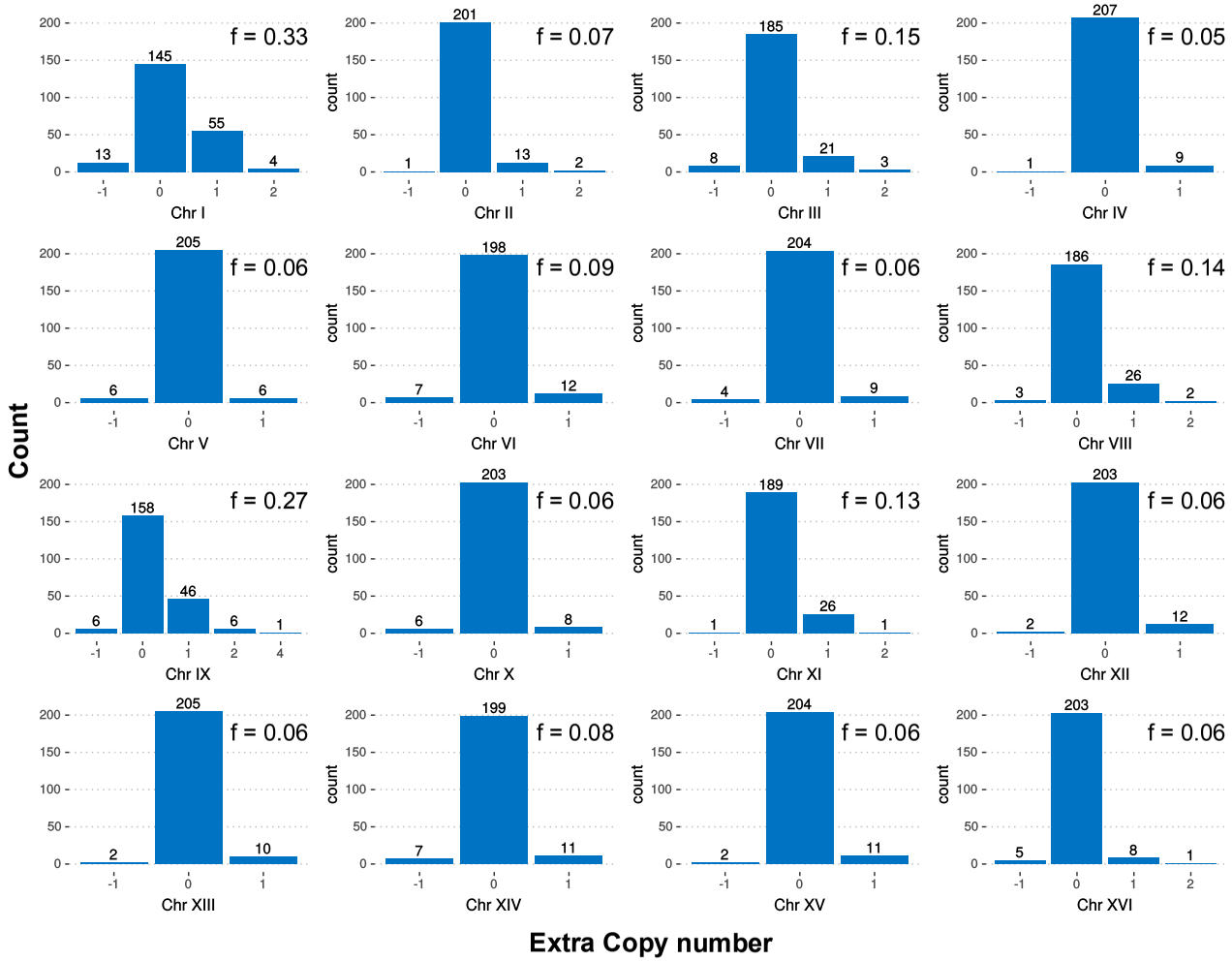

FIGURE 3.4 Aneuploidy frequency of the chromosomes in *S. cerevisiae*. For each chromosome we count the number of strains in Peter et al. (2018) that had euploid (0) or different copy numbers. The frequency was caclulated as the number of strains harboring abnormal copy number (independently of the gain or loss) divide by the number of strains harboring normal copy number of each chromosome.

characteristics could be important to select against or relax selection against aneuploidy. We looked at several characteristics of the genes that could influence the chromosomal dynamics. Physical interactions between proteins are essential for the cell. Changing the protein content of interacting partners can be detrimental and therefore needs to be regulated, and could be an important negative selection factor for aneuploidy. Epistasis is another factor to consider. If a chromosome contain highly connected genes in the interaction network, the effect of aneuploidy could be bigger. On the other hand, if the genes in the chromosomes are less connected the rewiring of the interaction network would be more tolerable. A possible positive effect of changing the chromosome copy number is if it contains genes that responds to a certain stress. We looked at different RNA-seq experiments to search for the upregulated genes in four stresses: lactic acid, ethanol, glycerol and oxalacetate.

To decipher if a pattern was recurrent in the chromosomes, we performed a bootstrap analysis. We generated 10,000 chromosomes of the same gene number as the actual chromosomes but with random arrangement of genes as normal distribution. We calculated the actual mean number of interactions or upregulated genes for each stress condition and then calculated the integral of the distribution at the left of the actual value. This value is low if the actual mean is lower than what would be expected by chance and tends to 1 if its higher than expected. The heatmap in Figure 3.5A represents all the features analysed for each chromosome. Interestingly no clear pattern emerges. None of the interactions and stress expressions seem to be clearly associated with the aneuploidy frequency. The chromosomes III, VI and XI have different kind of content that could be of relevance for the loss in VIN7 but are not shared between them.

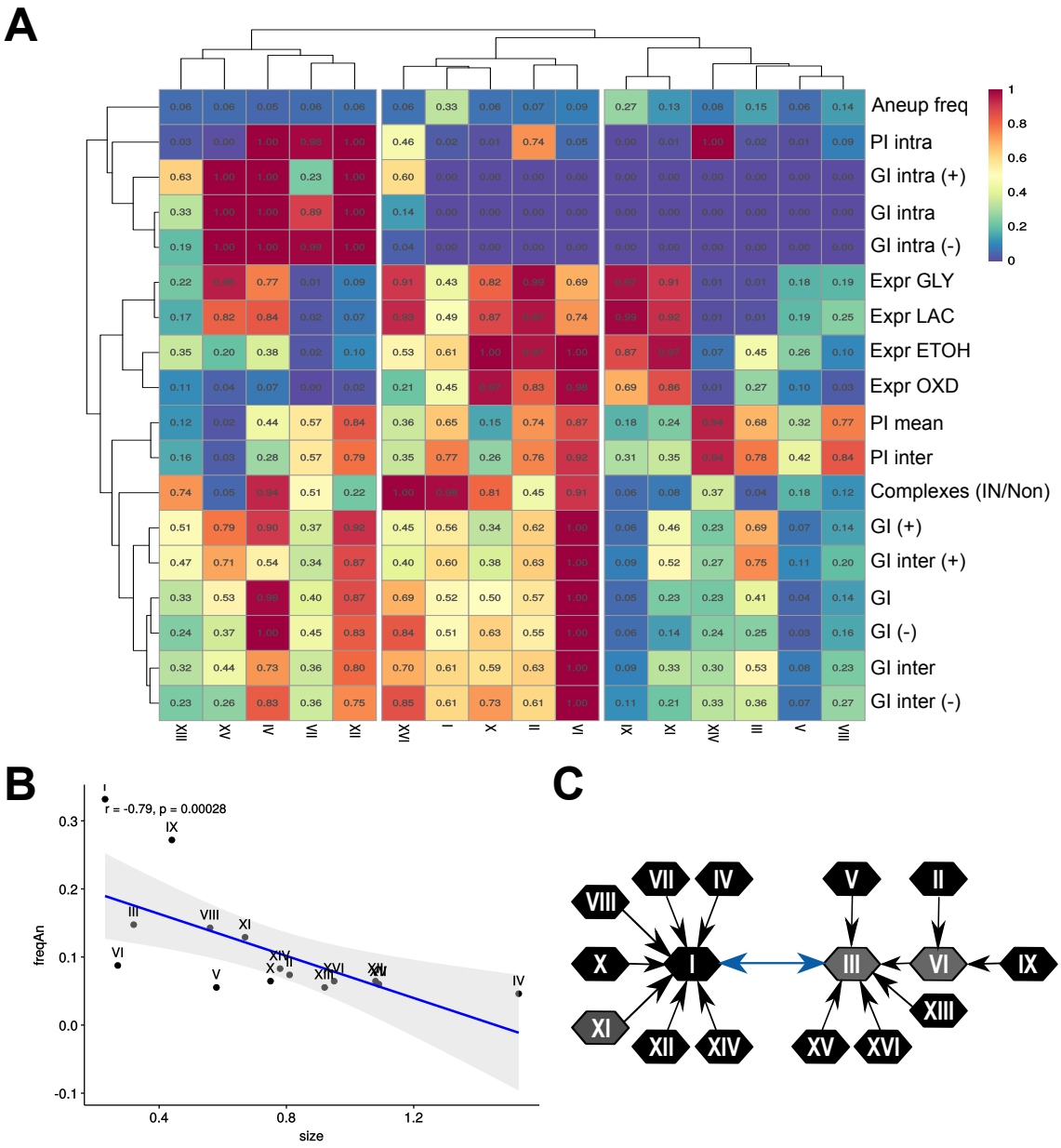

FIGURE 3.5 Factors affecting chromosome aneuploidy. **A**. p-value heatmap of the characteristics analysed for each chromosome. Values near to 1 mean higher than expected and near to 0 lower than expected. *Aneup freq*: aneuploidy frequency; *PI intra*: intra-chromosomal physical interactions; *GI intra (+)*: positive intra-chromosomal genetic interacitons; *GI intra*: intra-chromosomal genetic interactions; *GI intra (-)*: negative intra-chromosomal interacitons; *Expr GLY*: expression on glycerol; *Expr LAC*: expression on lactate; *Expr ETOH*: expression on ethanol; *Expr OXD*: expression on oxaloacetate; *PI mean*: mean physical interactions; *PI inter*: inter-chromosomal physical interactions; *GI (+)*: positives genetic interactions; *GI inter (+)*: positive inter-chromosomal genetic interactions; *GI*: genetic interactions; *GI (-)*: negative genetic interactions; *GI inter* : negative genetic interactions; *GI inter (-)*: negative inter-chromosomal genetic interactions **B**. Aneuploidy frequency of the chromosomes compared to the chromosome size. **C**. Less interacting chromosomes network. The arrow between chromosomes represent the chromosome with which the interaction tends to be neutral.

Another possible influence on the tolerance of aneuploidy of certain chromosome is the chromosome size. If changes affect smaller chromosomes, the effects on the number of proteins affected, the DNA content change or the interactions affected are smaller. We found a negative correlation between chromosome size and the aneuploidy frequency ($r = -0.79$, $p = 0.00028$, Figure 3.5B). In VIN7 chromosomes III and VI are lost and also are among the smallest chromosomes. Interestingly chromosome VI is not in the most frequently aneuploid even if it is one of the smallest, pointing out that different characteristics, or a combination of them, can be important.

To have different aneuploidies at the same time is a complex genotype. We asked if interactions between chromosomes could explain if certain chromosomes could appear in abnormal number together as the ones seen in VIN7. To do so we calculated the number of interactions between genes of each chromosomes. For each chromosome we search which other chromosome had the least number of interactions with. With this information we draw a network in which each node is a chromosome and a directed arrow pointed to its least interacting partner (Figure 3.5C). An interesting pattern emerged. The network showed two principal hubs, the first one centred on chromosome I and the second one on chromosome III. Moreover, chromosome I and III are connected in both directions as one is the least interacting to the other and vice versa. Chromosome VI is connected to chromosome III but it also is a centre of the network. The chromosomes at the centre are the less connected in the interaction network which would mean that the effects on the interaction network of the cell would be less detrimental. It is noteworthy that two of the three aneuploid chromosomes in VIN7 are central hubs of the network. It also is to notice that both chromosome I and III are frequently aneuploids in *S. cerevisiae* strains.

3.4 Discussion

Hybridization in the *Saccharomyces* genus is recurrent in industrial environments. With hybrids inheriting properties from both parental species means that this phenomenon is increasingly used by companies to improve the characteristics of wines and beers. The genome evolution of hybrids is still an open question. Here we sequenced the genome of the commercial *Sc* x *Sk* hybrid VIN7 in order to decipher its genome structure and dynamics. The VIN7 genome has been reported several times in different publications, first using CGH and then with two different sequencing technologies. All agreed that VIN7 is allotriploid with two copies of *Sc* and one of the *Sk*, cohered with what we observe here. However, two differences were detected between the sequencing experiments and the CGH analysis. In Peris et al, part of the *Sk* chromosome VII was replaced by Sc and *Sk* chromosome III was lost compared to Borneman et al. We observed more different in our genome compared to what was reported in other experiments. In addition to what was found in Peris et al, and using the same glycerol stock population, *Sk* chromosomes VI and XI were apparently lost in the majority of our sequenced population. Moreover, we could determine that the changes in the karyotype of the VIN7 occurred in a short number of generations. This points out to a highly dynamic genome.

We here looked if the different aneuploid variants of VIN7 presented differences in phenotypic traits. Even if only two growth condition were tested, we observed that ethanol tolerance was different between them, confirming that phenotype could change with the genomic content and

in very few generations. In agreement with this, it has been shown in *S. cerevisiae* that different aneuploidies could change the phenotype of strains and be of relevance for industry . In hybrids, a study that analysed karyotypic variants of an *S. pastorianus* strain , a hybrid between *S. cerevisiae* and S. eubayanus, showed that these brewing hybrids also have a dynamic genome. Interestingly, they found that these variants had different phenotypic characteristics. Similar results have been observed in a study with artificial hybridization, where it was suggested that the general ploidy of hybrids could affect the fermentation characteristics and the production of flavour components in beer. These evidences point to two important features. Firstly, it emphasizes the importance of genome stability of the strains in the possible outcome of industrial fermentation. If changes can occur in so few generations it is of relevance to further study what could be the factors that trigger it and the connection it has on the stability in the final product characteristics. Secondly, it would be of interest to better understand the relevance of such a dynamic genome in the domestication process of yeasts.

Aneuploidy in yeast have different outcomes. It can slow down growth dramatically, hence decreasing cell fitness in certain conditions (Torres et al., 2007). However, evidences for advantages of changes in certain chromosome copy number to increase fitness in stress conditions are increasing, here among increase ethanol tolerance (Morard et al., 2019), heat stress (Yona et al., 2012) and to other industrial related stresses (Gorter et al., 2017). We investigated the frequency of aneuploidy for each chromosome in a large dataset of *S. cerevisiae* strains (Peter et al., 2018). We looked at different features of the gene content of the chromosomes and observe, as other authors (Gilchrist and Stelkens, 2019; Peter et al., 2018), that the most related characteristics is the chromo-

some size. Other characteristics as interactions or expression of genes does not seem to explain the aneuploidy frequency. We therefore hypothesis that the higher frequency of smallest chromosomes is due to a lower effect on DNA replication stress and on the number of genes affected. If this is so then the fact that these are observed more frequently is a by-product of what is important in these strain: genome instability.

With genome instability complex patterns of aneuploidy can arise. It has been shown that in cells with induced chromosomal instability the pattern of chromosomal aneuploidies that emerge is not random but governed by interactions between chromosomes(Ravichandran et al., 2018) . Other studies showed that chromosome VI aneuploidy is more stable if it does not appear alone (Torres et al., 2007). It has also been shown that specific karyotype can lead to more instable genomes and others are relatively more stables (Zhu et al., 2012). We found that chromosome I and III have the least interactions impact which could be an explanation to their high aneuploidy frequency. These would be less detrimental and therefore appear more frequently in the population, but not necessary a reflection of selection. The appearance of aneuploidies would be stochastic and more frequent in industrial strains as certain changes in dose would permit a faster adaptation, on top of a broader phenotype landscape of the population attributed to a higher genome instability.

In industrial environments both hybrids and aneuploid strains appear more frequently than in nature. Moreover, higher ploidies are more frequent in these environments than in “natural” ones (Peter et al., 2018) and aneuploidies are more tolerated in in poliploids than in lower ploidies. A recent study claims that both hybridization and aneuploidy are adaptation mechanisms to perturbed

environments (Gilchrist and Stelkens, 2019). We here presented a case study of an allopoliploid strain with an instable genome that broaden its phenotypic space within a short time. In conclusion, hybrid genome instability can be an interesting adaptation mechanism that improve phenotypic diversity allowing to colonize their environment. More needs to be studied on the mechanisms and interactions that mould complex aneuploid genome but it becomes clearer that this can be an interesting a widely used adaptation mechanism.

CHAPTER 4

Genomic and transcriptomic analysis of short-term evolution of artificial hybrids between *S. cerevisiae* and *S. kudriavzevii*

4.1 Introduction

Hybridization has been used consciously and/or unconsciously by humans when selecting for improved traits in different organisms. This is particularly seen in agriculture, where hybridization – or crossing as it would often be rereferred to in this field – has been used for improving or generating new crops. An example of this is the cultivated strawberry which is an octoploid hybrid between two wild octoploids species, themselves products of ancient hybridization events (Edger et al., 2019). In yeast, the importance of hybridization in both evolution and domestication is now

becoming clear. The whole genome duplication event giving rise to the *Saccharomyces* ancestor was recognized as a hybridization event (Marcet-Houben and Gabaldón, 2015) or domestication gave rise to the hybrids of chapter 2, to cite two examples. Hybridization is also used recurrently in laboratory to improve *Saccharomyces* strains for commercial purposes(Gibson et al., 2017; Su et al., 2019; Pérez-Torrado et al., 2017; García-Ríos et al., 2019; Pérez-Través et al., 2012).

