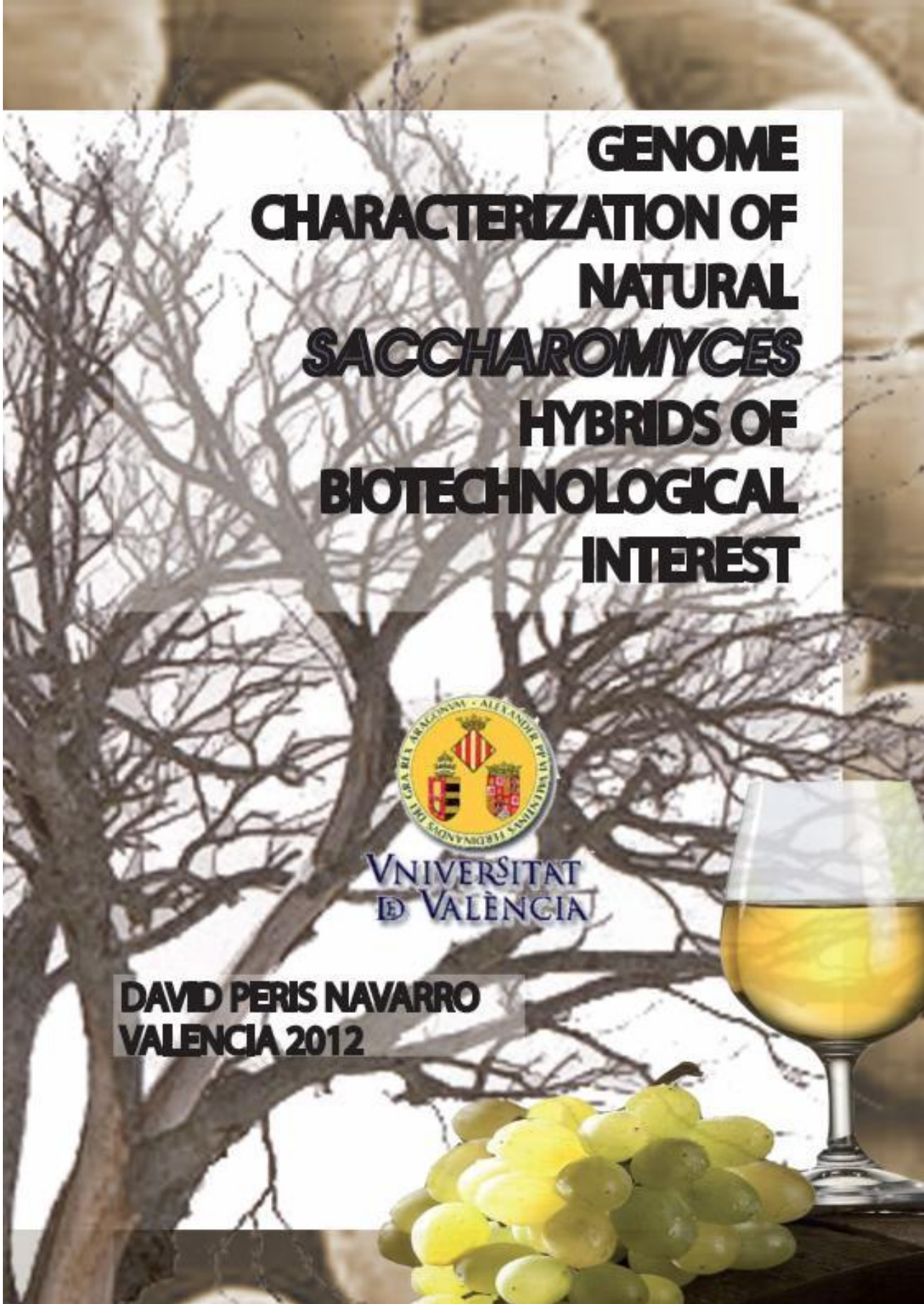
GENOME CHARACTERIZATION OF NATURAL *SACCHAROMYCES* HYBRIDS OF BIOTECHNOLOGICAL INTEREST



VNIVERSITAT DE VALÈNCIA

DAVID PERIS NAVARRO
VALENCIA 2012

VNIVERSITAT DE VALÈNCIA
Institut Cavanilles de Biodiversitat i Biologia Evolutiva

GENOME CHARACTERIZATION OF NATURAL *SACCHAROMYCES* HYBRIDS OF BIOTECHNOLOGICAL INTEREST

Tesis doctoral presentada por:

DAVID PERIS NAVARRO

para optar al grado de doctor en Biotecnología

Valencia, 12 de Septiembre 2012

VNIVERSITAT DE VALÈNCIA

Institut Cavanilles de Biodiversitat i Biologia Evolutiva

El Dr. Eladio Barrio Esparducer, Profesor Titular de Departamento de Genética de la Universitat de València y la Dra. Carmela Belloch Trinidad, Científico Titular del Consejo Superior de Investigaciones Científicas (CSIC) en el Instituto de Agroquímica y Tecnología de los Alimentos (IATA)

INFORMAN

Que la presente memoria "Genome characterization of natural *Saccharomyces* hybrids of biotechnological interest" constituye la tesis doctoral de Don. David Peris Navarro para optar al grado de doctor en 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, firmamos el presente informe en Valencia a 12 de SEPTIEMBRE de 2012

Fdo. Eladio Barrio Esparducer Fdo. Carmela Belloch Trinidad

ABSTRACT

At present, the genus *Saccharomyces* comprises seven species according to their patterns of breeding. The species boundaries are not clear due to the description of several reticulate events due to introgression and hybridization. In the last decade, new natural hybrids have been described in wine and brewing, such as *S. cerevisiae* x *S. kudriavzevii*. Due to new practices in wine and beer production, together with consequences in grape properties due to climatic change, led biotechnological companies to search for new yeast strains. In this context, hybrids have become of importance to biotechnological industries because they show good fermentative performance at low temperatures and produce new organoleptic compounds of industrial interest.

This doctoral thesis explores the evolution of the natural *S. cerevisiae* x *S. kudriavzevii* hybrids and the importance of hybridization in the evolution of the *Saccharomyces* species. This study was performed by using different molecular approaches combined with bioinformatic tools for phylogenetic tree/networks reconstruction and data analysis. Understanding the origin and genome characteristics of natural *S. cerevisiae* x *S. kudriavzevii* hybrids are our priority for obtaining, in the future, personalized yeasts with new properties of biotechnological interest.

AGRADECIMIENTOS

Cuando uno se sienta a escribir estas últimas líneas de su tesis, tiene la sensación de estar ya al final del camino de la primera fase en la vida como investigador. Una periodo inicial que, en el mejor de los casos, dura lo que su beca/contrato marca, y en otros un poquito más. En este camino, uno se encuentra con momentos de todo tipo, desde los más felices y divertidos a otros más difíciles. Durante ese recorrido mucha gente ha estado a mi lado, haciendo que los difíciles no parezcan tan difíciles, y donde los más felices se convertirán en algo inolvidable. Por ello, estas líneas van dedicadas a todos ellos.

En primer lugar, me gustaría agradecer a aquellas personas que han hecho que un sueño pudiera materializarse en forma de este trabajo. Gracias al Dr. Eladio Barrio y a la Dra. Amparo Querol por haber confiado en mí y haberme concedido la beca que abría la posibilidad de realizar esta tesis. Agradecer sobre todo a mi director de tesis, el Dr. Eladio Barrio, por las sentadas en el despacho para ir analizando y especulando sobre los datos generados, e ir definiendo los resultados finales. También tengo un especial agradecimiento hacia mi codirectora, la Dra. Carmela Belloch, quien ha estado para ayudarme en cualquier cosa que le he pedido, tanto a nivel profesional como moral. Carmela, has sido mi motor y he aprendido muchas cosas de ti, muchas gracias. Tampoco me gustaría olvidar, porque sería injusto, a las dos personas que me dieron la oportunidad de empezar mis primeros pasos en Ciencia. Guardo un bonito recuerdo de la Dra. Amparo Latorre y la Dra. Araceli Lamelas, quienes me permitieron hacer las primeras PCRs trabajando en el campo de la endosimbiosis. Y también agradecer a la Dra. Daniela Delneri y Dra. Elzbieta Piatkowska por permitirme una estancia en la Universidad de Manchester para profundizar en la proteómica de los híbridos.

Cuando uno llega al laboratorio por primera vez, se encuentra como perdido y con miedos, pero hay gente que te reconduce y consigues

rehacerte en cuestión de semanas. Por ello tengo que agradecer, al Dr. Armando Arias por esas primeras semanas o meses de aguante. También a todos los que fueron pasando por el 2.6.I; en especial a Chuy, con el que he pasado momentos increíbles. Además guardo un especial recuerdo de Christian, “el cosito”, con el que he compartido sentadas de bancada y muchos momentos buenos. Y como no, a mi amigo Sandi, quien ha sido para mí un gran apoyo. Todos ellos han sido grandes compañeros de laboratorio.