When hybridization event occurs, two distant genomes come together for the first time after a long period of diversion. As observed in industrial hybrids (chapter 2, Borneman et al. (2016) genomic recombination can occur between these two distant genomes. These changes have been shown to be responsible for fitness improvement in certain stress conditions (Pérez-Través et al., 2016; Origone et al., 2018; Krogerus et al., 2015). The use of artificial hybridization in yeast is a simple model to investigate what kind of changes occur just after the moment of the hybridization. Several studies looked at different important questions regarding hybridization using different *Saccharomyces* species (Zhang et al., 2019; Lancaster et al., 2019; Smukowski Heil et al., 2019). Nevertheless, most of them used spore to spore mating to generate the hybrids and we saw in chapter 2 that this is not the most common mating mechanism used outside of laboratory conditions.

Here we used a *Sc* x *Sk* hybrid obtained by rare-mating to perform an experimental evolution in three conditions: cold temperature, ethanol and control. The conditions were selected with the reasoning, that one of the stress conditions may favour one or the other sub-genome, as *Sc* is known to have a higher temperature and ethanol tolerance than *Sk*, which on the contrary is cold tolerant (Salvadó et al., 2011; Arroyo-López et al., 2010). We evolved the hybrid for 400 generations and

performed genomic and transcriptomic analysis to evaluate the evolution condition effect on these allotriploid hybrids.

4.2 Materials and methods

4.2.1 Experimental evolution of artificial hybrids

The procedure is summarized in Figure 4.1. A colony of the artificial hybrid obtained in chapter 2 was inoculated in an eppendorf tube with 1 mL of liquid GPY as a starter cultivate. The experimental evolution was carried out in three different conditions with three replica per condition: control (GPY at 25ºC), cold (GPY at 12ºC) and ethanol (GPY at 25ºC with 10% ethanol). The first step to begin the evolution was set in an eppendorf tube with 1mL of the corresponding medium inoculated with 2x106 cells/mL. After seven days of incubation on agitation, an aliquot was then plated on solid medium of the corresponding condition. The three first colonies were then selected to continuing the evolution. Tubes with 5mL of liquid medium were inoculated with the mix of the three colonies at a concentration of 2x106 cells/mL and incubated for seven days in each condition. After this, an aliquot was plated in solid medium and the three first colonies selected to continue the evolution. This process was repeated 16 times. Each step is of approximately 25 generations: around 5 in liquid medium and 20 in solid medium.

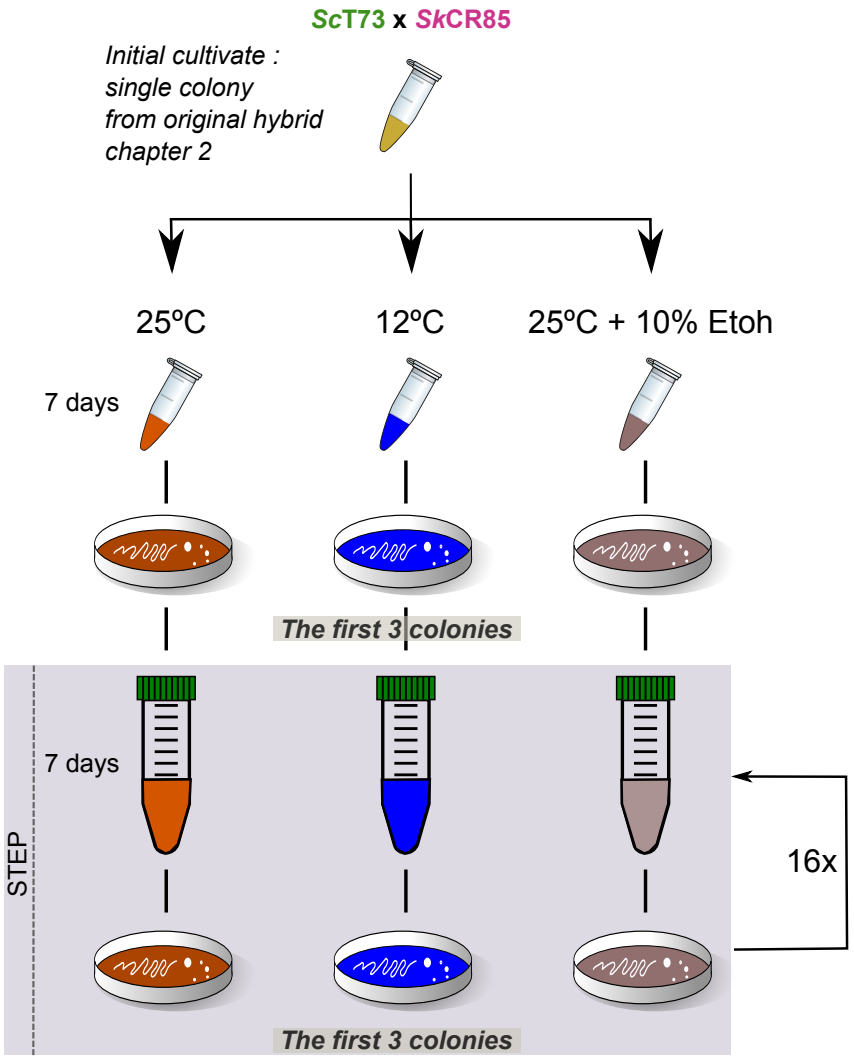

FIGURE 4.1 Schematic representation of the experimental evolution experiment. The grey box represent one step of evolution. This was repeated 16 times, representing aproximatively 400 generations.

4.2.2 DNA extraction and sequencing

At steps 4, 9 and 16 of the experimental evolution DNA from the three colonies selected to continue the process was extracted according to the protocol described by Querol et al. (1992). The genomes were sequenced with paired-end libraries of 150 nt on an Illumina NextSeq instrument.

4.2.3 Genome analysis

The reads were trimmed with Sickle v1.2 (Joshi and Fass, 2011) using a minimum quality per base of 28, and filtered with a minimum read length of 100 nt. Mapping of the sequences was performed with bowtie2 v2.3.0 (Langmead and Salzberg, 2012)on a concatenated reference of the parental strains *Sc* T73 and *Sk* CR85.

The analysis of genome content was performed as follows. Read depth (RD) was computed with bedtools v2.17.0 (Hung and Weng, 2016). Mean RD in 10-kb sliding windows of 1 kb steps was calculated for the representation of the complete genome coverage. Using the annotation, read depth of each gene was then calculated. The ratio of *Sc*/*Sk* content was calculated for each homeologous genes present in both subgenomes using the following equation:

$$R_{gene} = \frac{\overline{RD_{Sc}}}{(\overline{RD_{Sc}} + \overline{RD_{Sk}})}$$

This ratio tends to 0 if only *Sk* is present and to 1 if *Sc* is the only allele. The ratio at the end of

the evolution at 25ºC was used as the control condition. The deviation ratio ($devR$) is therefore defined as the difference between the ratio at the condition of interest and the ratio at 25ºC, at step 16:

$$devR_{gene} = R_{condition} - R_{25C,step16}$$

A variant analysis was performed with freebayes v1.1.60 (https://github.com/ekg/freebayes) with the parameters: -pooled-continuous, -F 0.15 -C 10 -E -1. Indels were removed and vcftools v0.1.13 was used to filter SNPs with a quality lower than 200. The frequency of each SNP was then calculated in all the samples. To analyze the differences at the end of the evolution the frequency of each SNP in the sample of the 25ºC evolution at step 16 was subtracted to the frequency at step 16 of the other evolution conditions. To analyze the changes of SNPs frequency throughout the evolution in each condition we performed a clustering analysis. First, SNPs frequencies of each sample were transformed to categorical states: homozygous absent ($f \leqslant 0.1$), heterozygous ($0.1 < f < 0.9$), homozygous fixed ($f \geqslant 0.9$). For each evolution condition clustering on the state in the three steps was performed with NbClust v3.0 (Charrad et al., 2014) using the euclidean distance and the Ward method. This clusters SNPs in each condition in the way they behave along the evolution. For example, one cluster would be: heterozygous at step 4, heterozygous at step 9, homozygous fixed at step 16, meaning that the SNPs in this cluster were heterozygous and are fixed at the end of the evolution experiment. Letters were assigned to the clusters to compare the different conditions. These clusters allowed to further filter SNPs depending on their comportment in the different evolution conditions. As an example, SNPs that are fixed in cold temperature at the end of the evolution and

heterozygous in the first two steps will only be kept in the analysis if they remain heterozygous in all steps of the evolution in the other conditions. Another behavior would show that these SNPs are in regions prone to errors and therefore unreliable.

4.2.4 RNA extraction and sequencing

The hybrids evolved at 12ºC and 25ºC at step 4 and step 16 of the experiment were used to perform growth experiments in the three conditions: GPY at 12ºC, GPY at 25ºC and GPY with 10% ethanol. All the experiments were performed by triplicate. Fermentations were performed in 100 mL bottles with 80mL of GPY or GPY supplemented with 10% ethanol and incubated at 25ºC and 12ºC. Sample extraction was realized at the middle of the exponential phase. Cells were harvested by centrifugation, and then stored at -80ºC. Total RNA was then extracted by the phenol-chloroform procedure. Samples were treated with phenol-tris, phenol-chloroform (5:1) and chloroform-isoamyl alcohol (24:1), and two precipitations with LiCl and ethanol with sodium acetate respectively. Samples were then sequenced on an Illumina HiSeq instrument and 150nt paired-end reads were obtained.

4.2.5 RNA-seq analysis and differential gene expression

Reads were mapped to the same reference as above with bowtie2 v2.3.0 (Langmead and Salzberg, 2012). Reads with alignment quality score lower than 2 and aligned multiple times were removed. For each gene, read count was performed with HTSeq-count v0.9.0 (Anders et al., 2015)

with parameter -m union.

To perform the ratio analysis, the transcript per million (tpm) was calculated for each gene. Then the tpm of the *Sc* allele was divided by the sum of the tpm of the *Sc* and *Sk* allele, similarly as done to obtain the genome ratio. To compare the ratio in the transcriptome and the genome content, for each gene the DNA ratio corresponding to the evolution step of the RNA sample was subtracted to the RNA ratio. Differential gene expression analysis was performed with DESeq2 (Love et al., 2014). Principal component analysis (PCA) was performed using the variance stabilizing transformation dataset of $log_2 fold$ normalized data given by the package. Given the results of the PCA analysis, only one paired comparison was performed. The growth at cold temperature of the hybrid at step 4 of the cold evolution was compared to the hybrid at step 16 using the Wald test. Benjamini-Hochberg correction was used to adjust $p - values$ for multiple comparisons. Functional analysis and GO term enrichment were performed with FunSpec (http://funspec.med.utoronto.ca/), with the Bonferroni correction and $p - value$ cutoff of 0.01 parameters. Analysis of transcription factors explaining the expression of the significant differentially expressed gene was performed with PheNetic (http://bioinformatics.intec.ugent.be/phenetic/#/index) (De Maeyer et al., 2013) with default parameters, the yeast interaction network obtained from the program, and the upstream option.

4.3 Results

4.3.1 Changes at the nucleotide levels are mostly due to LOH

In this work, we focused on the early evolution of an artificial *Sc* x *Sk* hybrid in different conditions. The hybrid obtained in chapter 2 was evolved in conditions thought to favor one or the other sub-genome to untangle the early changes that occur. The hybrid obtained by rare-mating between a wine diploid heterozygous *Sc* (T73) and a diploid homozygous natural *Sk* strain (CR85) was triploid with two copies of *Sc* and one of *Sk*. As previously shown (Figure 2.4) this genome architecture is explained by the mating of a spore of *Sk* with a mating competent diploid *Sc*. The competence to rare-mating was acquired by the loss of one copy of the chromosome III in the *Sc* population and is observed in the original hybrid and evolved strains.

The hybrid was evolved in three conditions: control, cold and ethanol. We considered the evolution at 25ºC as a control condition that should be the nearest to a balanced environment in industrial conditions. To understand what and where changes occur at cold temperatures, another line was evolved at 12ºC which is a condition in which *Sk* grows better than *Sc* (Salvadó et al., 2011). Ethanol tolerance is higher in *Sc* than in *Sk* (Arroyo-López et al., 2010) , to understand how hybrids will be affected by ethanol the evolution was carried out at 25ºC with 10% of ethanol. For each of the evolution line, DNA sequencing was performed at step 4, 9 and 16. In total, the evolution experiment covers approximately 400 generations.

SNPs detection in heterozygous yeast and polyploids is challenging and subjected to a high false discovery rate. To cope with it we calculated the frequency of each SNP detected in each sample; then calculated the deviation from the control condition (25ºC) subtracting the frequency at step 16 at 25ºC to the frequency in cold and ethanol evolution at step 16. This represents the differences at the end of the evolution between conditions. Plotting the deviation of both conditions one against the other permits to observe if the frequency change of the mutation is condition-specific (Figure 4.2).

First, we looked at which subgenome was affected by the variants. In Figure 4.2A we can see that most of the variants are observed on the *Sc* subgenome. As the *Sc* parental strain was diploid and heterozygous, most of what is detected are clouds of points around +/- 0.5 in one of the conditions. We further investigated these regions and showed that they are from the same chromosome (Figure 4.2B), representing Loss of Heterozygosity (LOH) events. In the ethanol evolution, a LOH event on *Sc* chromosome XVI is seen as well as a small one on chromosome VII. We also observed a LOH on the *Sc* chromosome VII in the control evolution at 25ºC. Contrary to this, we observed no LOH in the cold evolution. On the *Sk* subgenome, really few changes are observed. As its contribution is haploid only frequency changes of +/- 1 should be present. Only one of these is present in the cold temperature evolution and few in the other lines. SNPs on the *Sk* subgenome at intermediates frequency changes will most likely be errors and false positives.

In addition to investigating the changes at the end of the evolution experiment, we aimed to observe when these changes occur throughout the experiment. We classified the variants observed

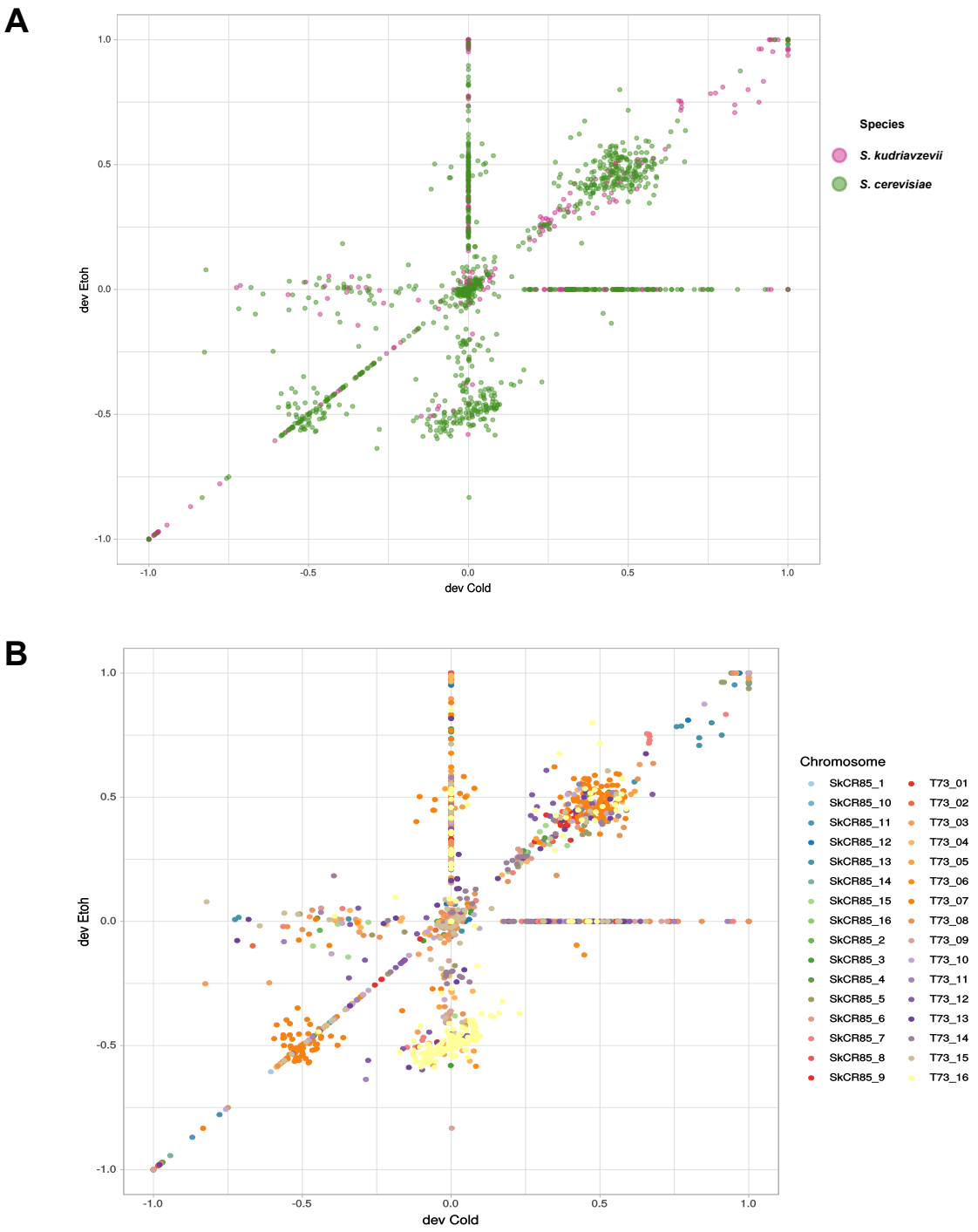

FIGURE 4.2 SNPs at the end of the evolution. The deviation is calculated as the difference between the frequency of each SNP in the cold (devCold, axis x) or the ethanol (devEtoh, axis y) and the frequency in the control evolution. devCold and devEtoh are plotted one against the other. The points horizontally plotted are then SNPs which frequency change in the cold evolution only. The points along the vertical axis are changing exclusively in the ethanol condition, and the SNPs on the oblique line, in the control condition. **A**. The colour of the points represents the subgenome. **B**. SNPs are coloured by chromosome.

in each condition using a clustering method that represented its variation along the evolution (see Materials and Methods). We grouped all of them in 20 clusters representing each a different comportment. We had two purposes with this analysis: filter the variants based on the reliability of their comportment and better understand the timing of their appearance. With this clustering, we could filter the variants selected to be condition-specific in Figure 4.2 depending on how they appear in the other conditions. Then these SNPs were filtered and only the ones that had a comportment reliable in all the experiment were further considered. In Figure 4.3 we represent the frequency of the selected SNPs in the condition of interest along the evolution. As previously shown, in the cold evolution really few changes are observed at the nucleotide level. From the 46 selected variants only 26.1 % (12) affected the *Sk* subgenome. This suggests that in the cold evolution, in this study, the *Sk* subgenome does not accumulate more changes than the *Sc*. Among the 235 SNPs we considered valid in the ethanol experiment, 162 are changes in the frequency of heterozygotic variants and more specifically, 106 of these variants correspond to a LOH event on *Sc* chromosome XVI. Only 12,7 % (30) of the valid SNPs were detected on the *Sk* subgenome, which is an even small proportion than in the cold temperature. Interestingly, the changes observed during both evolutionary experiments seems to happen between step 9 and 16. Specifically, 74% of the variants in the cold evolution and 86% in the ethanol occurred between these two time points.

Altogether we observed that the fixation of new point mutations are rare in the evolution experiments in this hybrid, as expected. LOH events are the most important phenomenon and affected the evolution at 25ºC and in ethanol and were not observed in the cold temperature evolution. As most of the observed changes at the nucleotide level appeared between step 9 and 16, it seems that

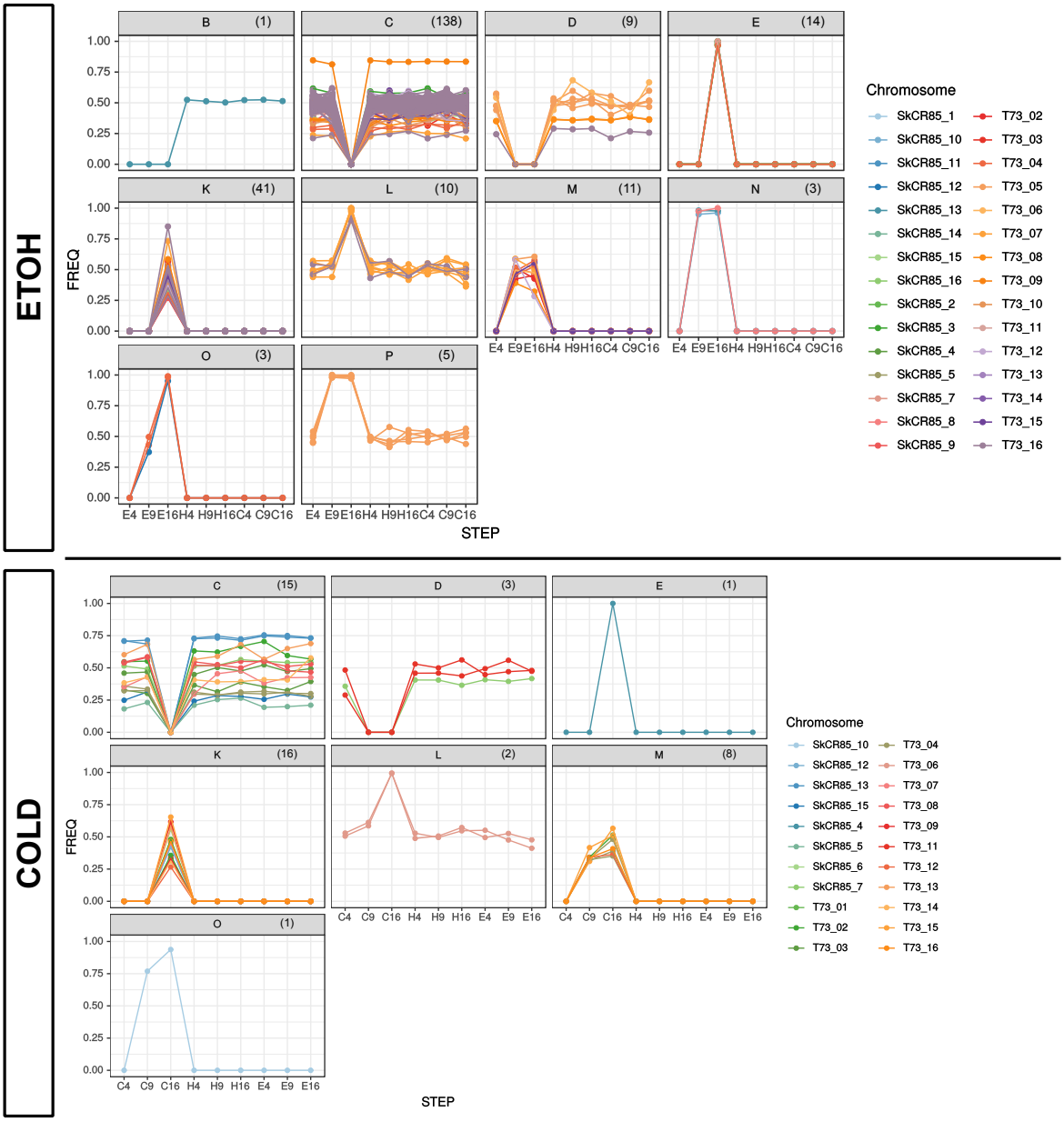

FIGURE 4.3 Clustering of the SNPs by comportment. Clusters of the selected SNPs in each condition. Each box represent a cluster that showed a logical comportment in the evolution experiment. The frequency is plotted of the SNP is plotted at each step (4,9,16) in the three evolution conditions (E: ethanol, C: cold temperature, H:25ºC). The first three points plotted in the x axis corresponds to the condition of interest, and the six following the frequency along evolution in the other two conditions. Numbers at the top represent the number of SNPs in the cluster. The colours correspond to the chromosome.

a certain waiting time exists in the access of this level of genomic variability.

4.3.2 Genome content is stable between subgenomes and cold evolution explored aneuploidy for adaptation.

Besides point mutations, interspecific hybrids have access to other sources of variability. They can change the chromosome dosage by aneuploidy and also change the number of *Sc* or *Sk* allele dosage by LOH but between homeologous chromosomes. As low temperature is thought to be more favorable to *Sk* and ethanol to *Sc*, we asked if changes in the genome content of one of the subgenome would appear in the evolution in these conditions.

To investigate the differences in genome content between the conditions we calculated a ratio similar to what was done in chapter 2. For each gene in the genome, the read-depth of the *Sc* allele was divided by the sum of the read depth of the *Sk* and *Sc* allele. As above we used step 16 of the evolution at 25ºC as control condition and subtracted it to step 16 in the conditions of interest. This deviation from the control was then plotted in a heatmap to observe the changes between conditions (Figure 4.4A). The deviation from the control, both in cold temperature and ethanol, is normally distributed and centered around 0 (Figure 4.4B). There are no skews in any directions which indicates that there is no directed loss as we observed in natural hybrids in the time and conditions of the experiment. The distribution is similar in both stress conditions refuting the hypothesis that cold temperature would fix more *Sk* allele and ethanol more *Sc*.

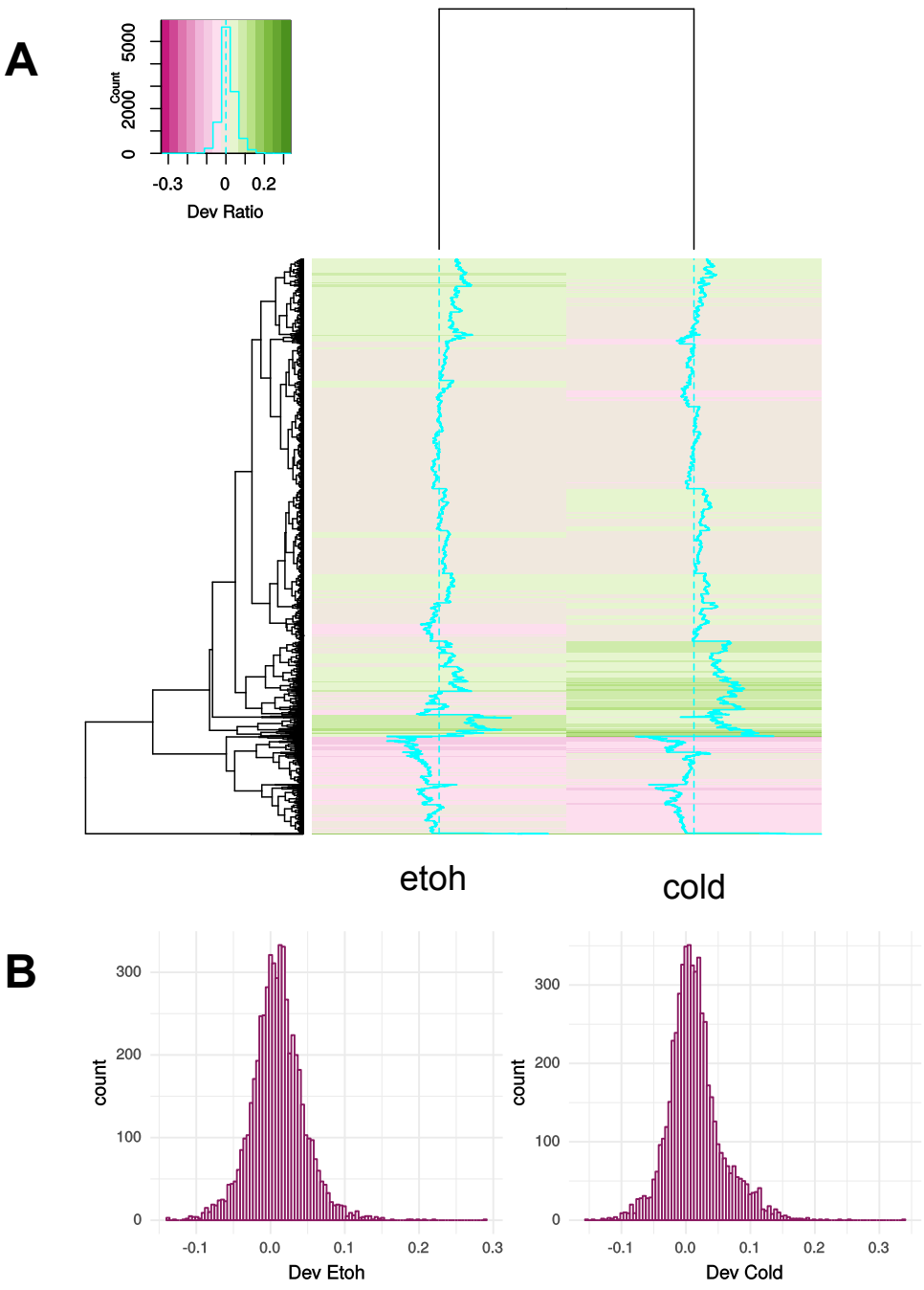

FIGURE 4.4 Genome ratio analysis. We here calculated the Sc/Sk ratio. The devRatio reprents the difference between the ratio at the end of the evolution in the condition of interest and the control condition. **A**. Heatmap representing the devRatio in ethanol and cold evolution. If the colour turns green then there is a change to *Sc* (ratio tends to 1) and if it turns pink, to *Sk* (ratio tends to 0). **B**. Histogram of the same data. Dev Etoh : ethanol condition, Dev Cold: cold temperature

The greatest part of the variation observed increased the *Sc* content (Figure 4.4A). We observed 65 genes with a deviation ratio higher than +0.08 to *Sc* in both ethanol and cold evolutions. These are scattered in the genome and the deviation is due to a lower ratio in the control condition rather than an increase in the condition of interest (mean ratio: 25ºC: 0.53; etoh: 0.65; 12ºC: 0.66). In the cold temperature evolution, more ratio deviations are observed. These are interestingly an increase in the *Sc* proportion. We found 354 genes with a deviation ratio higher than +0.08 compared to the control exclusively in the 12ºC evolution condition. The increase in ratio in these genes in this condition is due to an increment of the *Sc* content (mean ratio: 25ºC, 0.63; etoh, 0.66; 12ºC, 0.74). No GO enrichment was found for these genes. Nevertheless, 299 of these were from chromosome XII. The ratio of this group of genes and the mean deviation (ratio 0.75, deviation 0.11) are compatible with the increment in the number of copies of the overall *Sc* chromosome XII. To confirm the aneuploidy and determine when it occurred, we plot the coverage of the genome of the three steps of the cold temperature evolution (Figure 4.5). At the beginning of the evolution, the genome content is allotriploid with two *Sc* copies and one *Sk* copy, except for chromosome III which contains one copy of each subgenome, as the initial hybrid. At step 9, we observe an increase of read depth on the *Sc* chromosome XII. The rest of the genome does not show any other dosage change but the chromosome XII aneuploidy remains until the end of the evolution.

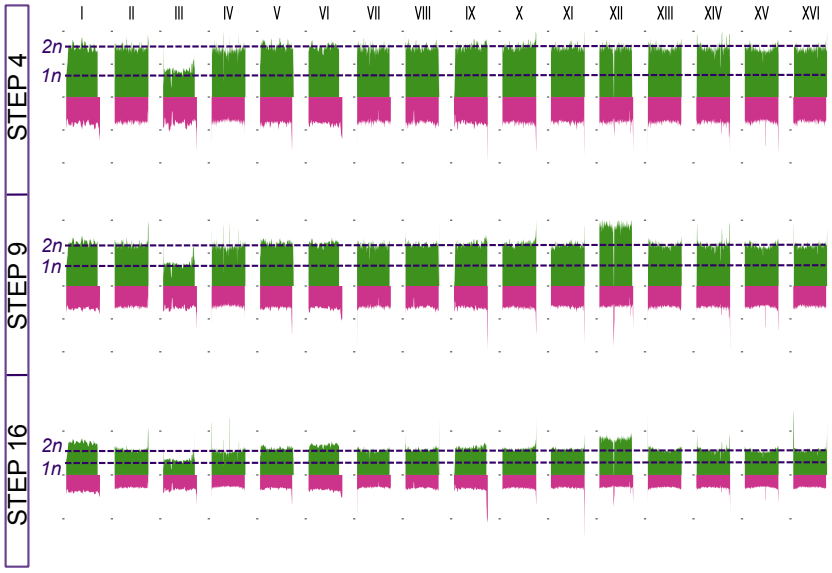

FIGURE 4.5 Genome content during the cold temperature evolution. The read depth is plotted for each step of the evolution at 12ºC. Subgenome are coloured in green, *Sc* and in pink *Sk*. The level of the ploidy is indicated by the horizontal line. The aneuploidy of the chromosome XII is observed at the step 9 of the evolution and maintained at step 16.

4.3.3 Global transcriptome is correlated with the genome content and the growth condition

We performed an RNA-seq experiment to investigate the comportment of the transcriptome in the evolution of hybrids and how it is related to the hybrid genome content. The artificial hybrids at step 4 and 16 of the evolution at 12ºC and 25ºC were grown in three conditions: 12ºC, 25ºC, and 25ºC with 10% ethanol.