El equipo del lab 2.6.I siempre estuvo bien acompañado del laboratorio 307 y 303 del IATA y otra gente del “Cava”. Todos ellos han sido compañeros de batalla en la bancada, compartiendo programas y libros para poder avanzar en nuestras investigaciones. Todos juntos supimos desconectar y celebrar cenitas para poder evadirnos del trabajo.

Son muchos y necesitaría varias páginas para nombrarlos a todos, pero merecen mención especial Tronchoni, Silvia, Roberto, Amparito, Bruno, Jiří, Lupita, Anto, Rosana y María. Y, cómo no, David Lázaro, porque para mí ha sido un gran ejemplo de dedicación y del trabajo bien hecho.

Dar las gracias a mi amigo de pueblo, carrera y curro, Rafeta, por su apoyo. A Manzano, por su ayuda informática, y a todos aquellos con los que he pasado muy buenos momentos: Diego, Sergio, Noé, Vanesa, Conchi, Mireia, Teresa, Carmen, Elisa, Pilar, Joan.

Grandes en el camino han sido Quelo y Laurette, lástima que esta última estuviera sólo unos meses en el lab, pero ha sido muy divertido compartir todo tipo de momentos con ellos.

Y Laura, con la que he compartido discusiones científicas, festivas y de organización desde el minuto 0, y que sin nombrarla, mi tesis no estaría al 100% completa.

Además, nombrar a las personas que han sido un plus en mi recta final, por su grandísimo apoyo en esta tesis y que sin ellas no sé como hubiera acabado: Adri, María Noel, Eli y Clara. Hemos hecho una buena piña, la cual me encantaría mantener a pesar de la distancia.

Y no olvidar mi viaje a Manchester. Allí conocí a gente con la que pasé muy buenos momentos, como Toni, Elena, Bea, Ana, Thifeen y Samina, mis mancunianos favoritos. Ellos también forman parte de esta aventura.

Por último lugar y no menos importantes, agradecerles a mis padres el esfuerzo que han realizado a lo largo de mi vida, por haberme dado la oportunidad de alcanzar este nivel de estudios, el cual se vea culminado con esta tesis. También a mis suegros, a todos mis familiares y amigos, que han sido capaces de escuchar, alguna vez, la palabra híbridos, vino o cerveza, en clave evolutiva, y no han salido corriendo.

Y como no, a la persona más importante en mi vida. Esa persona que ha estado junto a mí desde el inicio hasta el final de este trabajo de investigación. La admiro por haber disfrutado junto a mí los mejores momentos y haber sabido aguantar y comprender los malos durante estos 5 años de tesis. Sin su apoyo no creo que hubiera sido capaz de finalizar en estas condiciones de calma, ya que ella es mi paz interior. Por todo ello, esta tesis va dedicada a mi mujer. Te agradezco de todo corazón el haber estado siempre a mi lado, para lo mejor y para lo peor, muchísimas gracias Penélope.

Con constancia y tenacidad se obtiene lo que se desea; la palabra imposible no tiene significado

(Napoleón Bonaparte)

Para Penélope

INDEX

RESUMEN AMPLIO

Introducción

El género *Saccharomyces* incluye siete especies, teniendo en cuenta el concepto biológico de especie: *Saccharomyces cerevisiae, Saccharomyces paradoxus, Saccharomyces cariocanus, Saccharomyces mikatae, Saccharomyces arboricolus, Saccharomyces kudriavzevii* y *Saccharomyces bayanus*. No obstante, los límites entre las especies del género no son claros debido a la presencia de eventos reticulados como Transferencia Génica Horizontal (HGT), introgresiones e hibridaciones.

En los últimos años se han descrito híbridos en el género, entre *S. cerevisiae*, *S. bayanus* y *S. kudriavzevii*. Los mayores esfuerzos se han centrado en investigar el híbrido de cerveza "lager", *Saccharomyces pastorianus*, originado por la hibridación entre las especies *S. cerevisiae* y *S. bayanus* (*S. eubayanus*). Sin embargo, en la última década se han aislado híbridos entre *S. cerevisiae* x *S. kudriavzevii* en vino y cerveza. La naturaleza quimérica de los cromosomas del genoma nuclear de los híbridos, se ha podido describir utilizando diferentes métodos moleculares, como los Polimorfismos de Longitud de Fragmentos de Restricción (RFLPs) y los chips de Hibridación Genómica Competitiva (aCGH). La caracterización fisiológica de las especies parentales *S. cerevisiae* y *S. kudriavzevii,* como también de los diferentes híbridos *S. cerevisiae* x *S. kudriavzevii,* ha demostrado el posible papel de cada parental en el genoma quimérico. Estos estudios revelaron que los híbridos tienen propiedades intermedias heredadas de ambos parentales. Así pues, la resistencia al etanol parece haber sido heredada de *S. cerevisiae* y la capacidad de crecer a bajas temperaturas y la producción de más glicerol podría haberse heredado del parental *S. kudriavzevii*.

El interés en innovar en el proceso de producción de vino y cerveza, está obligando a las compañías biotecnológicas a buscar nuevas cepas fermentadoras, que sean capaces de producir vino y cerveza con nuevas propiedades organolépticas. Además, se está buscando mantener las

propiedades aromáticas en ambos procesos, que se puede conseguir llevando a cabo fermentaciones a más bajas temperaturas. Por otro lado, el cambio climático está generando mostos de uva con mayor concentración de azúcares fermentables y alto pH, lo que influye en las propiedades organolépticas finales del vino.

Las nuevas prácticas en producción de vino y cerveza necesitan nuevas cepas para llevar a cabo una fermentación alcohólica exitosa bajo condiciones de fermentación donde se puedan mejorar las propiedades aromáticas. Además, los híbridos de levaduras se han convertido en cepas con un alto interés biotecnológico, ya que pueden fermentar a bajas temperaturas mientras producen muy buenos perfiles fermentativos. A esto hay que sumarle la capacidad de estos híbridos en desviar el metabolismo del azúcar a la producción de glicerol, solucionando los problemas de astringencia y alta concentración de etanol final como consecuencia del cambio climático.

Objetivos y Metodología

Entender la evolución y el posible role de la cepas parentales en los genomas de los híbridos es crucial para la producción futura de híbridos comerciales a la carta. El conocimiento del origen de los híbridos naturales y sus propiedades nos abrirá la posibilidad de generar híbridos con aquellas propiedades fermentativas que las empresas biotecnológicas deseen.

En esta tesis, nos centraremos en la estructura y evolución de los genomas de los híbridos *S. cerevisiae* x *S. kudriavzevii* como también en su origen. Los principales objetivos y la metodología utilizada se enumeran a continuación.

1. Identificación de nuevos híbridos naturales S. cerevisiae *x* S. kudriavzevii

Los híbridos *S. cerevisiae* x *S. kudriavzevii* utilizados en esta tesis han sido aislados de vino y cerveza producidos en los países del Centro de Europa y de la Zona Mediterránea Europea. La identificación de los nuevos híbridos naturales *S. cerevisiae* x *S. kudriavzevii* se realizó usando el método de los RFLPs, el cual ha sido previamente utilizado en la caracterización de otros híbridos *S. cerevisiae* x *S. kudriavzevii* de vino y cerveza. Alrededor de 35 genes distribuidos a lo largo de los dieciséis cromosomas de las especies de *Saccharomyces* fueron amplificados por PCR y digeridos con enzimas de restricción capaces de discriminar entre las diferentes especies. Esta metodología garantiza la identificación de los parentales *Saccharomyces* y como contribuyen a cada gen en cada cromosoma. Además podemos identificar si el híbrido comercial utilizado en producción de vino, sidra, suplemento dietético o aislados de pacientes clínicos, está formado por dos o más parentales.

Para apoyar los datos de la naturaleza doble o triple híbrida se secuenciaron 7 genes nucleares (*BRE5*, *CAT8*, *EGT2*, *GAL4*, *MET6*, *CYR1* y *CYC3*). La secuenciación de un gen mitocondrial (*COX2*) se realizó para conocer que parental ha contribuido al genoma mitocondrial. Las diferentes secuencias fueron ensambladas usando el paquete *Staden* y los alineamientos se realizaron en MEGA4 (o MEGA5). La reconstrucción del árbol filogenético de *BRE5* y *COX2* se hizo mediante el método de máxima verosimilitud (ML) implementado en el programa PhyML 3.0, usando los modelos evolutivos obtenidos en el programa jModeltest, para deducir la contribución de cada parental a los genes secuenciados.