One question we were interested in was if in the different conditions we could observe a deregulation of one of the subgenomes compared to its genome content. The calculation for the *Sc*/*Sk* transcription ratio of each gene, was done in a similar way as for the genome content above. We then subtracted to each transcription ratio its corresponding genome ratio. Interestingly we ob-

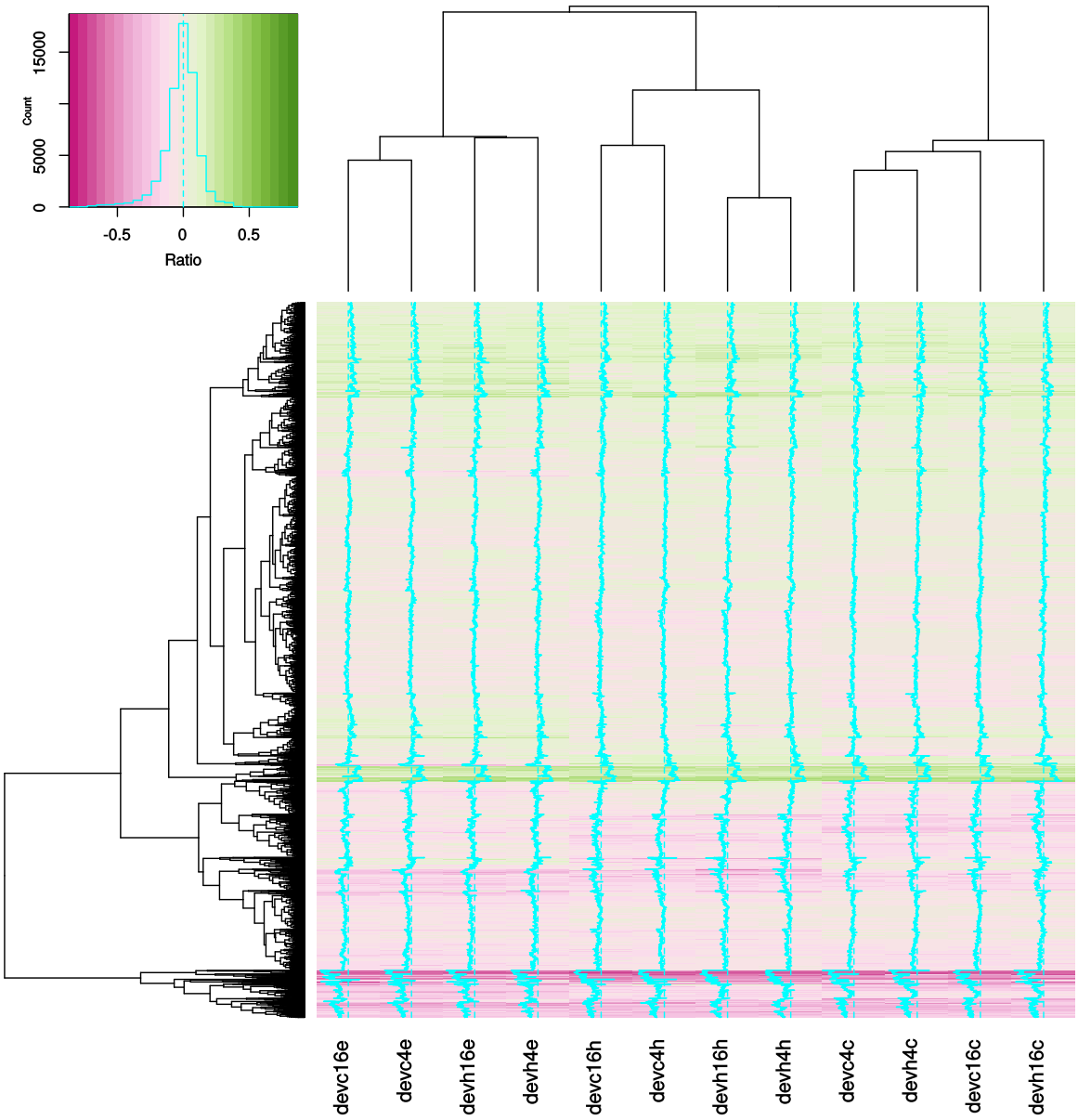

FIGURE 4.6 Deviation between the transcription and the genome ratio. the deviation ratio is here calculated as the difference between the ratio in the transcriptome and its corresponding genome ratio. The samples are named as follow: dev, then (c, or h) is the condition of evolution, then (4 or 16) is the step of evolution, then (e,h,c) are the condition of growth in the RNA-seq experiment (c: cold, e: ethanol, h: control at 25ºC). Then *devc126e* is the hybrid at step 16 of the evolution at cold temperature when growth in ethanol. If the colour turns green then there is a change to *Sc* (ratio tends to 1) and if it turns pink, to *Sk* (ratio tends to 0).

served that the transcription level is mostly related to the genome content (Figure 4.6). The deviation between the transcription ratio and the genome ratio is centered around 0 indicating that the ploidy is responsible for most of the transcript level difference between subgenome. Noteworthy, neither of the subgenomes have a notable different expression profile in the different growth conditions, hence in this study we do not observe a link between growth conditions and expressional changes of subgenome.

We further investigated the correlation between the ratio deviation between samples (Figure 4.7). We observed that the correlations are generally high as none of them are less than 0.6. The highest correlated samples corresponded to the ones grown in the same condition. This indicates that the evolution step and conditions had a smaller effect on the *Sc*/*Sk* ratio than the growth condition. Altogether we observe that the *Sc*/*Sk* transcription is well predicted by the genome ratio, as no global differences are observed in the different growth conditions, and that small changes that was observed group the different growth conditions together.

4.3.4 Evolution at cold temperature changes the transcriptome

We then investigated the transcription levels independently of the subgenome it comes from. All genes transcription levels were counted independently and a principal component analysis was performed (Figure 4.8). Overall, the samples clustered very nicely by growth condition. In the heat and ethanol conditions, no further clustering is observed. All the samples and replicates are mixed with no evidence of differentiation in these conditions. In the cold growth condition, however, we

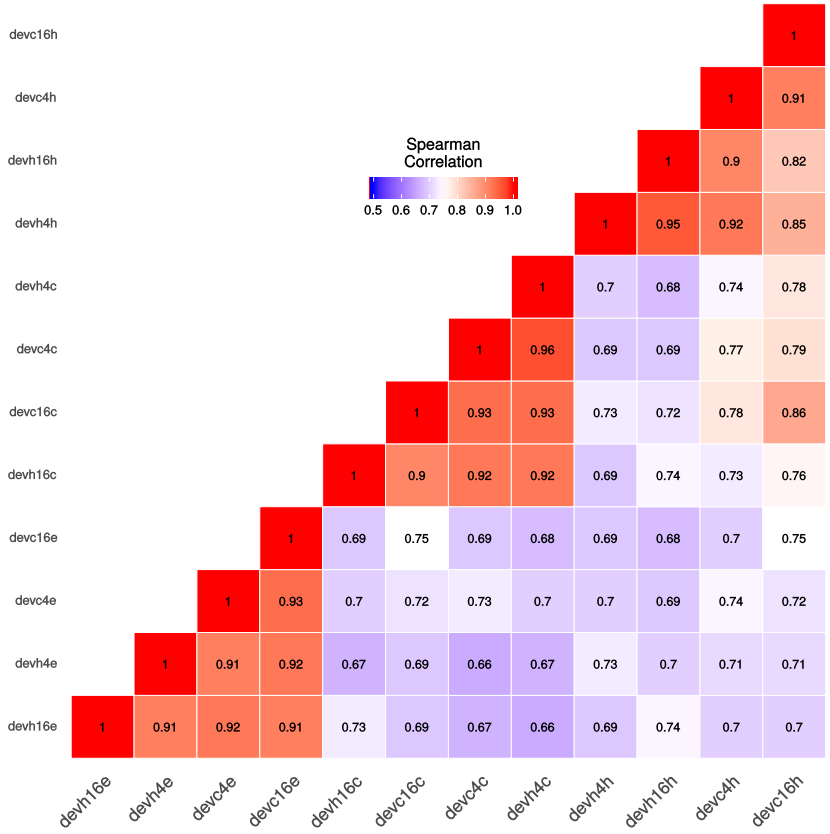

FIGURE 4.7 Correlation of the deviation RNA-DNA ratio between samples. The deviation ration between the transcriptome and the genome. Spearman correlation is calculated between each samples. The samples are named as follow: dev, then (c, or h) is the condition of evolution, then (4 or 16) is the step of evolution, then (e,h,c) are the condition of growth in the RNA-seq experiment (c: cold, e: ethanol, h: control at 25ºC). Then *devc126e* is the hybrid at step 16 of the evolution at cold temperature when growth in ethanol. The correlations are higher between the samples in the same growth condition.

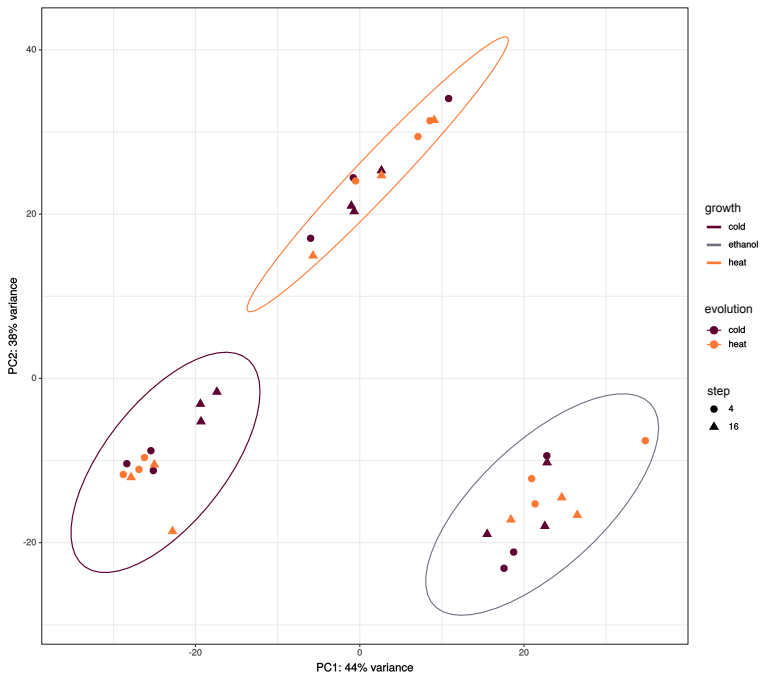

FIGURE 4.8 Principal Component Analysis of the complete transcriptome. The genes from the different subgenomes are here taken separately. The PCA shows the separation of the transcriptome is principally due to the growth condition (circle). The evolution step factor separates the three replicates of the cold temperature evolved hybrid from the rest of sample in the cold temperature growth condition.

observed two different clusters. The first cluster is formed by the samples of the hybrids evolved at 25ºC and both evolution steps together with the hybrids of the cold evolution at step 4. The second cluster in this growth condition is formed by the three replicates of the hybrids evolved at 12ºC at step 16. This indicates that the evolution at cold temperatures affects the transcription in this growth condition.

As the only evolution that gave rise to a transcriptional change was the evolution at cold temperature, we performed a differential gene expression analysis between the hybrid evolved at 12ºC at step 4 and the one at step 16 in the growth at the same temperature. In total 525 genes were significantly differentially expressed at an adjusted p-value lower than 0.05. From these, 323 were up- and 202 down-regulated. Among the upregulated genes, 205 are from the *Sc* chromosome XII

which is aneuploid at step 16. Surprisingly, among the 525 genes, only 47 are both *Sc* and *Sk* alleles differentially expressed. Among these, only one being discordant. It is the gene MAS1/YLR163C, a mitochondrial processing enzyme responsible for cleavage of targeting sequences from imported proteins, which is upregulated in *Sc* and downregulated in *Sk*. Within the genes in which only one of the alleles is differentially expressed, 314 are from *Sc* and 117 are from *Sk*. Considering that 201 of the *Sc* genes are upregulated due to the aneuploidy of chromosome XII, the contribution of each subgenome in the difference is similar.

A GO term enrichment analysis was performed with the DE genes. In the upregulated genes we found GO term enrichment for biological processes such as rRNA processing, ribosome biogenesis and other translation machinery related terms (Table 4.1). The downregulated genes shared biological process terms as electron transport chain, ATP biosynthetic process, aerobic respiration (Table 4.1). These results indicate that the cold temperature evolved hybrids upregulate the translation processes and downregulate the respiration and electronic transport chain compared to the original hybrid at cold temperature.

We investigated if the functional categories' enrichment had differences if DEG were from different subgenomes. We first analyzed the genes that appeared from both subgenomes. The upregulated were only 14, and no enrichment was found neither in molecular function nor in biological processes, but seven of them shared the cellular component nucleolus. Similarly, we found no enrichment for the downregulated genes, however, they were annotated to the categories of electron transport chain, ATP synthesis, as observed in the complete analysis. We then analyzed the genes

TABLE 4.1 Enriched GO terms for all differentially expressed genes after evolution at cold temperature

UP		Down	
Term	***p-value***	**Term**	***p-value***
ribosome biogenesis [GO:0042254]	*1E-014*	mitochondrial translation [GO:0032543]	*1E-014*
rRNA processing [GO:0006364]	*1E-014*	ATP synthesis coupled proton transport [GO:0015986]	*2.02061E-014*
endonucleolytic cleavage in ITS1 to separate SSU-rRNA from 5.8S rRNA and LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA) [GO:0000447]	*1.12433E-010*	electron transport chain [GO:0022900]	*9.35552E-012*
endonucleolytic cleavage to generate mature 5'-end of SSU-rRNA from (SSU-rRNA, 5.8S rRNA, LSU-rRNA) [GO:0000472]	*1.38874E-010*	proton transport [GO:0015992]	*2.21822E-010*
endonucleolytic cleavage in 5'-ETS of tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA) [GO:0000480]	*8.25357E-010*	aerobic respiration [GO:0009060]	*2.39085E-010*

that were differentially expressed only in one of the subgenomes (Table 4.2). Even if the genes were different, the upregulated genes in both *Sc*and *Sk* were related to ribosome biogenesis and rRNA processing. A similar pattern was observed for the downregulated genes. The enriched categories are the same for *Sc* and *Sk* subgenome even if the affected genes are different, and correspond to mitochondrial translation, mitochondrial proteins responsible for the aerobic respiration.

We used phenetics (De Maeyer et al., 2013) to identify transcription factors responsible for the differential expressed genes of each subgenome (Figure 4.9). We observed that the network resulting from the analysis is bigger in *Sc* than in *Sk*. Another observation is that *Sc* has more upregulated genes than *Sk*. Both of these facts are due to the *Sc* chromosome XII aneuploidy, as many of the genes linked are actually from chromosome XII in the *Sc* network. We focused on the central nodes (TABLE L), which are the transcription factors of interest, and detected three that were shared between *Sc* and *Sk*: YKL112W/ABF1, a general transcription regulator; YOR028C/CIN5, related to drug resistance and salt tolerance; and YOR358W/HAP5, a global regulator of respiratory genes. Interestingly, all three are downregulated in both *Sc* and *Sk*. An interesting point is that even these three transcription factors found in both networks regulate different genes. Among the central nodes in the *Sc* network, the second most connected gene (after CIN5 which is shared) was an interesting transcription factor: YLR223C/IFH1. This gene (IFH1) is located on chromosome XII and here appear as explaining an important part of the upregulated genes. Ifh1p regulates the transcription of nearly all ribosomal genes. It is interesting that it is located on chromosome XII and appears only in the *Sc* network. Summing up, we found that the regulation network that explains the differential expression is different in the two subgenome, but some factors are shared. These

TABLE 4.2 Enriched GO term in the genes differentially expressed only in one subgenome

UP				DOWN			
Sc		***Sk***		***Sc***		***Sk***	
Term	***p-value***	**Term**	***p-value***	**Term**	***p-value***	**Term**	***p-value***
ribosome biogenesis [GO:0042254]	*1.85283E-007*	ribosome biogenesis [GO:0042254]	*1.56652E-013*	mitochondrial translation [GO:0032543]	*1.9385E-009*	mitochondrial translation [GO:0032543]	*4.35283E-010*
rRNA processing [GO:0006364]	*1.3136E-007*	rRNA processing [GO:0006364]	*2.08339E-011*	aerobic respiration [02.13.03]	*1.55154E-006*	aerobic respiration [02.13.03]	*3.52695E-008*
		endonucleolytic cleavage to generate mature 5’-end of SSU-rRNA from (SSU-rRNA, 5.8S rRNA, LSU-rRNA) [GO:0000472]	*2.22539E-009*				
		endonucleolytic cleavage in ITS1 to separate SSU-rRNA from 5.8S rRNA and LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA) [GO:0000447]	*2.48135E-008*				

factors shared between both networks, nevertheless, regulates different genes. Moreover, we found a transcription factor in the aneuploid *Sc* chromosome XII that could explain the upregulation of the ribosomal protein and therefore part of the functional category's enrichment discovered here.

4.4 Discussion

Here we report the genomic and transcriptomic analysis of the experimental evolution of a *Sc* x *Sk* triploid artificial hybrid. We focused our questions on the way the genome was changing over the 400 generations at cold temperature and ethanol stress on two levels: nucleotide level and genome content. We also looked at the changes in the transcriptome focusing on two main questions. First, how the transcription is related with the ploidy and the impact of the stress condition. Second, how the evolution at stress changes the transcriptome to respond to the stress.