2. Caracterización genómica de los híbridos naturales S. cerevisiae *x* S. kudriavzevii

Los chips de Hibridación Genómica Comparada (aCGH) se utilizaron con el objetivo de confirmar la contribución de cada genoma parental al genoma del híbrido y conocer las diferencias entre híbridos *S. cerevisiae* x *S. kudriavzevii* de cerveza y de vino. Además, determinar la estructura genómica de los híbridos fue clave para conocer qué mecanismo es más probable para la formación de estos híbridos.

La información obtenida por la técnica de aCGH está limitada al estudio del genoma parental de *S. cerevisiae* ya que los chips utilizados están basados en el genoma de *S. cerevisiae.* La hibridación se llevó a cabo a alta temperatura, 65ºC (temperatura restrictiva), donde sólo los genes de *S. cerevisiae* podrían hibridar. La técnica consiste en marcar el DNA genómico del híbrido con un fluoróforo, y la de *S. cerevisiae* (cepa de referencia) con otro. Los dos DNAs genómicos se hibridaron en el mismo chip a 65ºC. Después de la hibridación la fluorescencia de cada fluoróforo (genoma) se detectó mediante el escáner GenePix Personal 4100A. Diferencias en la intensidad de fluorescencia de cada marcaje nos permitió estimar el número de copias de cada gen de *S. cerevisiae* (ORF) en el híbrido comparado con la cepa de referencia (*S. cerevisiae* haploide). El programa ChARM nos permitió representar los datos del subgenoma *S. cerevisiae* en diagramas "caryoscopes" para cada cromosoma, detectándose aquí aneuploidías o segmentos con diferente número de copias (delecciones, duplicaciones o cromosomas quiméricos). La confirmación del número de cromosomas en el genoma de los híbridos tuvo que realizarse por citometría de flujo, usando para marcar el DNA el fluoróforo "SYTOX Green" y comparándola contra la señal obtenida en la cepa haploide de *S. cerevisiae*.

El genoma de *S. kudriavzevii* se encuentra en sintenia con el genoma de *S. cerevisiae*. Por tanto, podemos estimar la estructura de los

subgenomas de *S. cerevisiae* y *S. kudriavzevii* combinando la información obtenida por las técnicas de PCR-RFLPs, aCGH y citometría de flujo. Además, pudimos describir la presencia de cromosomas quiméricos y la contribución del parental *S. kudriavzevii* al genoma de los híbridos.

Observamos que la pérdida de genes de *S. kudriavzevii* en el genoma de los híbridos era generalizada. Sin embargo, encontramos un grupo de genes de *S. kudriavzevii* común a todos los híbridos. Este grupo de genes se analizaron usando GeneMAPP y YeastMine para obtener los términos de "Gene Ontology" (GO) enriquecidos en los genomas de los híbridos *S. cerevisiae* x *S. kudriavzevii*. El análisis de enriquecimiento de términos GOs reveló el potencial papel del subgenoma de *S. kudriavzevii* en la resistencia al frío.

3. Estudio del origen de los híbridos naturales S. cerevisiae *x* S. kudriavzevii

El objetivo de esta parte es descubrir qué tipo de parentales dio lugar a los híbridos naturales de *S. cerevisiae* x *S. kudriavzevii* y estimar el número mínimo de eventos de hibridación necesarios para generar el conjunto de estos híbridos naturales. La naturaleza vínica o no vínica del parental *S. cerevisiae* se estudió reconstruyendo redes filogenéticas con el programa Networks 4.6 usando el método de "Median-Joining" (MJ) para cuatro genes nucleares (*BRE5*, *CAT8*, *EGT2* and *GAL4*). Además, el conjunto de genes del parental *S. cerevisiae,* con menos copias en los híbridos naturales, comparado con otras cepas de *S. cerevisiae,* se obtuvieron de los análisis de aCGH. Combinando las redes de MJ y el grupo de genes con menos copias pudimos conocer qué tipo de parental *S. cerevisiae* dio origen a estos híbridos.

Las secuencias génicas de siete genes nucleares (*BRE5, CAT8, EGT2, GAL4, MET6, CYR1* y *CYC3*) sirvieron para la reconstrucción de los árboles filogenéticos, con el programa MEGA5, usando el método de

"Neighbor-Joining" (NJ) y las super redes filogenéticas para cada tipo de alelos (*S. cerevisiae* y *S. kudriavzevii*). El método de Z-closure, utilizado en la reconstrucción de super redes permite usar como dato de entrada una colección de árboles filogenéticos con diferente número de taxones, por tanto nos permite utilizar híbridos que hayan perdido alguno de los alelos parentales. Con la información obtenida de las super redes, SNPs y los árboles filogenéticos reconstruidos usando el método de Neighbor-Joining y Máxima Parsimonia pudimos definir el número de parentales de *S. cerevisiae* o *S. kudriavzevii* que podrían haber dado lugar a los híbridos naturales.

Las secuencias del gen *COX2* fueron utilizadas para reconstruir la red filogenética usando el método "Neighbor-net". Tanto las redes filogenéticas como las super redes fueron reconstruidas con el programa SplitsTree 4.

Los análisis filogenéticos, los niveles de ploidía y la información obtenida de los "caryoscopes" y la técnica de PCR-RFLPs fueron combinadas para obtener el número mínimo de eventos de hibridación necesarios para explicar la diversidad encontrada en los genomas de los actuales híbridos *S. cerevisiae* x *S. kudriavzevii*.

4. Reconstrucción de eventos de hibridación ancestrales entre las especies de *Saccharomyces*

La transferencia génica horizontal y las introgresiones son huellas que quedan en el genoma, indicativos, en algunos casos, de hibridaciones ancestrales. Elementos egoístas, como las "homing endonucleases", pueden expandirse rápidamente en la población, pudiendo ser utilizadas para identificar hibridaciones ancestrales. El objetivo de esta sección es analizar si las especies de *Saccharomyces* han hibridado en el pasado, para ello usaremos las secuencias de los genes *COX2*, *ORF1* y *COX3* que se encuentran en el genoma mitocondrial.

Herramientas informáticas, como los programas RDP3 y SplitsTree 4, revelaron un punto común de recombinación en el gen *COX2* para varias especies del género *Saccharomyces,* incluyendo algunos híbridos. El gen *ORF1* codifica para una "free-standing homing endonucleasa", anotada como un pseudogen o como no funcional. La secuencia de *ORF1* se encuentra dentro de la secuencia del gen *COX2,* en el extremo 3', pero en diferente pauta de lectura. El gen *COX3* se encuentra en una unidad de transcripción diferente a *COX2* y *ORF1*. Las cepas representativas de la secuenciación del gen *COX2*, que incluyen diferentes especies del género *Saccharomyces*, fueron utilizadas para secuenciar el gen *ORF1* y *COX3*. Se diseñaron cebadores específicos para amplificar y secuenciar el *ORF1* y *COX3* de las diferentes especies, usando las herramientas de IDT Scitools. Las secuencias fueron alineadas usando el algoritmo MUSCLE y refinadas manualmente en Jalview 4.9b2. Se detectaron secuencias repetidas en tándem, en los genes *ORF1* y *COX2*, con el programa Tandem Repeat Finder. Las anotaciones de los dominios se basaron en descripciones previas y utilizando la herramienta Conserved Domain en el NCBI. La conservación de los sitios aminoacídicos de *ORF1* se detectó utilizando el programa WebLogo 2.8.2. Los haplotipos de *COX2, ORF1* y *COX3* se clasificaron con el programa DnaSP v5. Y la detección de selección positiva en la secuencia del gen *ORF1* se hizo utilizando Datamonkey. Los árboles filogenéticos y las redes filogenéticas fueron realizadas mediante PhyML/MEGA5 y SplitsTree4, respectivamente.

Conclusiones

1. Identificación de los nuevos híbridos naturales S. cerevisiae *x* S. kudriavzevii

La distribución geográfica de los híbridos naturales *S. cerevisiae* x *S. kudriavzevii* está limitada a climas Oceánicos y Continentales, que se caracterizan por tener inviernos fríos y veranos cálidos y secos. En esta

tesis, se extendió el límite geográfico de los híbridos *S. cerevisiae* x *S. kudriavzevii,* aislados de vino, a zonas más sureñas del clima Oceánico Europeo. Por primera vez, se han identificado híbridos *S. cerevisiae* x *S. kudriavzevii* en ambientes no vínicos ni cerveceros. Estas nuevas fuentes de aislamiento corresponden a un suplemento dietético y a un aislado clínico.

2. Caracterización genómica de híbridos naturales S. cerevisiae *x* S. kudriavzevii

La estructura genómica de los híbridos naturales *S. cerevisiae* x *S. kudriavzevii* apunta a una gran diversidad genómica. El contenido de DNA variaba de 3.00C a 4.00C comparado con la cepa de referencia haploide *S. cerevisiae*. Todos los híbridos tienden a mantener al menos 1 copia de cada cromosoma del parental *S. cerevisiae* y a perder cromosomas completos o partes de cromosomas del parental *S. kudriavzevii*. El papel del subgenoma de *S. cerevisiae* en los híbridos parece ser el mantenimiento del poder fermentativo y la resistencia al etanol, y por parte del subgenoma de *S. kudriavzevii* la resistencia al frío.

La secuenciación del gen *COX2* muestra que la mayoría de los híbridos naturales *S. cerevisiae* x *S. kudriavzevii* han heredado el genoma mitocondrial de *S. kudriavzevii,* lo que genera una restricción evolutiva a la pérdida masiva de genes del parental *S. kudriavzevii*. Esto se apoya debido a que los pocos híbridos naturales *S. cerevisiae* x *S. kudriavzevii* que heredaron el mitocondrial del parental *S. cerevisiae* han sufrido una mayor pérdida de genes del parental *S. kudriavzevii*. Además, estos resultados indican que pueden existir incompatibilidades citonucleares que favorezcan la existencia de una barrera postzigótica en los híbridos del género *Saccharomyces*, y por ello una baja viabilidad de sus esporas.

El nivel de ploidía junto con la estructura genómica de los híbridos naturales *S. cerevisiae* x *S. kudriavzevii* indicaría que el mecanismo más

probable para la formación de híbridos es el “rare-mating”. En la mayoría de casos, este “rare-mating”, se produjo posiblemente entre una cepa diploide de *S. cerevisiae* y una cepa haploide de *S. kudriavzevii*. Sin embargo, en otros casos como PB7 el cruce se habría dado entre dos cepas diploides, generando una cepa tetraploide con 2 cromosomas de *S. cerevisiae* y dos de *S. kudriavzevii* para el conjunto de los 16 cromosomas. En el caso de la cepa AMH, su origen parece haber sido una doble hibridación, entre una cepa diploide de *S. cerevisiae* y una haploide *S. kudriavzevii* y el híbrido derivado de este cruce habría vuelto a hibridar con una cepa diploide de *S. cerevisiae*.

La aparición de cromosomas quiméricos parece ser el resultado del entrecruzamiento de los cromosomas homeólogos mediado por secuencias altamente recombinantes, como son las ARS, elementos Ty, elementos Y’, regiones de rRNA y regiones génicas muy conservadas. Esto activaría el mecanismo de reparación MMR que generaría los cromosomas quiméricos.

3. Estudio del origen de los híbridos naturales S. cerevisiae *x* S. kudriavzevii

La secuenciación de genes nucleares y mitocondriales, combinado con los datos de aCGH, PCR-RFLPs, citometría de flujo y análisis bioinformáticos reveló que al menos existen seis eventos de hibridación diferentes que han generado el conjunto de los doble híbridos naturales estudiados en esta tesis. Varias cepas parentales podrían haber dado origen a los híbridos naturales, tanto dobles como triples. En todo caso, estas cepas están relacionadas con el grupo de las *S. cerevisiae* vínicas y Europeas, y con *S. kudriavzevii* de Europa. En el caso de los híbridos cerveceros, la cepa *S. cerevisiae* parental podría ser un heterocigoto emparentado con las *S. cerevisiae* cerveceras. El lugar exacto del origen de los diferentes grupos de híbridos es difícil de asegurar debido a la

expansión de *S. cerevisiae* alrededor del mundo; sin embargo, los datos apuntan a un origen Europeo. Se identificaron seis grupos de híbridos según su origen. Las cepas cerveceras (excepto CECT11003 y CECT11004), junto con el aislado clínico MR25, parecen haberse originado del mismo evento de hibridación. Las cepas vínicas Suizas y dos cerveceras (CECT11003 y CECT11004) podrían haber evolucionado de la misma célula híbrida original. Otro grupo engloba a los híbridos Austriacos, Vin7 y SOY3. El suplemento dietético IF6 parece haberse formado de cepas parentales muy parecidas a las cerveceras y vínicas. Y en el caso de PB7, AMH y los triples híbridos (CBS 2834 y CID1) sus linajes parecen ser independientes.

4. Reconstrucción de eventos de hibridación ancestrales entre las especies de Saccharomyces

La presencia de un punto caliente de recombinación en el gen mitocondrial *COX2* y su cercanía al gen *ORF1* parece indicar que la proteína *Orf1p* podría estar implicada en la recombinación en esta región. El gen *ORF1* es un elemento egoísta, el cual está bajo evolución neutral, que podría haberse perdido varias veces en los linajes de las cepas del género *Saccharomyces*, y recuperado posteriormente, tras la fusión de mitocondrias, como consecuencia de un evento de hibridación.

La hibridación entre especies del género *Saccharomyces* parece ser un fenómeno bastante frecuente, complicando el concepto biológico actual de especie, en levaduras. Los datos apuntan a que la hibridación es un mecanismo adaptativo muy importante en la evolución de levaduras, tanto en el pasado como en la actualidad. Los híbridos generados, podrían estar mejor adaptados a la variación ambiental o a las condiciones nuevas de fermentación, que sus progenitores, desplazándolos exitosamente donde las condiciones son más extremas para ambos parentales.

GENERAL INTRODUCTION

1. ASCOMYCETE YEASTS AND EVOLUTION

1.1 Yeasts and ecology

Ascomycete yeasts (phylum Ascomycota: subphylum Saccharomycotina: class Saccharomycetes: order Saccharomycetales) comprise a monophyletic lineage with a single order of about 1500 known species (Kurtzman *et al.,* 2011). Pasteur was the first to put forward the notion that yeast are necessary components of the microbiota of fermenting wine or beer by effecting the conversion of sugar to ethanol, while. Hansen provided the first insights on the distribution of yeasts in their natural habitats, being recognized as the founder of yeasts systematic (Phaff *et al.,* 1978).

Present whole genome sequencing projects are involved in sequencing Ascomycete yeast genomes. About 40 different yeast species have been sequenced so far (figure 1) and special attention has been directed to the Saccharomycotina (or Hemiascomycetes) (Casaregola *et al.,* 2011).

Yeasts are found in association with plants, animals and their interfaces. The characteristics of yeast habitats are usually rich in simple organic carbon, liquid or very high moisture, acidic or occasionally alkaline, and nutritionally complex. Such conditions are found in plant tissue undergoing various forms of decays, as well as exudates of roots, leaves, or flowers. Moreover, some yeasts are adapted to conditions met in association with the body of certain animals, usually acting as intestinal commensals. Yeasts are continuously found in habitats depending of serendipitous situations due to scarce previous knowledge (Kurtzman & Fell 1998).

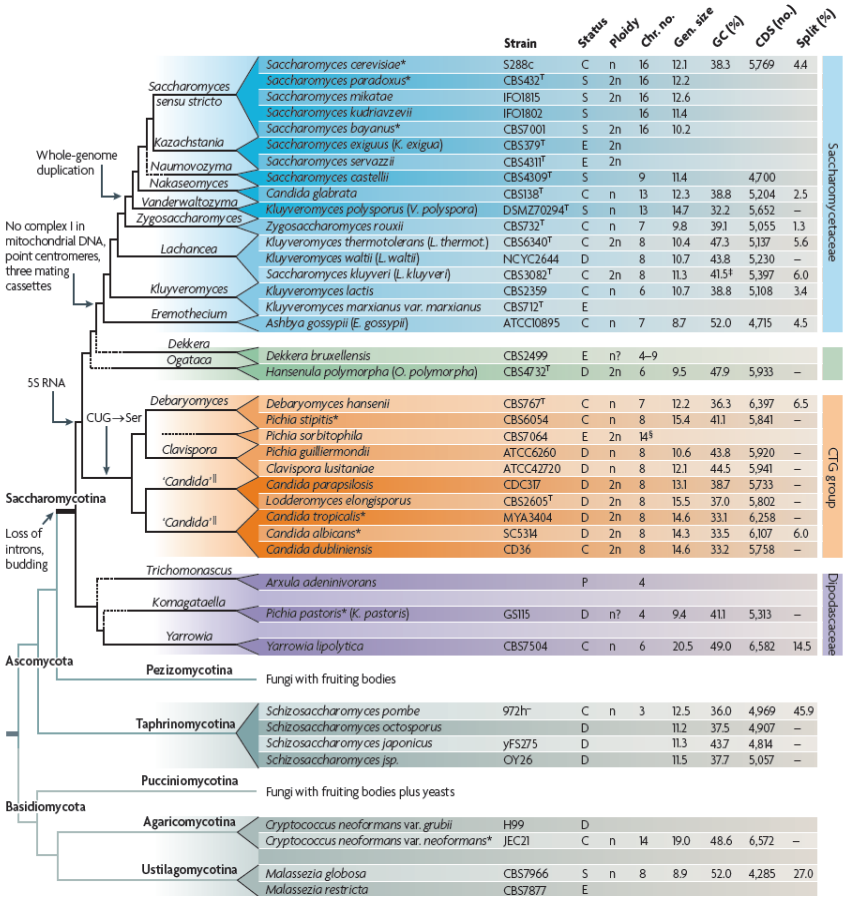


Figure 1. Tree topology of Ascomycetes yeasts with a complete sequenced genome. (adapted from Bernard Dujon 2010)).