Variant detection is almost a classical analysis in the next generation sequencing era. Nevertheless, its reliability is not frequently discussed, and many studies only filter SNPs based on the quality score. This is particularly challenging when the organism studied is heterozygotic. Here we are reporting a two-step pipeline for the analysis of variants in a polyploid heterozygotic strain. In the first step we are using variants frequency at the end of the evolution experiment and the frequency comparison between evolution conditions, to detect which SNPs are condition-specific. The second step consists of clustering these SNPs based on their heterozygotic state. This permits us to further filter them based on reliability due to their comportment in the other conditions of

TABLE 4.3 Central nodes of the phenetics network for both subgenomes

Sc network			***Sk network***		
Systematic name	name	*name description*	*Systematic name*	name	*name description*
YKL112W	***ABF1***	**ARS-Binding Factor 1**	**YKL112W**	***ABF1***	**ARS-Binding Factor 1**
YOR028C	***CIN5***	**Chromosome INstability**	**YOR028C**	***CIN5***	**Chromosome INstability**
YOR358W	***HAP5***	**Heme Activator Protein**	**YOR358W**	***HAP5***	**Heme Activator Protein**
YDR146C	*SWI5*	SWItching deficient	YDR174W	*HMO1*	High MObility group (HMG) family
YDR216W	*ADR1*	Alcohol Dehydrogenase II synthesis Regulator	YGL254W	*FZF1*	Five Zinc Fingers
YIR017C	*MET28*	METhionine	YML007W	*YAP1*	Yeast AP-1
YKL020C	*SPT23*	SuPpressor of Ty	YOR344C	*TYE7*	Ty1-mediated Expression
YLL004W	*ORC3*	Origin Recognition Complex			
YLR182W	*SWI6*	SWItching deficient			
YLR183C	*TOS4*	Target Of Sbf			
YLR223C	*IFH1*	Interacts with Fork Head			
YLR451W	*LEU3*	LEUcine biosynthesis			
YMR042W	*ARG80*	ARGinine requiring			
YNL167C	*SKO1*	Suppressor of Kinase Overexpression			
YOR172W	*YRM1*	Yeast Reveromycin resistance Modulator			

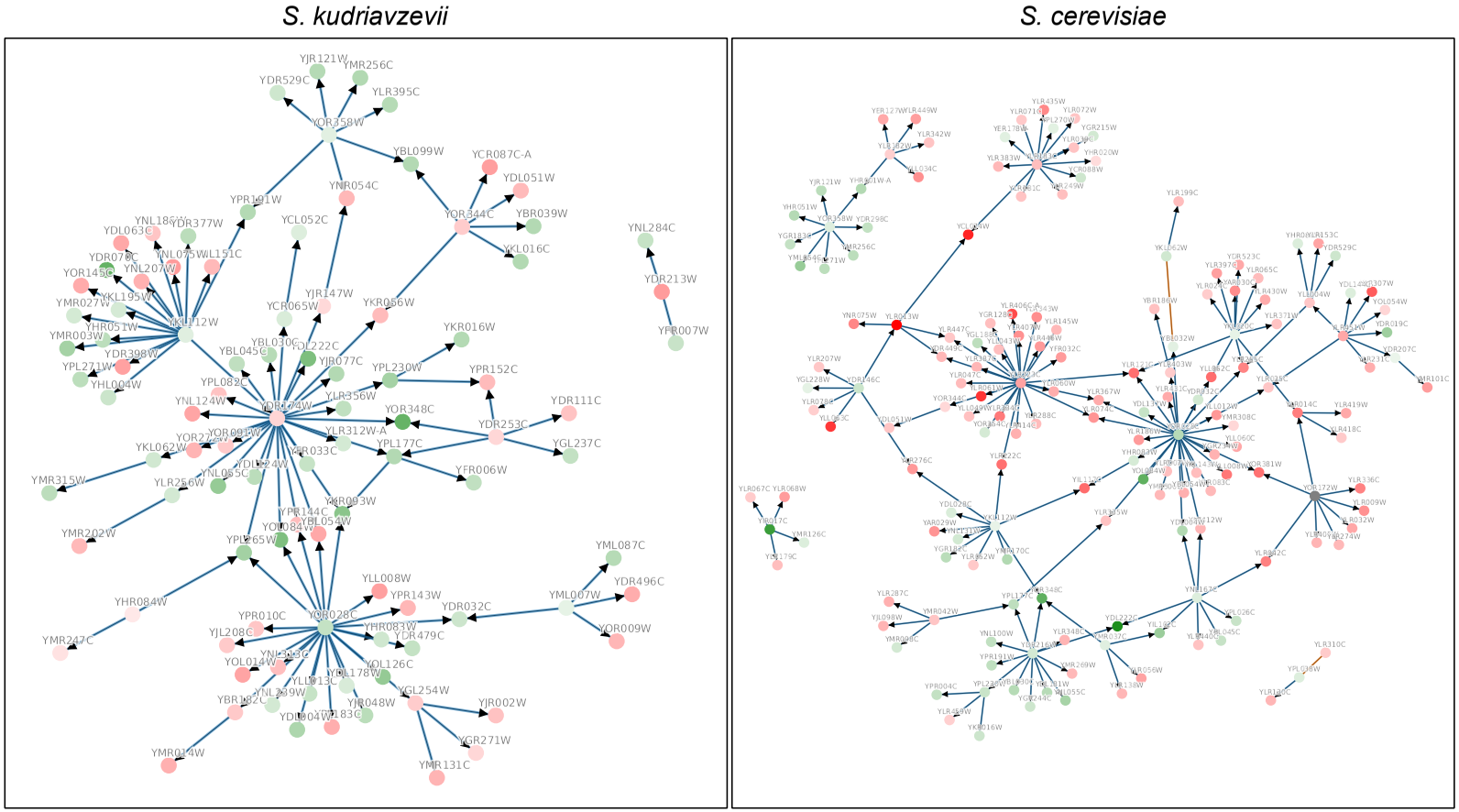

FIGURE 4.9 Phenetic analysis of the differentially expressed genes by subgenome. Phenetic was used to analyse the transcription factors explaining the differential expression of the cold evolved hybrid. The deferentially expressed genes in *Sc* and *Sk* were submitted separately.We observed that only three central nodes are shared between the subgenomes network.

the experiment. This analysis pipeline had the advantage to be conservative, what is important when dealing with low coverage sequencing and high genome complexity. The dynamic of the frequency can be therefore an interesting parameter to take into account in variant calling pipelines in heterozygous polyploids evolutionary experiments.

The SNPs analysis showed that the principal mechanism that caused variation in the genome sequence was loss of heterozygosity in the *Sc* subgenome. A recent study with diploid hybrids (Tattini et al., 2019) showed that LOH rate was lower in interspecific hybrids than in intraspecific hybrids. We here have a triploid hybrid with two copies of the *Sc* and observe high rate of LOH in the *Sc* subgenome but no LOH between the *Sc* and the *Sk*. This could be evidence of a lower LOH rate between more distant subgenome, what supports the results by Tattini et al 2019. As directional homeologous recombinations are observed in industrial hybrids (chapter 2) we sought to observe them in this experiment. Likewise, other studies with experimental evolution in hybrids saw directional LOH between the subgenomes in different stresses (Zhang et al., 2019), and the loss of *Sc* to the cryotolerant species in cold temperature evolution (Smukowski Heil et al., 2019). However, this correlation between environment and LOH was not observed here. As the other experiments used diploid hybrids, we can speculate that the access to certain variation differ between diploids and triploids, in the same way as diploids, haploids and polyloids differ in the adaptation mechanisms in *Sc* (**?**) Marad et al. (2018). Another difference between these studies and the present work is that our experimental procedure includes strong bottlenecks at each step what gives an important drift effect. Further work will need to be done to decipher what is the importance of these factors to observe directional changes in the genome.

Here we observed a strong correlation between the ploidy and the expression between subgenomes. We showed that the ratio of *Sc* and *Sk* could explain the transcription ratio. The ratio of transcription is also correlated with the aneuploidy which increases the expression of *Sc* chromosome XII. In a study in the hybrid Zygosaccharomyces parabailii (Ortiz-Merino et al., 2018) authors used the same ratio and claimed that lactic acid stress unbalances it. Here we wanted to determine whether the ratio was affected by the condition and the evolution in this condition. Indeed, we showed that the condition explained the differences in ratio. What we did not observe was a significant increase in expression of *Sk* in cold and of *Sc* in ethanol, which could have been expected if the specialized subgenome was overexpressed in their preferred condition. In lager yeast hybrids it was shown that ploidy influenced the transcription levels and the attributes of the final products (Krogerus et al., 2017). The changes observed here in the ratios confirms that the broader changes are due to the ploidy and the number of copies of each chromosomes.

The hybrids used in this study were phenotyped in each condition in another thesis of our group (Ortiz Tovar, 2018). In that work it was shown that the only condition in which the evolution improved the growth ability of the hybrid was cold temperature. Here we observed that the biggest change at the genome level during the evolution occurred at cold temperature and was an increase in the number of copies of *Sc* chromosome XII. We also showed that this was the only condition of evolution that affected the transcriptome. These were, summarizing, an upregulation of the translation related processes and a downregulation of the respiration. These results agree with what was previously observed. In *Sk* it was shown that the increase of translation efficiency improved the cold temperature tolerance (Tronchoni et al., 2014). This was also observed in *Sc* were it seems that

this upregulation of translation related protein, ribosomal proteins, aims to cope with the block of it at coldest temperature (García-Ríos et al., 2016). It was shown that to cope with cold temperature it was necessary to increase the resistance to oxidative stress using the sulfur assimilation pathway (García-Ríos et al., 2014). The reduction of the respiratory function also reduces the apparition of ROS. We therefore hypothesis that what we have observed here is a different path to the same objective: dealing with the oxidative stress that appears in this condition.

We were surprised that the differential expression analysis revealed the same processes were affected even though the genes involved were different. We did a network analysis with phenetics which revealed that even if some of the transcription factors were the same in both *Sc* and *Sk*, the genes regulated were different. This points to that *Sc* and *Sk* have different regulation networks and that further work has to be done to understand it. Nevertheless, it seems that this could be an interesting way to understand heterosis if in the hybrid both networks are used it could improve the adaptability of the hybrids. The analysis revealed that the transcription factor IFH1 from the chromosome XII was one of the central node in *Sc* network. Ifh1p regulates transcription of ribosomal proteins, and so is related with the translation process we saw was important at cold temperature. Also the ribosomal tracks are located on chromosomes XII. It was shown in other conditions that aneuploidy was a fast mechanism to overexpress certain genes related to stress (Yona et al., 2012). Here the chromosome XII aneuploidy could increase both ribosomal protein and the expression of the regulator Ifh1p to improve the translation efficiency at cold tolerance.

In conclusion, we observed that alotriploid *Sc* x *Sk* had a higher LOH rate between *Sc* copies

than between homeologous *Sc* and *Sk* subgenomes along the 400 generations of the experiment. The ploidy explained the expression of the chromosomes. The evolution at cold temperature gave rise to aneuploidy and transcriptomic adaptation. Even though more needs to be done to understand ploidy influence and transcription regulation in the context of hybridization, our work supports the idea that these are important factors that can determine the accessibility to adaptive mechanisms and the outcome of hybrid evolution.

General Discussion

What makes an organism fitter to an environment is an intriguing question in biology. At the genomic level, many mechanisms are available that will impact the phenotype. With the explosion of genome sequencing technologies, mechanisms such as aneuploidy, polyploidy, and hybridization are emerging as being more frequent and relevant than what was considered earlier. To study such mechanisms a good model organism is *Saccharomyces* yeast. *Saccharomyces* have small genomes and can be easily cultivated. *Saccharomyces cerevisiae* is probably among the best-understood organisms and many different data are available to question different aspects of genome evolution. In this doctoral thesis, we aimed to investigate different aspects of the adaptive value of aneuploidy and interspecific hybridization in *Saccharomyces*.

In the first chapter, we were interested in studying what genomic differences were underlying the different ethanol tolerance observed in *S. cerevisiae* strains (Arroyo-López et al., 2010). The most interesting genomic change we observed was a shared polysomy of chromosome III in the

highest ethanol tolerant strains. We could investigate that the correlation between the number of copies of chromosome III with ethanol tolerance was also found in multiple strains backgrounds. Aneuploidy is increasingly being scrutinized as a fast and important mechanism to the resistance of different stresses (Gorter de Vries et al., 2017). What we bring here is that we could demonstrate that this specific chromosome polysomy was an adaptive mechanism by removing the extra copy and observing the phenotype change. It is, however, worth to say that more is needed to understand what gene (genes) or what mechanism is responsible for the increase in ethanol tolerance.

In industrial environments, different *S. cerevisiae* x *S. kudriavzevii* hybrids were isolated and studied in our group (González et al., 2008; Peris, Lopes, Belloch, Querol and Barrio, 2012*b*; Belloch et al., 2008). In the second chapter of this work, we asked how these hybrids mate and how this mechanism would influence the genomic and adaptability outcome of these hybrids. We first could confirm that the isolates were from different origins and therefore different hybridization events. The genome structure also revealed that the most frequent mechanism used to mate was rare-mating. The heterozygosity in the *S. cerevisiae* subgenome in most of the strains and the haploid contribution of *S. kudriavzevii* could only be explained by this mechanism and could be experimentally confirmed by artificial hybridization. Howerver, we also showed that other mechanisms could be used such as spore to spore mating. The case we showed here had a tetraploid genome and the genome structure was explained by genome doubling after the hybridization. The fertility of this hybrid was also high, which makes this hybrid an example of fertility restoration by genome duplication as observed in other hybrids in the genus (Charron et al., 2019). We observed that most of the hybrids had different, uniques, homeologous recombinations in their genome, changing cold

temperature phenotype (Ortiz-Tovar et al., 2018) , and that the number of these events reduces the fertility.

One of the hybrids we studied in chapter 2, VIN7, was previously sequenced by another group (Borneman et al., 2012) . However, the genome content that they observed was different from what we observed using different methodologies (Peris, Lopes, Arias and Barrio, 2012). In chaper 3, we studied in detail the genome of this strain and showed that its genome was unstable. We could see that depending on the treatment and isolation source, certain *S. kudriavzevii* chromosomes were lost, and that this impacted the phenotype of the strain when growing under ethanol stress. We studied the influence of different factors of the different chromosomes such as the length, genetic interactions or physical interactions content in their loss or gain probability. We found that the length and the interactions between them were the characteristics that better explained their aneuploidy probability. What this chapter suggests is that genomic instability, here probably triggered by hybridization, is an important factor of phenotypic variability, and therefore to adaptability.

The last chapter of this work deals with short-term evolution of artificial hybrids between *S. cerevisiae* and *S. kudriavzevii*. We wanted to know how the genome content changed in conditions in which the species that form the hybrid had different phenotype : ethanol, were *S. cerevisiae* is better fit, and cold temperature, where the best species is *S. kudriavzevii* (Arroyo-López et al., 2010; Salvadó et al., 2011). We surprisingly did not observe directed loss of genomic regions of the unfit subgenome in the length of our experiment. We instead saw that recombination between the *S. cerevisiae* copies. Therefore homeologous recombination is probably not the first accessible

mechanism for adaptation. More will have to be done to answer if it is because this mechanism is more detrimental than others and then is observed on more long-term adaptation, or if it is an effect of drift in our experiment. We also stand that the ploidy has an important effect on the adaptive mechanisms available. We also wanted to know how the evolution and growth conditions affected the expression of the genes. The RNA-seq data also pointed out the importance of the ploidy. The ratio of *S. cerevisiae* / *S. kudriavzevii* expression was mostly explained by the chromosomal genomic ratio. The evolution at cold temperature was the only condition that changed the expression profile. We once again saw that aneuploidy was an important factor in the adaptation, here in hybrids. Chromosome XII aneuploidy in the hybrid evolved at cold temperature could be related to a higher translation efficiency that helped to cope with this stress. We also observed that the genes differentially expressed after the evolution in both subgenomes were different but the processes affected were similar. This points to an interesting, direction, and where future work could deepen: adaptability is more related to systems as a whole than to specific genes.

To recapitulate, in this work we presented different shreds of evidence of the importance of ploidy and its variability in the adaptability of *Saccharomyces*. We emphasized the relevance of aneuploidy as a mechanism to change gene expression and increase stress resistance. We also evidenced that the mating mechanism between species influences the ploidy of the offspring. This ploidy is then influencing the accessibility to certain mechanisms of generating genome variability and therefore the way hybrids can explore the phenotypic landscape.

Conclusions

1) Different subpopulations of wine *S. cerevisiae* strains are observed. The flor strains showed higher ethanol tolerance than other strains but intrapopulation differences were also present. This points out that the ethanol tolerance phenotype is variable even within populations.
2) The heterozygosity in *S. cerevisiae* showed a correlation with the population origins. It is related to different meiosis frequency, outcrossing and self-fertilization usage.
3) An aneuploidy on chromosome III was shared by high ethanol tolerant strains. By removing the extra copy of this chromosome ethanol tolerance is reduced. The number of chromosome copies also correlated with ethanol tolerance in multiple backgrounds. Therefore chromosome III aneuploidy is an adaptive mechanism to cope with ethanol stress.
4) The phylogenomic analysis of *S. cerevisiae* x *S. kudriavzevii* hybrids confirmed that they were originated by different hybridization events. As at least three different events occurred, indicating that hybridization is a recurrent phenomenon in *Saccharomyces* adaptation to industrial fermentations.

5) The ploidy of hybrids is a direct result of the mating mechanism involved in hybridization. Consequently, different mating mechanisms will result in different genome compositions, which constrain further genome evolution.

6) Allotriploidy is the most common genome composition in *S. cerevisiae* x *S. kudriavzevii* hybrids. The levels of heterozygosity observed in the hybrid S. cerevisiae subgenome indicates that rare-mating is the most frequent mechanism of hybridization. Artificial hybridization by rare-mating resulted in a similar genome structure, which confirms the main role of this conjugation type in the generation of hybrids.

7) One of the hybrids, PB7, resulted to be a perfect homozygous allotetraploid hybrid exhibiting high fertility. Therefore, spore-to-spore conjugation and allotetraploidization by genome duplication is unfrequent but is also occurring in *Saccharomyces* hybridization.

8) Homeologous recombination was found to be an important contributor to the genomic variability in industrial *S. cerevisiae* x *S. kudriavzevii* hybrids. This could have improved the phenotypic variability between strains to respond to selective pressures but also improves the post-zygotic barrier, as hybrids with higher recombinations show lower fertility.