A few yeast species are human pathogens, and fewer than 10 species are plant pathogens. Yeasts are responsible for important industrial and biotechnological processes, including baking, brewing, wine, bioethanol production and synthesis of recombinant proteins (Suh *et al.,* 2006).

1.2 The Saccharomyces *genus*

The *Saccharomyces* genus (previously called *Saccharomyces sensu stricto*) currently includes the species *Saccharomyces cerevisiae*, *Saccharomyces paradoxus, Saccharomyces bayanus* (Naumov 1987), *Saccharomyces cariocanus, Saccharomyces mikatae, Saccharomyces*

kudriavzevii (Naumov *et al.,* 2000) and *Saccharomyces arboricolus* (Naumov *et al.,* 2010).

The ecology of *Saccharomyces* species is diverse. Several species of this genus have been only found in natural environments, this is the case of *S. mikatae* (in partially decayed leaf), *S. kudriavzevii* (decayed leaf, soils and oaks) and *S. arboricolus* (oak trees); whereas *S. cerevisiae*, *S. paradoxus* and *S. bayanus* have been found associated to both natural and biotechnological environments.

Previous studies on *S. cerevisiae* indicated that this domesticated yeast, chiefly adapted to man-made fermentations (wine, beer, sake, baker) and normally absent in natural ecosystems, might have evolved from wild *S. paradoxus* (Naumov 1996; Naumov *et al.,* 1997; Martini 1993; Vaughan-Martini & Martini 1995; Ciani *et al.,* 2004).

However, recent studies suggest that *S. cerevisiae* is a natural species which has been 'domesticated' to perform superbly on man-made fermentations (Fay & Benavides 2005). Industrial *S. cerevisiae* strains are highly specialized organisms, which have evolved to growth in the different environments or ecological niches that have been provided by human activity. This specialization has been associated with some genome characteristics, such as diploidy genome with the presence of aneuploidies or polyploidies, high level of chromosome length polymorphism, homotallism, genome renewal and allopolyploid/hybrid genomes (Mortimer *et al.,* 1994; Querol *et al.,* 2003). In addition, wine strains have been characterized by the presence of a set of duplication and depletion genes referred as "commercial wine yeast signature" (Dunn *et al.,* 2005; Carreto *et al.,* 2008). Recently, genome sequencing of a *S. cerevisiae* wine yeast have revealed the presence of horizontal gene transfers that could be involved in adaptation to industrial environment (Novo *et al.,* 2009).

1.2.1 Population studies

Many *Saccharomyces* species have been sequenced with the purpose to unveil the population structure (Liti *et al.,* 2009; Schacherer *et al.,* 2009) (figure 2).

In *S. cerevisiae,* five “pure” populations have been described: North American, Sake, Malaysian, West African and Wine/European (Liti *et al.,* 2009). In *S. paradoxus* three populations depending on the geographic isolation were found: American (includes *S. cariocanus*), Far Eastern and European (Liti *et al.,* 2006, 2009). In the case of *S. kudriavzevii* two different populations have been described: European and Japanese (Sampaio & Gonçalves 2008; Hittinger *et al.,* 2010; Lopes *et al.,* 2010). *S. bayanus* includes two varieties: *uvarum* and *bayanus. S. bayanus* var. *bayanus* strains have been shown to be hybrids between *S. cerevisiae* and other unknown yeast close to *S. bayanus* var. *uvarum* (Rainieri *et al.,* 2006). Recently, a “pure” strain of *S. bayanus* has been described as the new species *S. eubayanus* (Libkind *et al.,* 2011). The *S. eubayanus* like-strain genome has been found in the former *S. pastorianus*, a hybrid between *S. cerevisiae* and *S. bayanus* var. *bayanus* (now *S. eubayanus*), which is found in lager-brewing fermentation (Libkind *et al.,* 2011). Future debate on definition of *S. eubayanus* and *S. uvarum* as different species (Pulvirenti *et al.,* 2000; Nguyen & Gaillardin 2005) or different varieties of the same specie *S. bayanus* (Naumov 2000) is expected.

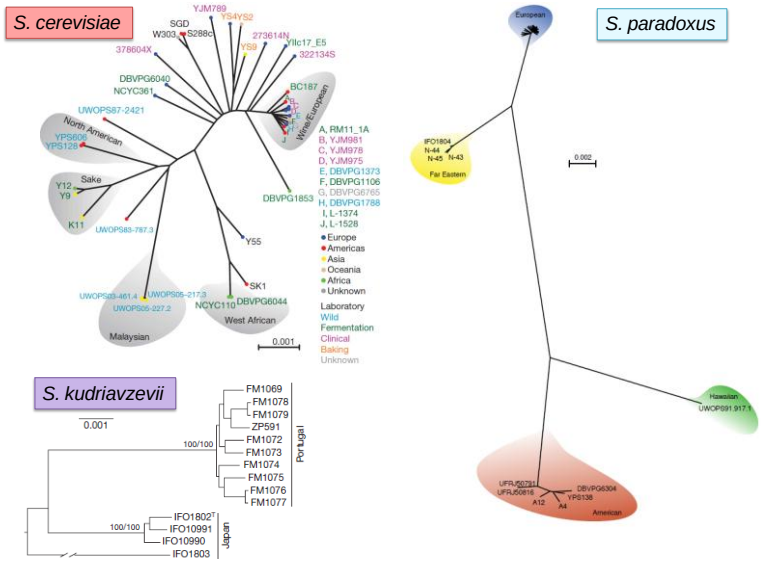


Figure 2. Population structure of some *Saccharomyces* species (adaptation from Liti *et al.,* 2009 and Hittinger *et al.,* 2010)

1.2.2 *Saccharomyces* hybrid yeasts

The development of molecular methods of yeast characterization has demonstrated that some wine and beer *Saccharomyces* strains have complex genomes composed by genetic elements from two or more species (Masneuf *et al.,* 1998; Groth *et al.,* 1999; de Barros Lopes *et al.,* 2002; Liti *et al.,* 2005; González *et al.,* 2006, 2008). These strains are widely known as interspecific hybrids (figure 3).

The best known industrial interspecies hybrid is the lager yeast *S. pastorianus*, originated from hybridization between *S. cerevisiae* and *S. eubayanus*-related yeast strain. *S. bayanus* strains have long been recognized as a cryotolerant yeast species (Sato *et al.,* 2002; Rainieri *et al.,* 2006), therefore the hybridization between *S. cerevisiae* and a cryotolerant *S. eubayanus* might be the result of selective pressures derived from brewing at low temperatures (Libkind *et al.,* 2011).

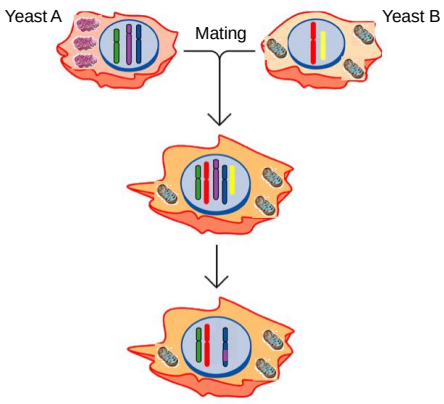


Figure 3. Interspecific hybrid formation.

Other natural hybrids are those originated from hybridization between *S. cerevisiae* and *S. kudriavzevii* (González *et al.,* 2006, 2008; Lopandic *et al.,* 2007). The role of *S. kudriavzevii* genome in hybrids is unclear, since the known strains of this species have been found in decaying leaves from Japan and oak trees from Portugal and Spain (Naumov *et al.,* 2000; Sampaio & Gonçalves 2008; Lopes *et al.,* 2010) but not associated to fermentative environments. Physiological evaluation of some *S. kudriavzevii* isolates has shown that this species is characterized by a higher cryotolerance and a lower ethanol tolerance than *S. cerevisiae* (Arroyo-López *et al.,* 2009; Salvadó *et al.,* 2011).

Albeit differences between *S. cerevisiae* x *S. eubayanus* and *S. cerevisiae* x *S. kudriavzevii* hybrids, the role of the *S. eubayanus* or *S. kudriavzevii* genomes in the hybrid seems to be similar, that is, maintenance of good fermentative performance at low temperatures (Belloch *et al.,* 2008).

Commercial yeasts *S. cerevisiae* x *S. kudriavzevii* hybrids have been identified in wine from Switzerland, Austrian and Germany, and lager beer from Belgium, England and New Zealand (González *et al.,* 2006, 2008; Bradbury *et al.,* 2006; Lopandic *et al.,* 2007) (figure 4). Genome diversity of

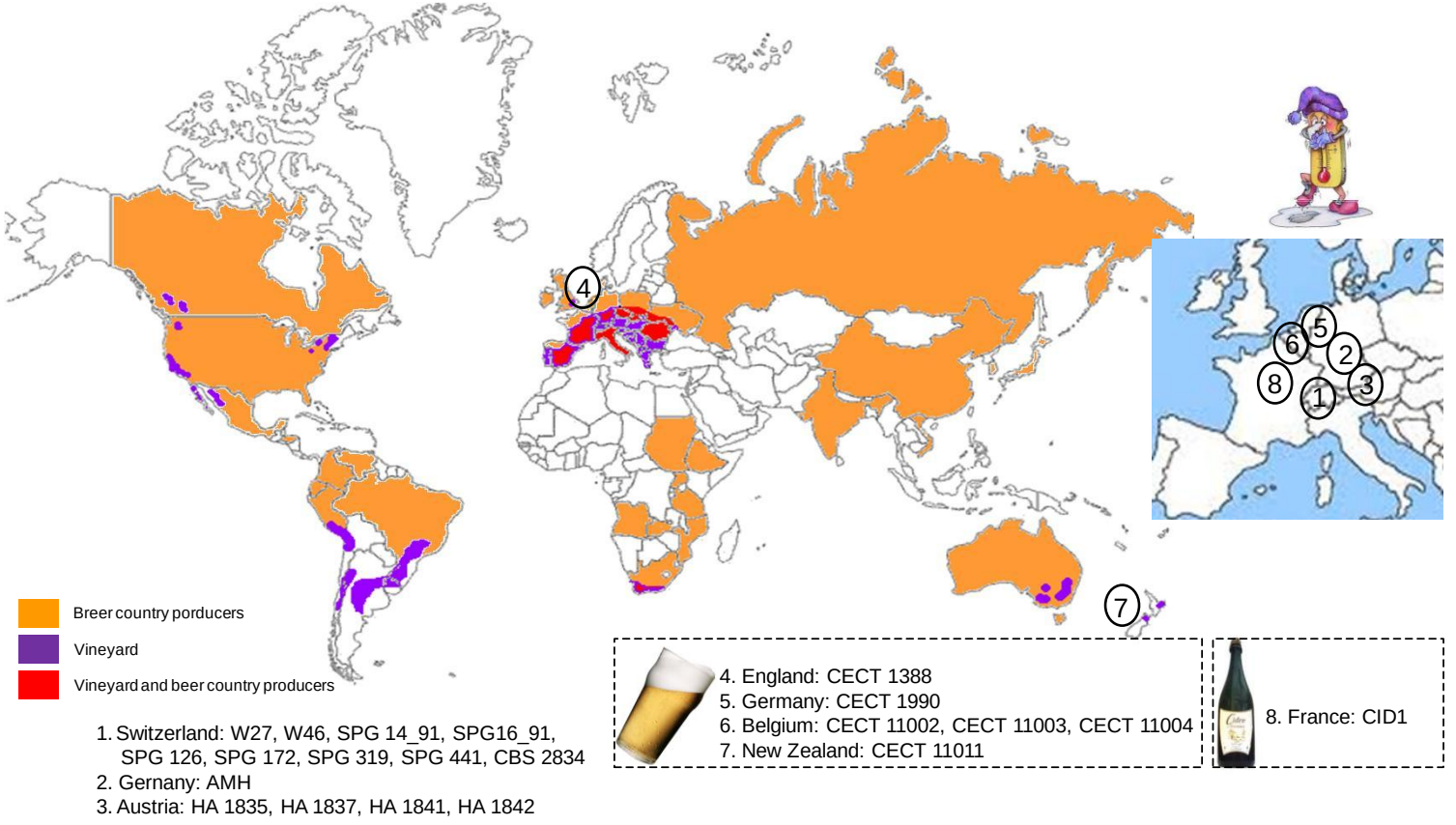


Figure 4. *S. cerevisiae* x *S. kudriavzevii* hybrid distribution. Orange and violet colors indicate brewing companies and vineyards distribution around the world, respectively. Red color indicates regions where vineyards and brewing companies are located. More detailed is done for European region and thermometer indicates low temperature where hybrids where isolated.

commercial *S. cerevisiae* x *S. kudriavzevii* hybrids has been analysed by Restriction Fragment Length Polymorphisms (RFLPs) of 35 genes distributed in the 16 chromosomes of these *Saccharomyces* hybrids (figure 5).

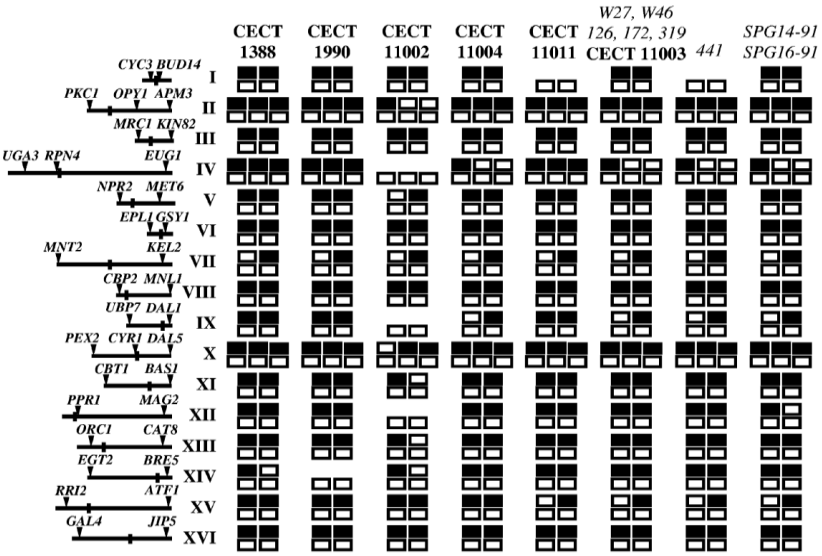


Figure 5. RFLPs data from *S. cerevisiae* x *S. kudriavzevii* hybrids. Each square corresponds to a copy of each gene region according to its chromosome location, indicated at the left. White and black squares represent alleles of *S. cerevisiae* and *S. kudriavzevii* origin, respectively. Brewing and wine hybrids are indicated in bold and italics, respectively. The presence or absence of alleles coming from each parent species was determined by restriction analysis of the 35 gene regions amplified by PCR with general primers (adapted from Gonzalez *et al.,* 2008).

Genome structure of a Swiss commercial *S. cerevisiae* x *S. kudriavzevii* hybrid was explored by array Comparative Genome Hybridization (aCGH), flow cytometry and Real Time qPCR (RT-qPCR) (Belloch *et al.,* 2009). The results of this study indicated that the genome of this hybrid is diploid with a trend to loss *S. kudriavzevii* genes (figure 6).

1.2.3 Yeast hybrids and the biological species concept

The definition of species is a central concept in biological sciences (Mallet 1995; Coyne & Orr 1998). There are different species concepts depending on which criteria are used (Mallet 2007).

The biological species concept (BSC) is based on patterns of breeding. Species are groups of interbreeding natural populations that are reproductively isolated from other such groups (Mayr 1970), but within which interbreeding and genetic recombination reduce the possibility of divergence.