9) Genome instability in hybrids can lead to different genomic outcomes in a few generations. The loss of different chromosomes in the VIN7 hybrid affected the phenotype of the strain.This indicates that genome instability generates variation allowing a faster adaptation, that may be promoted under the selective pressures present in industrial environments.

10) Aneuploidy patterns are governed by the number of interactions among chromosomes and the size of the chromosomes involved. Besides these factors, many others can influence the

frequency of the aneuploidy of each chromosome individually, but no other general patterns are found.

11) Experimental evolution of artificial hybrids showed that Loss of Heterozygosity and aneuploidy were the main genomic changes observed at short evolution times.

12) In the allotriploid hybrids, the LOH rate between the allelic copies of the hybrid *S. cerevisiae* chromosomes is higher than between the homeologous *S. cerevisiae* and *S. kudriavzevii* chromosomes. The different stress conditions do not affect the replacement of one subgenome by the other even if the parental species is fitter in each condition.

13) Expression at the chromosomal level is strongly affected by the ploidy.

14) Cold stress is the only condition affecting the transcription pattern. Cold stress promoted upregulation of translation, favored by the polysomy of chromosome XII, where the ribosomal genes are located, and downregulation of the respiration process.

15) The processes affected in the transcriptional rewiring observed at the end of the evolution are the same in both subgenomes even though the specific genes are different. Even if regulators are similar, the genes affected can change in both subgenomes. This means that the adaptation occurs at the system level instead than at the gene-specific level.

Bibliography

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., Jones, J., Keller, B., Marczewski, T., Mallet, J., Martinez-Rodriguez, P., Möst, M., Mullen, S., Nichols, R., Nolte, A. W., Parisod, C., Pfennig, K., Rice, A. M., Ritchie, M. G., Seifert, B., Smadja, C. M., Stelkens, R., Szymura, J. M., Väinölä, R., Wolf, J. B. W. and Zinner, D. (2013), 'Hybridization and speciation', *Journal of Evolutionary Biology* **26**(2), 229–246.

Ackermann, R. R., Arnold, M. L., Baiz, M. D., Cahill, J. A., Cortés Ortiz, L., Evans, B. J., Grant, B. R., Grant, P. R., Hallgrimsson, B., Humphreys, R. A., Jolly, C. J., Malukiewicz, J., Percival, C. J., Ritzman, T. B., Roos, C., Roseman, C. C., Schroeder, L., Smith, F. H., Warren, K. A., Wayne, R. K. and Zinner, D. (2019), 'Hybridization in human evolution: Insights from other organisms', *Evolutionary Anthropology: Issues, News, and Reviews* **28**(4), 189–209.

Adamczyk, J., Deregowska, A., Skoneczny, M., Skoneczna, A., Kwiatkowska, A., Potocki, L., Rawska, E., Pabian, S., Kaplan, J., Lewinska, A. and Wnuk, M. (2016), 'Adaptive response to chronic mild ethanol stress involves ros, sirtuins and changes in chromosome dosage in wine yeasts', *Oncotarget* **7**(21), 29958–29976.

Albertin, W., Chernova, M., Durrens, P., Guichoux, E., Sherman, D. J., Masneuf-Pomarede, I. and Marullo, P. (2018), 'Many interspecific chromosomal introgressions are highly prevalent in Holarctic Saccharomyces uvarum strains found in human-related fermentations', *Yeast* **35**, 141–156.

Alix, K., Gérard, P. R., Schwarzacher, T. and Heslop-Harrison, J. S. P. (2017), 'Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants', *Annals of Botany* **120**(2), 183–194.

Almeida, P., Barbosa, R., Bensasson, D., Gonçalves, P. and Sampaio, J. P. (2017), 'Adaptive divergence in wine yeasts and their wild relatives suggests a prominent role for introgressions and rapid evolution at noncoding sites', *Molecular Ecology* **26**(7), 2167–2182.

Almeida, P., Gonçalves, C., Teixeira, S., Libkind, D., Bontrager, M., Masneuf-Pomarède, I., Albertin, W., Durrens, P., Sherman, D. J., Marullo, P., Todd Hittinger, C., Gonçalves, P. and Sampaio, J. P. (2014), 'A Gondwanan imprint on global diversity and domestication of wine and cider yeast Saccharomyces uvarum', *Nature Communications* **5**(1), 4044.

Anders, S., Pyl, P. T. and Huber, W. (2015), 'HTSeq–a Python framework to work with high-throughput sequencing data', *Bioinformatics* **31**(2), 166–169.

Anderson, J. B., Funt, J., Thompson, D. A., Prabhu, S., Socha, A., Sirjusingh, C., Dettman, J. R., Parreiras, L., Guttman, D. S., Regev, A. and Kohn, L. M. (2010), 'Determinants of Divergent Adaptation and Dobzhansky-Muller Interaction in Experimental Yeast Populations', *Current Biology* **20**(15), 1383–1388.

Ansanay-Galeote, V., Blondin, B., Dequin, S. and Sablayrolles, J.-M. (2001), 'Stress effect of ethanol on fermentation kinetics by stationary-phase cells of Saccharomyces cerevisiae', *Biotechnology Letters* **23**(9), 677–681.

Arnold, M. L. (1997), *Natural Hybridization and Evolution*, Oxford Series in Ecology and Evolution, Oxford University Press.

Arroyo-López, F. N., Salvadó, Z., Tronchoni, J., Guillamón, J. M., Barrio, E. and Querol, A. (2010), 'Susceptibility and resistance to ethanol in saccharomyces strains isolated from wild and fermentative environments', *Yeast* **27**(12), 1005–1015.

Auesukaree, C. (2017), 'Molecular mechanisms of the yeast adaptive response and tolerance to stresses encountered during ethanol fermentation', *Journal of Bioscience and Bioengineering* **124**(2), 133–142.

Baldauf, S. L. and Palmer, J. D. (1993), 'Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins.', *Proceedings of the National Academy of Sciences of the United States of America* **90**(24), 11558–62.

Barbosa, R., Almeida, P., Safar, S. V., Santos, R. O., Morais, P. B., Nielly-Thibault, L., Leducq, J. B., Landry, C. R., Gonçalves, P., Rosa, C. A. and Sampaio, J. P. (2016), 'Evidence of natural hybridization in Brazilian wild lineages of Saccharomyces cerevisiae', *Genome Biology and Evolution* **8**(2), 317–329.

Barton, N. H. (2001), 'The role of hybridization in evolution', *Molecular Ecology* **10**(3), 551–568.

Bell, M. A. and Travis, M. P. (2005), 'Hybridization, transgressive segregation, genetic covariation, and adaptive radiation', *Trends in Ecology & Evolution* **20**(7), 358–361.

Belloch, C., Orlic, S., Barrio, E. and Querol, A. (2008), 'Fermentative stress adaptation of hybrids within the Saccharomyces sensu stricto complex', *International Journal of Food Microbiology* **122**(1-2), 188–195.

Bennett, R. J., Forche, A. and Berman, J. (2014), 'Rapid mechanisms for generating genome diversity: Whole ploidy shifts, aneuploidy, and loss of heterozygosity', *Cold Spring Harbor Perspectives in Medicine* **4**(10).

Berman, J. (2016), 'Ploidy plasticity: a rapid and reversible strategy for adaptation to stress', *FEMS Yeast Research* **16**(3), fow020.

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. and Pirovano, W. (2011), 'Scaffolding preassembled contigs using SSPACE.', *Bioinformatics (Oxford, England)* **27**(4), 578–9.

Boetzer, M. and Pirovano, W. (2012), 'Toward almost closed genomes with GapFiller.', *Genome biology* **13**(6), R56.

Borneman, A. R., Desany, B. A., Riches, D., Affourtit, J. P., Forgan, A. H., Pretorius, I. S., Egholm, M. and Chambers, P. J. (2011), 'Whole-Genome Comparison Reveals Novel Genetic Elements That Characterize the Genome of Industrial Strains of Saccharomyces cerevisiae', *PLoS Genetics* **7**(2), e1001287.

Borneman, A. R., Desany, B. a., Riches, D., Affourtit, J. P., Forgan, A. H., Pretorius, I. S., Egholm, M. and Chambers, P. J. (2012), 'The genome sequence of the wine yeast VIN7 reveals an allotriploid hybrid genome with Saccharomyces cerevisiae and Saccharomyces kudriavzevii origins', *FEMS Yeast Research* **12**(1), 88–96.

Borneman, A. R., Forgan, A. H., Kolouchova, R., Fraser, J. A. and Schmidt, S. A. (2016), 'Whole genome comparison reveals high levels of inbreeding and strain redundancy across the spectrum of commercial wine strains of saccharomyces cerevisiae', *G3: Genes, Genomes, Genetics* **6**(4), 957–971.

Boynton, P. J. and Greig, D. (2014), 'The ecology and evolution of non-domesticated saccharomyces species', *Yeast* **31**(12), 449–462.

Brauner, A., Fridman, O., Gefen, O. and Balaban, N. Q. (2016), 'Distinguishing between resistance, tolerance and persistence to antibiotic treatment', *Nature Reviews Microbiology* **14**(5), 320–330.

Chambers, S. R., Hunter, N., Louis, E. J. and Borts, R. H. (1996), 'The mismatch repair system reduces meiotic homeologous recombination and stimulates recombination-dependent chromosome loss.', *Molecular and Cellular Biology* **16**(11), 6110–6120.

Charrad, M., Ghazzali, N., Boiteau, V. and Niknafs, A. (2014), 'Nbclust: An r package for determining the relevant number of clusters in a data set', *Journal of Statistical Software, Articles* **61**(6), 1–36.

Charron, G., Marsit, S., Hénault, M., Martin, H. and Landry, C. R. (2019), 'Spontaneous whole-genome duplication restores fertility in interspecific hybrids', *Nature Communications* **10**(1), 4126.

Chen, G., Rubinstein, B. and Li, R. (2012), 'Whole chromosome aneuploidy: Big mutations drive adaptation by phenotypic leap', *BioEssays* **34**(10), 893–900.

Chou, J.-Y. and Leu, J.-Y. (2010), 'Speciation through cytonuclear incompatibility: Insights from yeast and implications for higher eukaryotes', *BioEssays* **32**(5), 401–411.

Coi, A. L., Bigey, F., Mallet, S., Marsit, S., Zara, G., Gladieux, P., Galeote, V., Budroni, M., Dequin, S. and Legras, J. L. (2017), 'Genomic signatures of adaptation to wine biological ageing conditions in biofilm-forming flor yeasts', *Molecular Ecology* **26**(7), 2150–2166.

Costanzo, M., VanderSluis, B., Koch, E. N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S. D., Pelechano, V., Styles, E. B., Billmann, M., van Leeuwen, J., van Dyk, N., Lin, Z.-Y., Kuzmin, E., Nelson, J., Piotrowski, J. S., Srikumar, T., Bahr, S., Chen, Y., Deshpande, R., Kurat, C. F., Li, S. C., Li, Z., Usaj, M. M., Okada, H., Pascoe, N., San Luis, B.-J., Sharifpoor, S., Shuteriqi, E., Simpkins, S. W., Snider, J., Suresh, H. G., Tan, Y., Zhu, H., Malod-Dognin, N., Janjic, V., Przulj, N., Troyanskaya, O. G., Stagljar, I., Xia, T., Ohya, Y., Gingras, A.-C., Raught, B., Boutros, M., Steinmetz, L. M., Moore, C. L., Rosebrock, A. P., Caudy, A. A., Myers, C. L., Andrews, B. and Boone, C. (2016), 'A global genetic interaction network maps a wiring diagram of cellular function', *Science* **353**(6306).

Coste, A., Turner, V., Ischer, F., Morschhauser, J., Forche, A., Selmecki, A., Berman, J., Bille, J. and Sanglard, D. (2006), 'A mutation in Tac1p, a transcription factor regulating CDR1 and CDR2, is coupled with loss of heterozygosity at chromosome 5 to mediate antifungal resistance in Candida albicans', *Genetics* **172**(4), 2139–2156.

Crebelli, R., Conti, G., Conti, L. and Carere, A. (1989), 'A Comparative-Study on Ethanol and Acetaldehyde as Inducers of Chromosome Malsegregation in Aspergillus-Nidulans', *Mutation Research* **215**(2), 187–195.

Darling, A. E., Mau, B. and Perna, N. T. (2010), 'progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement.', *PloS one* **5**(6), e11147.

Dayarian, A., Michael, T. P. and Sengupta, A. M. (2010), 'SOPRA: Scaffolding algorithm for paired reads via statistical optimization.', *BMC bioinformatics* **11**(1), 345.

De Maeyer, D., Renkens, J., Cloots, L., De Raedt, L. and Marchal, K. (2013), 'PheNetic: network-based interpretation of unstructured gene lists in E. coli', *Molecular BioSystems* **9**(7), 1594.

De Queiroz, K. (2007), 'Species concepts and species delimitation', *Systematic Biology* **56**(6), 879–886.

Deatherage, D. E. and Barrick, J. E. (2014), *Identification of Mutations in Laboratory-Evolved Microbes from Next-Generation Sequencing Data Using breseq*, Springer New York, New York, NY, pp. 165–188.

Demuyter, C., Lollier, M., Legras, J.-L. and Le Jeune, C. (2004), 'Predominance of Saccharomyces uvarum during spontaneous alcoholic fermentation, for three consecutive years, in an Alsatian winery', *Journal of Applied Microbiology* **97**(6), 1140–1148.

Depotter, J. R., Seidl, M. F. and Wood, T. A. (2016), 'Interspecific hybridization impacts host range and pathogenicity of filamentous microbes', *Current Opinion in Microbiology* **32**, 7–13.

Dion-Côté, A.-M. and Barbash, D. A. (2017), 'Beyond speciation genes: an overview of genome stability in evolution and speciation', *Current Opinion in Genetics & Development* **47**, 17–23.

Dowling, T. E. and Secor, C. L. (1997), 'The role of hybridization and introgression in the diversification of animals', *Annual Review of Ecology and Systematics* **28**(1), 593–619.

Duan, S.-F., Han, P.-J., Wang, Q.-M., Liu, W.-Q., Shi, J.-Y., Li, K., Zhang, X.-L. and Bai, F.-Y. (2018), 'The origin and adaptive evolution of domesticated populations of yeast from Far East Asia', *Nature Communications* **9**(1), 2690.

Dujon, B. (2006), 'Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution', *Trends in Genetics* **22**(7), 375–387.

Dujon, B. (2015), 'Basic principles of yeast genomics, a personal recollection', *FEMS Yeast Research* **15**(5), 1–11.

Dujon, B. A. and Louis, E. J. (2017), 'Genome diversity and evolution in the budding yeasts (saccharomycotina)', *Genetics* **206**(2), 717–750.

Dunn, B., Paulish, T., Stanbery, A., Piotrowski, J., Koniges, G., Kroll, E., Louis, E. J., Liti, G., Sherlock, G. and Rosenzweig, F. (2013), 'Recurrent rearrangement during adaptive evolution in an interspecific yeast hybrid suggests a model for rapid introgression', *PLOS Genetics* **9**(3), 1–16.

Eberlein, C., Hénault, M., Fijarczyk, A., Charron, G., Bouvier, M., Kohn, L. M., Anderson, J. B. and Landry, C. R. (2019), 'Hybridization is a recurrent evolutionary stimulus in wild yeast speciation', *Nature Communications* **10**(1), 923.

Edger, P. P., Poorten, T. J., VanBuren, R., Hardigan, M. A., Colle, M., McKain, M. R., Smith, R. D., Teresi, S. J., Nelson, A. D. L., Wai, C. M., Alger, E. I., Bird, K. A., Yocca, A. E., Pumplin, N., Ou, S., Ben-Zvi, G., Brodt, A., Baruch, K., Swale, T., Shiue, L., Acharya, C. B., Cole, G. S., Mower, J. P., Childs, K. L., Jiang, N., Lyons, E., Freeling, M., Puzey, J. R. and Knapp, S. J. (2019), 'Origin and evolution of the octoploid strawberry genome', *Nature Genetics* **51**(3), 541–547.

Eldarov, M. A., Beletsky, A. V., Tanashchuk, T. N., Kishkovskaya, S. A., Ravin, N. V. and Mardanov, A. V. (2018), 'Whole-genome analysis of three yeast strains used for production of sherry-Like wines revealed genetic traits specific to flor yeasts', *Frontiers in Microbiology* **9**, 965.

Erny, C., Raoult, P., Alais, A., Butterlin, G., Delobel, P., Matei-Radoi, F., Casaregola, S. and Legras, J. L. (2012), 'Ecological success of a group of Saccharomyces cerevisiae/Saccharomyces kudriavzevii hybrids in the Northern European wine-making environment', *Applied and Environmental Microbiology* **78**(9), 3256–3265.

Esteve-Zarzoso, B., Peris-Torán, M. J., García-Maiquez, E., Uruburu, F. and Querol, A. (2001), 'Yeast population dynamics during the fermentation and biological aging of sherry wines', *Applied and Environmental Microbiology* **67**(5), 2056–2061.

Feurtey, A. and Stukenbrock, E. H. (2018), 'Interspecific gene exchange as a driver of adaptive evolution in fungi', *Annual Review of Microbiology* **72**(1), 377–398. PMID: 29927707.

Fisher, K. J., Buskirk, S. W., Vignogna, R. C., Marad, D. A. and Lang, G. I. (2018), 'Adaptive genome duplication affects patterns of molecular evolution in Saccharomyces cerevisiae', *PLOS Genetics* **14**(5), e1007396.