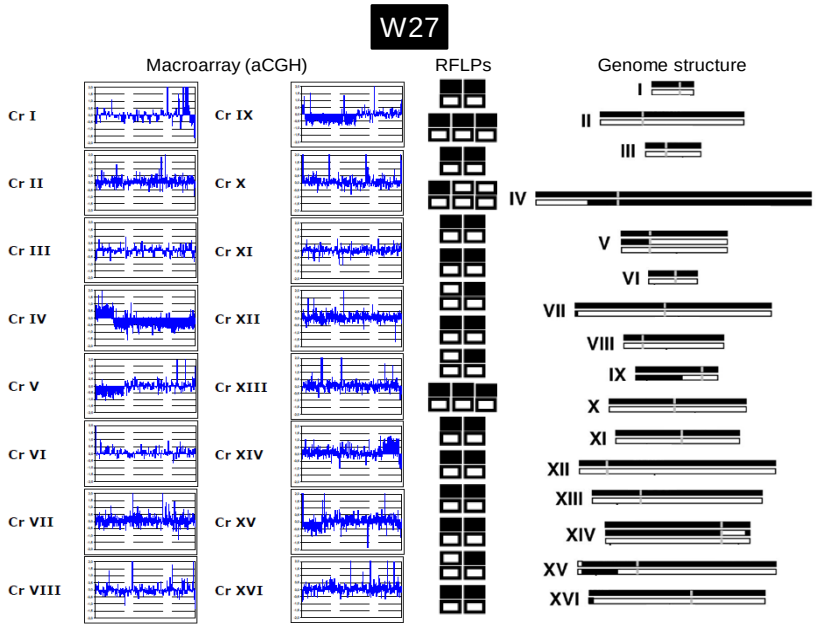


Figure 6. Genome structure of W27 *S. cerevisiae* x *S. kudriavzevii* hybrid inferred by the combination of macroarray data with RFLPs and RT-qPCR. In the genome structure black and white bars indicate *S. cerevisiae* and *S. kudriavzevii* subgenome, respectively (adapted from González (2006 and 2008) and Belloch *et al.,* (2009)).

Problems in the BSC are found due to asexual reproduction in many organisms including fungi (Taylor *et al.,* 2000), hybridization in plants (Rieseberg, 1997) and in yeast (de Barros Lopes *et al.,* 2002), and lateral

gene transfer between bacteria (Gogarten and Townsend 2005) and between yeast (Liti *et al.,* 2005).

Saccharomyces species have been classified according to the Biological and the Phylogenetic species (PSC) concepts being in agreement with both of them. The exception is the species *S. cariocanus* which is almost identical in genome sequence to the American *S. paradoxus* differing only by four translocations, therefore being the BSC not in accordance with the PSC.

The PSC (Cracraft, 1989) defines the species as an irreducible (basal) cluster of organisms, diagnosably distinct from other such clusters, and within which there is a parental pattern of ancestry and descent. However, a parental pattern of ancestry and descent is not clear when there are hybrids. It is necessary to define the genotypic cluster criterion (GCC), where separate species are recognized if there are several clusters separated by multilocus phenotypic or genotypic gaps, in a local area. A single species (the null hypothesis) is recognized if there is only a single cluster in the frequency distribution of multilocus phenotypes and genotypes. The genotypic gaps may be entirely vacant, or they may contain low frequencies of intermediate genotypes, or hybrids (Mallet 1995; Feder 1998).

1.2.4 Identification of hybrid yeasts

For hybrid identification, DNA-DNA hybridization (Vaughan Martini & Kurtzman 1985) and analysis of nuclear genes, such as *MET2* (Hansen & Kielland-Brandt 1994) unveiled the hybrid nature of *S. pastorianus* although failed in the identification of *S. monacensis.* In addition, gene sequencing of *MET2* and *ATP9*, and karyotypes were combined to infer the hybrid nature of CID1 and S6U (Masneuf *et al.,* 1998; Groth *et al.,* 1999). Analysis using multilocus markers, such as AFLPs o RAPDs have unveiled the hybrid nature of several *Saccharomyces* strains (de Barros Lopes *et al.,* 2002;

Fernández-Espinar *et al.,* 2003). *Saccharomyces* strains containing introgressed sequences or genomes containing two species sequences have been detected by PCR amplification and sequencing (Casaregola *et al.,* 2001), combined with microarrays analysis (Belloch *et al.,* 2009; Muller & McCusker 2009).

1.3 Genome evolution in yeasts

Yeasts offer unique advantages for evolutionary genomic studies among eukaryotic organisms. They are easily amenable to microbial genetic techniques, and the limited size and compactness of their genomes facilitate the characterization of naturally or artificially evolved populations using sequencing. Studies to infer evolutionary changes using the comparison of yeast genomes can be complemented by experimental analyses to elucidate the underlying molecular mechanisms. Yeasts were considered primitive unicellular eukaryotes, however they have repeatedly emerged from distinct phylogenetic lineages of 'modern' fungi (Kurtzman *et al.,* 2011).

The complete genome sequence of *S. cerevisiae* offers an unparalleled reference source for studying basic molecular mechanisms of eukaryotic cells, as more than 80% of its ~5780 protein-coding genes have been functionally characterized (Peña-Castillo & Hughes 2007).

Saccharomycetaceae yeasts are characterized by point centromeres (which are highly conserved) and triplicate mating-type cassettes that ensure the simultaneous presence of both mating-type alleles in haploid cells (with some exceptions). In this family a whole-genome duplication has been described (Kellis *et al.,* 2004), creating a subset of clades that have shorter chromosomes bearing the traces of the duplication followed by numerous gene deletions (Dujon 2010).

Genomes of the budding yeasts range in size from ~9 to 20 megabases (for the haploid set) and contain a limited number of protein-

coding genes (~4700-6500). They have few spliceosomal introns (~2-15% of split genes) and a variable number of tRNA genes (~160-510). The presence of a large number of paralogous gene copies is common to all yeast genome, which are highly diverged in their sequences and represent various types of ancestral duplications. Non-coding RNAs can be found within budding yeast genomes, in addition to limited numbers of mobile elements belonging to various families (mostly class I). iRNA machinery are generally absent, except in specific cases (Drinnenberg *et al.,* 2009). The presence of autonomous plasmids or viral elements is highly variable (Meinhardt *et al.,* 1990).

Genome comparisons of distinct yeast clades showed high differences. The orthologous proteins of *S. cerevisiae* and *S. paradoxus* are as different as those of humans and mouse (Dujon 2006). Classical Darwinian Theory proposes gradual evolutionary adaptations; however these high differences between yeast genomes can be only explained by repeated bottlenecks events that occurred during clonal divisions. This clonal mode of propagation, and the effects of bottlenecks, is important as it offers the possibility for non-optimized variants to survive and eventually colonize novel niches to which they may be better adapted. In addition it allows the involvement of different evolutionary mechanisms in reshaping the yeasts genomes (Dujon 2010) (figure 7).

A mechanism of loss of heterozigosity (LOH) has been described in *S. cerevisiae* (Butler *et al.,* 2009) and *Candida albicans* (Andersen *et al.,* 2008). *C. albicans* show a mosaic of heterozygous and homozygous regions in homologous chromosomes. LOH reduces heterozigosity in diploid cells or hybrids as has been shown in *S. cerevisiae* (Acuña *et al.,* 1994).

The genomes of wine strains of *S. cerevisiae* contain DNA fragments from different species, such as *S. paradoxus, S. kudriavzevii, S. uvarum*

and even the distantly related *Zygosaccharomyces bailii* (Liti *et al.,* 2006; Naumova *et al.,* 2005; Doniger *et al.,* 2008; Muller & McCusker 2009; Novo *et al.,* 2009). These foreign sequences called introgressions are the result of homologous recombinations after a hybridization event between two different yeast species. This is observed in *S. cerevisiae* and *S. paradoxus* where several introgressions have been described (Liti *et al.,* 2006; Wei *et al.,* 2007; Muller & McCusker 2009). This suggests that very recent introgressions have occurred. This phenomenon seems to be common in yeast genomes, although the importance of this mechanism in the domestication has not been determined (Dujon 2010).

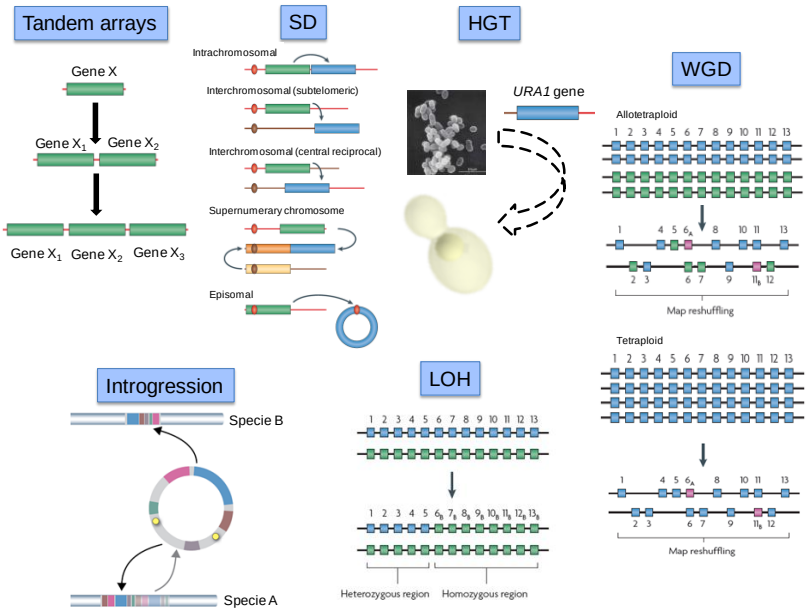


Figure 7. Evolutionary mechanisms involved in reshaping yeast genomes (adapted from Dujon (2010); Galeote *et al.,* (2011)).