Forche, A., Steinbach, M. and Berman, J. (2009), 'Efficient and rapid identification of Candida albicans allelic status using SNP-RFLP.', *FEMS yeast research* **9**(7), 1061–1069.

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. and Verstrepen, K. J. (2016), 'Domestication and divergence of Saccharomyces cerevisiae beer yeasts', *Cell* **166**(6), 1397–1410.e16.

García-Ríos, E., Guillén, A., de la Cerda, R., Pérez-Través, L., Querol, A. and Guillamón, J. M. (2019), 'Improving the cryotolerance of wine yeast by interspecific hybridization in the genus Saccharomyces', *Frontiers in Microbiology* **9**, 3232.

García-Ríos, E., Gutiérrez, A., Salvadó, Z., Arroyo-López, F. N. and Guillamon, J. M. (2014), 'The fitness advantage of commercial wine yeasts in relation to the nitrogen concentration, temperature, and ethanol content under microvinification conditions', *Applied and Environmental Microbiology* **80**(2), 704–713.

García-Ríos, E., Querol, A. and Guillamón, J. M. (2016), 'iTRAQ-based proteome profiling of Saccharomyces cerevisiae and cryotolerant species Saccharomyces uvarum and Saccharomyces kudriavzevii during low-temperature wine fermentation', *Journal of Proteomics* **146**, 70–79.

García Ríos, E., Nuévalos, M., Barrio, E., Puig, S. and Guillamón, J. M. (2019), 'A new chromosomal rearrangement improves the adaptation of wine yeasts to sulfite', *Environmental Microbiology* **21**(5), 1771–1781.

Gerstein, A. C., Fu, M. S., Mukaremera, L., Li, Z., Ormerod, K. L., Fraser, J. A., Berman, J. and Nielsen, K. (2015), 'Polyploid titan cells produce haploid and aneuploid progeny to promote stress adaptation.', *mBio* **6**(5), e01340–15.

Gibson, B., Geertman, J.-M. A., Hittinger, C. T., Krogerus, K., Libkind, D., Louis, E. J., Magalhães, F. and Sampaio, J. P. (2017), 'New yeasts - new brews: modern approaches to brewing yeast design and development', *FEMS Yeast Research* **17**(4). fox038.

Gibson, B. R., Storgårds, E., Krogerus, K. and Vidgren, V. (2013), 'Comparative physiology and fermentation performance of Saaz and Frohberg lager yeast strains and the parental species Saccharomyces eubayanus', *Yeast* **30**(7), 255–266.

Gilchrist, C. and Stelkens, R. (2019), 'Aneuploidy in yeast: Segregation error or adaptation mechanism?', *Yeast* **36**(9), 525–539.

Giudici, P., Solieri, L., Pulvirenti, A. M. and Cassanelli, S. (2005), 'Strategies and perspectives for genetic improvement of wine yeasts', *Applied Microbiology and Biotechnology* **66**(6), 622–628.

Goddard, M. R. and Greig, D. (2015), 'Saccharomyces cerevisiae: a nomadic yeast with no niche?', *FEMS Yeast Research* **15**(3).

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. and Oliver, S. G. (1996), 'Life with 6000 Genes', *Science* **274**(5287), 546–567.

Gonçalves, M., Pontes, A., Almeida, P., Barbosa, R., Serra, M., Libkind, D., Hutzler, M., Gonçalves, P. and Sampaio, P. (2016), 'Distinct domestication trajectories in top-fermenting beer yeasts and wine yeasts', *Current Biology* (26), 1–12.

Gonçalves, P., Valério, E., Correia, C., de Almeida, J. M. G. C. F. and Sampaio, J. P. (2011), 'Evidence for divergent evolution of growth temperature preference in sympatric Saccharomyces species', *PLoS ONE* **6**(6), e20739.

González, S. S., Barrio, E. and Querol, A. (2008), 'Molecular characterization of new natural hybrids of Saccharomyces cerevisiae and S. kudriavzevii in brewing.', *Applied and environmental microbiology* **74**(8), 2314–20.

Gordon, D. and Green, P. (2013), 'Consed: A graphical editor for next-generation sequencing', *Bioinformatics* **29**(22), 2936–2937.

Gorter de Vries, A. R., Pronk, J. T. and Daran, J.-M. G. (2017), 'Industrial relevance of chromosomal copy number variation in saccharomyces yeasts', *Applied and Environmental Microbiology* **83**(11).

Gorter, F. A., Derks, M. F., Van Den Heuvel, J., Aarts, M. G., Zwaan, B. J., De Ridder, D. and De Visser, J. A. G. (2017), 'Genomics of adaptation depends on the rate of environmental change in experimental yeast populations', *Molecular Biology and Evolution* **34**(10), 2613–2626.

Goulet, B. E., Roda, F. and Hopkins, R. (2017), 'Hybridization in plants: old ideas, new techniques.', *Plant physiology* **173**(1), 65–78.

Govindaraju, D. R. (2019), 'An elucidation of over a century old enigma in genetics—Heterosis', *PLOS Biology* **17**(4), e3000215.

Greig, D., Louis, E. J., Borts, R. H. and Travisano, M. (2002), 'Hybrid speciation in experimental populations of yeast', *Science* **298**(5599), 1773 LP – 1775.

Guillamón, J. M. and Barrio, E. (2017), 'Genetic polymorphism in wine yeasts: Mechanisms and methods for its detection', *Frontiers in Microbiology* **8**, 806.

Hagman, A. and Piškur, J. (2015), 'A study on the fundamental mechanism and the evolutionary driving forces behind aerobic fermentation in yeast', *PLOS ONE* **10**(1), e0116942.

Harrison, R. G. (1993), *Hybrid zones and the evolutionary process*, Oxford University Press.

Hebly, M., Brickwedde, A., Bolat, I., Driessen, M. R., de Hulster, E. A., van den Broek, M., Pronk, J. T., Geertman, J.-M., Daran, J.-M. and Daran-Lapujade, P. (2015), 'S. cerevisiae × S. eubayanus interspecific hybrid, the best of both worlds and beyond', *FEMS Yeast Research* **15**(3). fov005.

Hittinger, C. T. (2013), 'Saccharomyces diversity and evolution: a budding model genus', *Trends in Genetics* **29**(5), 309–317.

Homer, N., Merriman, B. and Nelson, S. F. (2009), 'BFAST: an alignment tool for large scale genome resequencing', *PLoS ONE* **4**(11), 1–12.

Hung, J.-H. and Weng, Z. (2016), 'Sequence alignment and homology search with BLAST and ClustalW', *Cold Spring Harbor protocols* **2016**(11), pdb.prot093088.

Huson, D. H. and Bryant, D. (2006), 'Application of phylogenetic networks in evolutionary studies', *Molecular Biology and Evolution* **23**(2), 254–267.

James, T. Y., Michelotti, L. A., Glasco, A. D., Clemons, R. A., Powers, R. A., James, E. S., Simmons, D. R., Bai, F. and Ge, S. (2019), 'Adaptation by loss of heterozygosity in saccharomyces cerevisiae clones under divergent selection', *Genetics* **213**(2), 665–683.

Jansen, G., Wu, C., Schade, B., Thomas, D. Y. and Whiteway, M. (2005), 'Drag&Drop cloning in yeast', *Gene* **344**, 43–51.

Joshi, N. and Fass, J. (2011), 'Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files'.
URL: *https://github.com/najoshi/sickle*

Katoh, K. and Standley, D. M. (2013), 'MAFFT multiple sequence alignment software version 7: Improvements in performance and usability', *Molecular Biology and Evolution* **30**(4), 772–780.

Keane, O. M., Toft, C., Carretero-Paulet, L., Jones, G. W. and Fares, M. A. (2014), 'Preservation of genetic and regulatory robustness in ancient gene duplicates of Saccharomyces cerevisiae', *Genome research* **24**(11), 1830–1841.

Kodama, Y., Kielland-Brandt, M. C. and Hansen, J. (2006), Lager brewing yeast, *in* 'Comparative genomics: using fungi as models', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 145–164.

Krause, D. J., Kominek, J., Opulente, D. A., Shen, X.-X., Zhou, X., Langdon, Q. K., DeVirgilio, J., Hulfachor, A. B., Kurtzman, C. P., Rokas, A. and Hittinger, C. T. (2018), 'Functional and evolutionary characterization of a secondary metabolite gene cluster in budding yeasts.', *Proceedings of the National Academy of Sciences of the United States of America* **115**(43), 11030–11035.

Krogerus, K., Magalhães, F., Vidgren, V. and Gibson, B. (2015), 'New lager yeast strains generated by interspecific hybridization', *Journal of Industrial Microbiology and Biotechnology* **42**(5), 769–778.

Krogerus, K., Magalhães, F., Vidgren, V. and Gibson, B. (2017), 'Novel brewing yeast hybrids: creation and application', *Applied Microbiology and Biotechnology* **101**(1), 65–78.

Krogerus, K., Preiss, R. and Gibson, B. (2018), 'A Unique Saccharomyces cerevisiae × Saccharomyces uvarum hybrid isolated From Norwegian farmhouse beer: Characterization and reconstruction', *Frontiers in Microbiology* **9**, 2253.

Kumaran, R., Yang, S. Y. and Leu, J. Y. (2013), 'Characterization of Chromosome Stability in Diploid, Polyploid and Hybrid Yeast Cells', *PLoS ONE* **8**(7).

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C. and Salzberg, S. L. (2004), 'Versatile and open software for comparing large genomes.', *Genome biology* **5**(2), R12.

Kurtzman, C. P. and Sugiyama, J. (2015), Saccharomycotina and Taphrinomycotina: The yeasts and yeastlike fungi of the Ascomycota, *in* 'Systematics and Evolution', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 3–33.

Kutyna, D. R., Cordente, A. G. and Varela, C. (2014), Genetic engineering of industrial Saccharomyces cerevisiae strains using a selection/sounter-selection approach, Humana Press, New York, NY, pp. 157–168.

Lancaster, S. M., Payen, C., Smukowski Heil, C. and Dunham, M. J. (2019), 'Fitness benefits of loss of heterozygosity in Saccharomyces hybrids.', *Genome research* .

Langmead, B. and Salzberg, S. L. (2012), 'Fast gapped-read alignment with Bowtie 2', *Nature Methods* **9**(4), 357–359.

Leducq, J.-B., Nielly-Thibault, L., Charron, G., Eberlein, C., Verta, J.-P., Samani, P., Sylvester, K., Hittinger, C. T., Bell, G. and Landry, C. R. (2015), 'Speciation driven by hybridization and chromosomal plasticity in a wild yeast', *Nature Microbiology* **1**(1), 027383.

Legras, J. L., Erny, C. and Charpentier, C. (2014), 'Population structure and comparative genome hybridization of European flor yeast reveal a unique group of Saccharomyces cerevisiae strains with few gene duplications in their genome', *PLoS ONE* **9**(10), 1–14.

Legras, J.-L., Galeote, V., Bigey, F., Camarasa, C., Marsit, S., Nidelet, T., Sanchez, I., Couloux, A., Guy, J., Franco-Duarte, R., Marcet-Houben, M., Gabaldon, T., Schuller, D., Sampaio, J. P., Dequin, S. and Wittkopp, P. (2018), 'Adaptation of S. cerevisiae to fermented food environments reveals remarkable genome plasticity and the footprints of domestication', *Molecular Biology and Evolution* **35**(7), 1712–1727.

Legras, J. L., Moreno-Garcia, J., Zara, S., Zara, G., Garcia-Martinez, T., Mauricio, J. C., Mannazzu, I., Coi, A. L., Zeidan, M. B., Dequin, S., Moreno, J. and Budroni, M. (2016), 'Flor yeast: New perspectives beyond wine aging', *Frontiers in Microbiology* **7**, 503.

Letunic, I. and Bork, P. (2016), 'Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees', *Nucleic Acids Research* **44**(W1), W242–W245.

Libkind, D., Hittinger, C. T., Valério, E., Gonçalves, C., Dover, J., Johnston, M., Gonçalves, P. and Sampaio, J. P. (2011), 'Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast.', *Proceedings of the National Academy of Sciences* **108**(35), 14539–14544.

Liti, G., Barton, D. B. H. and Louis, E. J. (2006), 'Sequence diversity, reproductive isolation and species concepts in Saccharomyces', *Genetics* **174**(2), 839–50.

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., Quail, M. a., Goodhead, I., Sims, S., Smith, F., Blomberg, A., Durbin, R. and Louis, E. J. (2009), 'Population genomics of domestic and wild yeasts', *Nature* **458**(7236), 337–341.

Llorente, B., Smith, C. E. and Symington, L. S. (2008), 'Break-induced replication: What is it and what is it for?', *Cell Cycle* **7**(7), 859–864.

Lopandic, K., Tafer, H. and Sterflinger, K. (2018), 'Draft genome sequence of the Saccharomyces cerevisiae × Saccharomyces kudriavzevii HA1836 interspecies hybrid yeast', *Genome announcements* **6**(20), e00343–18.

Love, M. I., Huber, W. and Anders, S. (2014), 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology* **15**(12), 550.

Macías, L. G., Morard, M., Toft, C. and Barrio, E. (2019), 'Comparative genomics between Saccharomyces kudriavzevii and S. cerevisiae applied to identify mechanisms involved in adaptation', *Frontiers in Genetics* **10**, 187.

Magwene, P. M., Willis, J. H. and Kelly, J. K. (2011), 'The statistics of bulk segregant analysis using next generation sequencing', *PLoS Computational Biology* **7**(11), 1–9.

Mallet, J. (1995), 'A species definition for the modern synthesis', *Trends in Ecology & Evolution* **10**(7), 294–299.

Mangado, A., Morales, P., Gonzalez, R. and Tronchoni, J. (2018), 'Evolution of a yeast with industrial background under winemaking conditions leads to diploidization and chromosomal copy number variation', *Frontiers in Microbiology* **9**, 1816.

Marad, D. A., Buskirk, S. W. and Lang, G. I. (2018), 'Altered access to beneficial mutations slows adaptation and biases fixed mutations in diploids', *Nature Ecology & Evolution* **2**(5), 882–889.

Marcet-Houben, M. and Gabaldón, T. (2015), 'Beyond the whole-genome duplication: Phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage', *PLoS Biology* **13**(8), 1–26.

Marsit, S., Leducq, J.-B., Durand, É., Marchant, A., Filteau, M. and Landry, C. R. (2017), 'Evolutionary biology through the lens of budding yeast comparative genomics', *Nature Reviews Genetics* **18**(10), 581–598.

Mattenberger, F., Sabater-Muñoz, B., Toft, C., Sablok, G. and Fares, M. A. (2017), 'Expression properties exhibit correlated patterns with the fate of duplicated genes, their divergence, and transcriptional plasticity in Saccharomycotina', *DNA Research* **24**(6), 559–570.

Mavárez, j. and Linares, M. (2008), 'Homoploid hybrid speciation in animals', *Molecular Ecology* **17**(19), 4181–4185.

Mayr, E. (1970), *Populations, Species and Evolution*, Harvard University Press, Cambridge.

Monerawela, C. and Bond, U. (2016), 'Brewing up a storm: The genomes of lager yeasts and how they evolved', *Biotechnology Advances* **35**(4), 512–519.

Monerawela, C. and Bond, U. (n.d.), 'Recombination sites on hybrid chromosomes in Saccharomyces pastorianus share common sequence motifs and define a complex evolutionary relationship between group I and II lager yeasts', *FEMS Yeast Research* (5).

Morales, L. and Dujon, B. (n.d.), *Microbiology and Molecular Biology Reviews* .

Morard, M., Macías, L. G., Adam, A. C., Lairón-Peris, M., Pérez-Torrado, R., Toft, C. and Barrio, E. (2019), 'Aneuploidy and ethanol tolerance in Saccharomyces cerevisiae', *Frontiers in Genetics* **10**, 82.

Naranjo-Ortiz, M. A. and Gabaldón, T. (2019), 'Fungal evolution: major ecological adaptations and evolutionary transitions', *Biological Reviews* **94**(4), 1443–1476.

Naseeb, S., James, S. A., Alsammar, H., Michaels, C. J., Gini, B., Nueno-Palop, C., Bond, C. J., McGhie, H., Roberts, I. N. and Delneri, D. (2017), 'Saccharomyces jurei sp. nov., isolation and genetic identification of a novel yeast species from Quercus robur', *International Journal of Systematic and Evolutionary Microbiology* **67**(6), 2046–2052.

Natesuntorn, W., Iwami, K., Matsubara, Y., Sasano, Y., Sugiyama, M., Kaneko, Y. and Harashima, S. (2015), 'Genome-wide construction of a series of designed segmental aneuploids in Saccharomyces cerevisiae.', *Scientific reports* **5**(1), 12510.

Naumov, G. (1996), 'Genetic identification of biological species in theSaccharomyces sensu stricto complex', *Journal of Industrial Microbiology* **17**(3), 295–302.

Naumov, G. I., Naumova, E. S. and Sniegowski, P. D. (1998), 'Saccharomyces paradoxus and Saccharomyces cerevisiae are associated with exudates of North American oaks.', *Canadian Journal of Microbiology* **44**(11), 1045–1050.

Naumov, G., Naumova, E., Masneuf, I., Aigle, M., Kondratieva, V. and Dubourdieu, D. (2000), 'Natural polyploidization of some cultured yeast Saccharomyces sensu stricto: auto- and allotetraploidy', *Systematic and Applied Microbiology* **23**(3), 442–449.

Naumova, E. S., Roberts, I. N., James, S. A., Naumov, G. I. and Louis, E. J. (2000), 'Three new species in the Saccharomyces sensu stricto complex: Saccharomyces cariocanus, Saccharomyces kudriavzevii and Saccharomyces mikatae.', *International Journal of Systematic and Evolutionary Microbiology* **50**(5), 1931–1942.

Navarro-Tapia, E., Nana, R. K., Querol, A. and Pérez-Torrado, R. (2016), 'Ethanol cellular defense induce unfolded protein response in yeast', *Frontiers in Microbiology* **7**(1), 189.

Navarro-Tapia, E., Perez-Torrado, R. and Querol, A. (2017), 'Ethanol effects involve non-canonical unfolded protein response activation in yeast cells', *Frontiers in Microbiology* **8**(1), 1–12.

Nilsson-Tillgren, T., Gjermansen, C., Kielland-Brandt, M. C., Petersen, J. G. L. and Holmberg, S. (1981), 'Genetic differences between Saccharomyces carlsbergensis and S. cerevisiae. Analysis of chromosome III by single chromosome transfer', *Carlsberg Research Communications* **46**(1), 65.

Origone, A. C., Rodríguez, M. E., Oteiza, J. M., Querol, A. and Lopes, C. A. (2018), 'Saccharomyces cerevisiae × Saccharomyces uvarum hybrids generated under different conditions share similar winemaking features', *Yeast* **35**(1), 157–171.

Ortiz-Merino, R. A., Kuanyshev, N., Braun-Galleani, S., Byrne, K. P., Porro, D., Branduardi, P. and Wolfe, K. H. (2017), 'Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch', *PLOS Biology* **15**(5), e2002128.

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. and Morrissey, J. P. (2018), 'Ploidy variation in Kluyveromyces marxianus separates dairy and non-dairy isolates', *Frontiers in Genetics* **9**(1), 1–16.

Ortiz-Tovar, G., Pérez-Torrado, R., Adam, A. C., Barrio, E. and Querol, A. (2018), 'A comparison of the performance of natural hybrids Saccharomyces cerevisiae × Saccharomyces kudriavzevii at low temperatures reveals the crucial role of their S. kudriavzevii genomic contribution', *International Journal of Food Microbiology* **274**(1), 12–19.

Ortiz Tovar, M. G. (2018), CARACTERIZACIÓN ENOLÓGICA Y MOLECULAR DE HÍBRIDOS NATURALES Y ARTIFICIALES ENTRE LAS ESPECIES SACCHAROMYCES CEREVISIAE Y SACCHAROMYCES KUDRIAVZEVII, PhD thesis, Universitat de València.

Otto, S. P. (2007), 'The evolutionary consequences of polyploidy', *Cell* **131**(3), 452–462.

Otto, T. D., Dillon, G. P., Degrave, W. S. and Berriman, M. (2011), 'RATT: Rapid annotation transfer tool', *Nucleic acids research* **39**(9), e57.

Payseur, B. A. and Rieseberg, L. H. (2016), 'A genomic perspective on hybridization and speciation', *Molecular Ecology* **25**(11), 2337–2360.

Pérez-Ortín, J. E., Querol, A., Puig, S. and Barrio, E. (2002), 'Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains.', *Genome research* **12**(10), 1533–9.

Pérez-Torrado, R., Barrio, E. and Querol, A. (2017), 'Alternative yeasts for winemaking: Saccharomyces non- cerevisiae and its hybrids', *Critical Reviews in Food Science and Nutrition* **58**(11), 1780–1790.

Pérez-Torrado, R., González, S. S., Combina, M., Barrio, E. and Querol, A. (2015), 'Molecular and enological characterization of a natural Saccharomyces uvarum and Saccharomyces cerevisiae hybrid', *International Journal of Food Microbiology* **204**, 101–110.

Pérez-Través, L., Lopes, C. A., Barrio, E. and Querol, A. (2012), 'Evaluation of different genetic procedures for the generation of artificial hybrids in Saccharomyces genus for winemaking', *International Journal of Food Microbiology* **156**(2), 102–111.

Pérez-Través, L., Lopes, C. A., Querol, A. and Barrio, E. (2014), 'On the complexity of the Saccharomyces bayanus taxon: Hybridization and potential hybrid speciation', *PLoS ONE* **9**(4), e93729.

Pérez-Través, L., Querol, A. and Pérez-Torrado, R. (2016), 'Increased mannoprotein content in wines produced by Saccharomyces kudriavzevii×Saccharomyces cerevisiae hybrids', *International Journal of Food Microbiology* **237**, 35–38.

Peris, D., Arias, A., Orlić, S., Belloch, C., Pérez-Través, L., Querol, A. and Barrio, E. (2017), 'Mitochondrial introgression suggests extensive ancestral hybridization events among Saccharomyces species', *Molecular Phylogenetics and Evolution* **108**, 49–60.

Peris, D., Lopes, C. a., Arias, A. and Barrio, E. (2012), 'Reconstruction of the evolutionary history of Saccharomyces cerevisiae x S. kudriavzevii hybrids based on multilocus sequence analysis', *PLoS ONE* **7**(9), 1–14.

Peris, D., Lopes, C. a., Belloch, C., Querol, A. and Barrio, E. (2012*a*), 'Comparative genomics among Saccharomyces cerevisiae × Saccharomyces kudriavzevii natural hybrid strains isolated from wine and beer reveals different origins', *BMC Genomics* **13**(1), 407.

Peris, D., Lopes, C. A., Belloch, C., Querol, A. and Barrio, E. (2012*b*), 'Comparative genomics among Saccharomyces cerevisiae × Saccharomyces kudriavzevii natural hybrid strains isolated from wine and beer reveals different origins', *BMC Genomics* **13**(1), 407.

Peris, D., Pérez-Torrado, R., Hittinger, C. T., Barrio, E. and Querol, A. (2017), 'On the origins and industrial applications of Saccharomyces cerevisiae x Saccharomyces kudriavzevii hybrids', *Yeast* **35**(1), 51–69.

Peris, D., Sylvester, K., Libkind, D., Gon??alves, P., Sampaio, J. P., Alexander, W. G. and Hittinger, C. T. (2014), 'Population structure and reticulate evolution of Saccharomyces eubayanus and its lager-brewing hybrids', *Molecular Ecology* **23**(8), 2031–2045.

Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergstrom, A., Sigwalt, A., Barré, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J. M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Liti, G. and Schacherer, J. (2018), 'Genome evolution across 1,011 Saccharomyces cerevisiae isolates', *Nature* **556**(7701), 339–344.

Pfliegler, W. P., Atanasova, L., Karanyicz, E., Sipiczki, M., Bond, U., Druzhinina, I. S., Sterflinger, K. and Lopandic, K. (2014), 'Generation of new genotypic and phenotypic features in artificial and natural yeast hybrids', *Food Technology and Biotechnology* **52**(1), 46–57.

Piškur, J., Rozpędowska, E., Polakova, S., Merico, A. and Compagno, C. (2006), 'How did Saccharomyces evolve to become a good brewer?', *Trends in Genetics* **22**(4), 183–186.

Pontes, A., Čadež, N., Gonçalves, P. and Sampaio, J. P. (2019), 'A Quasi-Domesticate Relic Hybrid Population of Saccharomyces cerevisiae × S. paradoxus Adapted to Olive Brine', *Frontiers in Genetics* **10**, 449.

Pulvirenti, A. and Zambonelli, C. and Todaro, A.and Giudici, P. (2002), 'Interspecific hybridisation by digestive tract of invertebrates as a source of environmental biodiversity within the Saccharomyces cerevisiae', *Annals of Microbiology* **52**(3), 245–255.

Querol, A., Barrio, E., Huerta, T. and Ramón, D. (1992), 'Molecular monitoring of wine fermentations conducted by active dry yeast strains.', *Applied and Environmental Microbiology* **58**(9), 2948–2953.

Querol, A., Fernández-Espinar, M. T., Del Olmo, M. and Barrio, E. (2003), 'Adaptive evolution of wine yeast', *International Journal of Food Microbiology* **86**(1), 3–10.

Quinlan, A. R. and Hall, I. M. (2010), 'BEDTools: a flexible suite of utilities for comparing genomic features', *Bioinformatics* **26**(6), 841–842.

Ravichandran, M. C., Fink, S., Clarke, M. N., Hofer, F. C. and Campbell, C. S. (2018), 'Genetic interactions between specific chromosome copy number alterations dictate complex aneuploidy patterns.', *Genes & development* **32**(23-24), 1485–1498.

Rodríguez, M. E., Pérez-Través, L., Sangorrín, M. P., Barrio, E. and Lopes, C. A. (2014), 'Saccharomyces eubayanus and saccharomyces uvarum associated with the fermentation of araucaria araucana seeds in patagonia', *FEMS Yeast Research* **14**(6), 948–965.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A. and Barrell, B. (2000), 'Artemis: sequence visualization and annotation', *Bioinformatics* **16**(10), 944–945.

Salvadó, Z., Arroyo-López, F. N., Guillamón, J. M., Salazar, G., Querol, A. and Barrio, E. (2011), 'Temperature adaptation markedly determines evolution within the genus Saccharomyces.', *Applied and environmental microbiology* **77**(7), 2292–22302.

Scott, A. L., Richmond, P. A., Dowell, R. D. and Selmecki, A. M. (2017), 'The influence of polyploidy on the evolution of yeast grown in a sub-optimal carbon source', *Molecular Biology and Evolution* **34**(10), 2690–2703.

Selmecki, A., Gerami-Nejad, M., Paulson, C., Forche, A. and Berman, J. (2008), 'An isochromosome confers drug resistance in vivo by amplification of two genes, ERG11 and TAC1', *Molecular Microbiology* **68**(3), 624–641.

Shen, X.-X., Opulente, D. A., Kominek, J., Zhou, X., Steenwyk, J. L., Buh, K. V., Haase, M. A., 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.-i., Čadež, N., Libkind, D., Rosa, C. A., DeVirgilio, J., Hulfachor, A. B., Groenewald, M., Kurtzman, C. P., Hittinger, C. T. and Rokas, A. (2018), 'Tempo and mode of genome evolution in the budding yeast subphylum', *Cell* (175), 1–13.

Sipiczki, M. (2008), 'Interspecies hybridization and recombination in Saccharomyces wine yeasts', *FEMS Yeast Research* **8**(7), 996–1007.

Sipiczki, M. (2018*a*), 'Interspecies hybridisation and genome chimerisation in Saccharomyces: Combining of gene pools of species and its biotechnological perspectives', *Frontiers in Microbiology* **9**, 3071.

Sipiczki, M. (2018*b*), 'Interspecies Hybridisation and Genome Chimerisation in Saccharomyces: Combining of Gene Pools of Species and Its Biotechnological Perspectives', *Frontiers in Microbiology* **9**, 3071.

Smukowski Heil, C. S., Large, C. R. L., Patterson, K., Hickey, A. S.-M., Yeh, C.-L. C. and Dunham, M. J. (2019), 'Temperature preference can bias parental genome retention during hybrid evolution', *PLOS Genetics* **15**(9), e1008383.

Snoek, T., Verstrepen, K. J. and Voordeckers, K. (2016), 'How do yeast cells become tolerant to high ethanol concentrations?', *Current Genetics* **62**(3), 475–480.

Špírek, M., Poláková, S., Jatzová, K. and Sulo, P. (2015), 'Post-zygotic sterility and cytonuclear compatibility limits in S. cerevisiae xenomitochondrial cybrids', *Frontiers in Genetics* **5**, 454.

Sprouffske, K. and Wagner, A. (2016), 'Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves', *BMC Bioinformatics* **17**(1), 172.

Stamatakis, A. (2014), 'RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies', *Bioinformatics* **30**(9), 1312–1313.

Stanke, M. and Morgenstern, B. (2005), 'AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints.', *Nucleic Acids Research* **33**(suppl_2), W465–W467.

Stefanini, I., Dapporto, L., Berná, L., Polsinelli, M., Turillazzi, S. and Cavalieri, D. (2016), 'Social wasps are a Saccharomyces mating nest', *Proceedings of the National Academy of Sciences* **113**(8), 2247–2251.

Stelkens, R. B., Brockhurst, M. A., Hurst, G. D. D. and Greig, D. (2014), 'Hybridization facilitates evolutionary rescue', *Evolutionary Applications* **7**(10), 1209–1217.

Stukenbrock, E. H. (2016), 'The role of hybridization in the evolution and emergence of new fungal plant pathogens', *Phytopathology* **106**(2), 104–112.

Su, Y., Gamero, A., Rodríguez, M. E., Lopes, C. A., Querol, A. and Guillamón, J. M. (2019), 'Interspecific hybridisation among diverse Saccharomyces species: A combined biotechnological solution for low-temperature and nitrogen-limited wine fermentations', *International Journal of Food Microbiology* **310**, 108331.

Tattini, L., Tellini, N., Mozzachiodi, S., D'Angiolo, M., Loeillet, S., Nicolas, A. and Liti, G. (2019), 'Accurate tracking of the mutational landscape of diploid hybrid genomes', *Molecular Biology and Evolution* .

Thomson, J. M., Gaucher, E. A., Burgan, M. F., De Kee, D. W., Li, T., Aris, J. P. and Benner, S. A. (2005), 'Resurrecting ancestral alcohol dehydrogenases from yeast', *Nature Genetics* **37**(6), 630–635.

Tilakaratna, V. and Bensasson, D. (2017), 'Habitat predicts levels of genetic admixture in Saccharomyces cerevisiae', *G3: Genes|Genomes|Genetics* **7**(9), 2919–2929.

Todd, R. T., Forche, A. and Selmecki, A. (2017), Ploidy variation in fungi: Polyploidy, aneuploidy, and genome evolution, *in* 'The Fungal Kingdom', American Society of Microbiology, pp. 599–618.

Torres, E. M., Sokolsky, T., Tucker, C. M., Chan, L. Y., Boselli, M., Dunham, M. J. and Amon, A. (2007), 'Effects of aneuploidy on cellular physiology and cell division in haploid yeast.', *Science* **317**(5840), 916–24.

Tronchoni, J., Medina, V., Guillamón, J., Querol, A. and Pérez-Torrado, R. (2014), 'Transcriptomics of cryophilic Saccharomyces kudriavzevii reveals the key role of gene translation efficiency in cold stress adaptations', *BMC Genomics* **15**(1), 432.

Vaughan Martini, A. and Kurtzman, C. P. (1985), 'Deoxyribonucleic acid relatedness among species of the genus Saccharomyces sensu stricto', *International Journal of Systematic Bacteriology* **35**(4), 508–511.

Voordeckers, K., Kominek, J., Das, A., Espinosa-Cantú, A., De Maeyer, D., Arslan, A., Van Pee, M., van der Zande, E., Meert, W., Yang, Y., Zhu, B., Marchal, K., DeLuna, A., Van Noort, V., Jelier, R. and Verstrepen, K. J. (2015), 'Adaptation to high ethanol reveals complex evolutionary pathways', *PLoS Genetics* **11**(11), e1005635.

Voordeckers, K. and Verstrepen, K. J. (2015), 'Experimental evolution of the model eukaryote Saccharomyces cerevisiae yields insight into the molecular mechanisms underlying adaptation', *Current Opinion in Microbiology* **28**(Figure 1), 1–9.

Wendland, J. (2014), 'Lager yeast comes of age', *Eukaryotic cell* **13**(10), 1256–1265.

Wickham, H. (2009), *ggplot2*, Springer New York, New York, NY.

Wiley, E. O. (1978), 'The evolutionary species concept reconsidered', *Systematic Zoology* **27**(1), 17.

Yakimowski, S. B. and Rieseberg, L. H. (2014), 'The role of homoploid hybridization in evolution: A century of studies synthesizing genetics and ecology', *American Journal of Botany* **101**(8), 1247–1258.

Yona, A. H., Manor, Y. S., Herbst, R. H., Romano, G. H., Mitchell, A., Kupiec, M., Pilpel, Y. and Dahan, O. (2012), 'Chromosomal duplication is a transient evolutionary solution to stress', *Proceedings of the National Academy of Sciences* **109**(51), 21010–21015.

Yu, G., Smith, D. K., Zhu, H., Guan, Y. and Lam, T. T.-Y. (2017), 'ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data', *Methods in Ecology and Evolution* **8**(1), 28–36.

Zerbino, D. R. and Birney, E. (2008), 'Velvet: algorithms for de novo short read assembly using de Bruijn graphs.', *Genome research* **18**(5), 821–829.

Zhang, K., Zhang, L. J., Fang, Y. H., Jin, X. N., Qi, L., Wu, X. C. and Zheng, D. Q. (2016), 'Genomic structural variation contributes to phenotypic change of industrial bioethanol yeast Saccharomyces cerevisiae', *FEMS Yeast Research* **16**(2), 1–12.

Zhang, Z., Bendixsen, D. P., Janzen, T., Nolte, A. W., Greig, D. and Stelkens, R. (2019), 'Recombining your way out of trouble: The genetic architecture of hybrid fitness under environmental stress', *Molecular Biology and Evolution* . msz211.

Zhu, J., Pavelka, N., Bradford, W. D., Rancati, G. and Li, R. (2012), 'Karyotypic determinants of chromosome instability in aneuploid budding yeast', *PLoS Genetics* **8**(5), e1002719.