The acquisition of genes from bacteria (Horizontal Gene Transfer or HGT) was considered to be rare in yeasts. However, some recent studies have reported the gaining of important functional innovations or reacquisition of bacterial genes (Gojković *et al.,* 2004). Furthermore, recent

analysis suggest that HGT is not rare in yeasts (Fitzpatrick 2011) and could be occurring between different species of yeast due to hybridization events. In this case, HGT and introgression could be considered as similar domestication mechanisms reshaping yeast genomes.

As postulated by Ohno 40 years ago (Ohno 1970), all genomes show numerous traces of gene duplications. In yeast, duplications can occur by different mechanisms: i) expansions of tandem gene arrays mediated by unequal homologous recombination (Fogel & Welch 1982; Despons *et al.,* 2010); ii) segmental duplication (SD) (Souciet *et al.,* 2009) mediated by dispersed repeated elements in genomes, such as remnants of Ty elements (class I retrotransposons) or microhomology/microsatellite-induced replication (MMIR) mechanism (Payen *et al.,* 2008); iii) single-gene duplications mediated by Ty retrotransposons (Schacherer *et al.,* 2004); iv) whole-genome duplications (WGD).

WGD hypothesis was confirmed by sequence comparisons of *S. cerevisiae* with other yeasts of Saccharomycetaceae family (Dietrich *et al.,* 2004; Kellis *et al.,* 2004). Genomes of the Saccharomycetaceae family have been the most extensively studied. 'Protoploid' Saccharomycetaceae have a large number of chromosomes (6-8) and 'duplicated' Saccharomycetaceae have twice as many (13-16). WGD could be studied to understand the evolutionary consequences of such duplications events. WGD has consequences for gene dosage and could affect the protein interaction networks. WGD could explain successive deletions of genes from the initial polyploidy stage creating phenotypically disadvantaged intermediates that could be maintained and evolved under several steps of bottlenecks (Presser *et al.,* 2008; Vinogradov & Anatskaya 2009). In *S. cerevisiae,* only ~550 duplicated pairs (ohnologues) have been retained (Byrne & Wolfe 2005), and similar or lower number of duplicates are observed for other yeasts coming from the same duplication event.

Although WGD is not a special feature of biotechnological strains, it provided new genes that played a direct role in the adaptation of *Saccharomyces* species toward highly efficient fermentation performance under anaerobic conditions (Piskur & Langkjaer 2004; Wolfe 2004). Hybridization has been not considered important in the evolution of yeast. However, WGD could be occurred due to a complete genome duplication (autopolyploidization) or by diploid mating (allopolyploidization), some authors supported the latter mechanism to be occur in the ancestor of post-WGD yeasts (Andalis *et al.,* 2004; Dunn & Sherlock 2008). This and recent descriptions support the hypothesis that hybridization is more important than previously has been recognized.

2. LIFE CYCLE OF YEASTS AND HYBRID FORMATION

Diploid yeasts reproduce asexually, frequently dividing by mitosis and budding off genetically identical cells, when they are grown in rich medium. But, when placed in medium lacking sufficient nitrogen to maintain mitosis, diploids can undergo meiosis producing tetrad of four haploid spores. Spores are dormant and resistant to many environmental conditions, but when returned to rich medium, they germinate into metabolically active haploid gametes of two mating types, *MATα* and *MATa*. Two gametes with different mating types can fuse together to produce a single diploid cell (Greig 2008). Sexual reproduction in yeast is facultative and not all diploids in a population will enter meiosis when deprived of the nutrients. In some situations, diploids, in starvation conditions, do not enter meiosis and can die or survive for many months (Fabrizio & Longo 2003). Meiosis and sporulation usually produce a tetrad of haploid spores, but in carbon source limitation during sporulation they can produce triads, dyads or monads (Taxis *et al.,* 2005).

In species with haploid mating types, e.g. *S. cerevisiae,* three principally different mating behaviors are possible: amphimixis, haplo-selfing and automoxis. Amphimixis is the mating of haploid cells derived from meiotic products of unrelated diploid cells. Haplo-selfing involves cells that derive from the same haploid cell, via mother-daughter mating upon mating type switching of one of the cells involved (Herskowitz 1988). Haplo-selfing is only possible when cells are able to change mating type, and leading to the formation of an entirely homozygous diploid cell. The population genetic implications have been characterized as 'renewal of the genome' (Mortimer 2000) because it efficiently selects for favorable combinations of alleles, and enables purify deleterious mutations. Automoxis, or intratetrad mating, is the mating of haploid cells originating from the same ascus (figure 8).

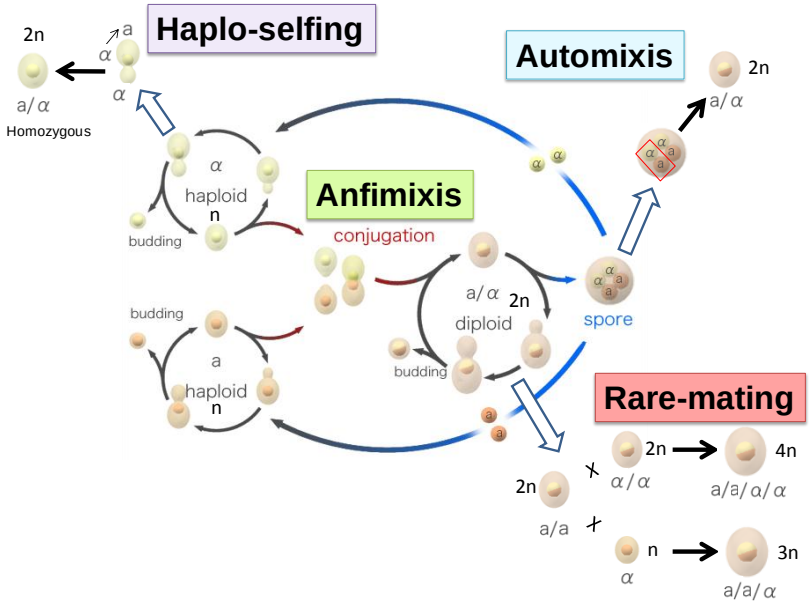


Figure 8. Life cycle of *Saccharomyces* with different types of mating.

Yeasts heterozygous for mating are generally lacking the ability to mate. Nevertheless, in rare circumstances these yeasts show rare-mating abilities due to conversion from heterozigosity to homozigosity for the

mating type alleles (Gunge & Nakatomi 1972). Rare-mating could have as consequence the formation of polyploid genomes.

Coexistence of *Saccharomyces* species in similar habitats (Sniegowski *et al.,* 2002) and the isolation of interspecific viable natural hybrids (de Barros Lopes *et al.,* 2002) indicates the absence of prezygotic isolation, despite the preference to mate with cells from the same species (Maclean & Greig 2008), and might have occurred in the wild. A postzygotic barrier may exist between *Saccharomyces* species due to the isolation of few viable spores (≤1%) in interspecies crossings (Naumov 1987). Apparently, gene order is not the only reason to postzygotic barrier, as seem to be between *S. paradoxus* and *S. cariocanus*, since many of the *Saccharomyces* species genomes are collinear (Fischer *et al.,* 2000; Kellis *et al.,* 2003; Scannell *et al.,* 2011).

The genome structure of several hybrids between *S. cerevisiae* x *S. bayanus* and *S. cerevisiae* x *S. kudriavzevii* have been described. In the first case two different groups of hybrids has been found. The group 1 indicates haploid x haploid hybridization and in the group 2, homozygous *S. cerevisiae* x haploid *S. bayanus* hybridization, (Dunn & Sherlock 2008; Nakao *et al.,* 2009). In the *S. cerevisiae* x *S. kudriavzevii* hybrids two haploids should have mated due to the diploid status of hybrids (Belloch *et al.,* 2009). However, these studies cannot elucidate which is the mechanism involved in the hybrids formation, being spore-spore, cell-spore or rare-mating equally probable.

In yeasts, mitochondrial DNA (mtDNA) is of uniparental inheritance, thus driving to a homoplasmic state. However, biparental inheritance of mitochondria (heteroplasmic state) has been documented in fungi of the genus *Neurospora* (Yang & Griffiths 1993). In *S. cerevisiae* the fusion of two mating yeast cells to form a diploid zygote is rapidly followed by the fusion of mitochondria to form a continuous mitochondrial reticulum where

mtDNA as well as other mitochondrial constituents derived from both parents are mixed (Berger & Yaffe 2000). A large body of data has confirmed that intraspecific mitochondrial recombination occurs readily in *S. cerevisiae* (Wilkie & Thomas 1973; Dujon *et al.,* 1974).

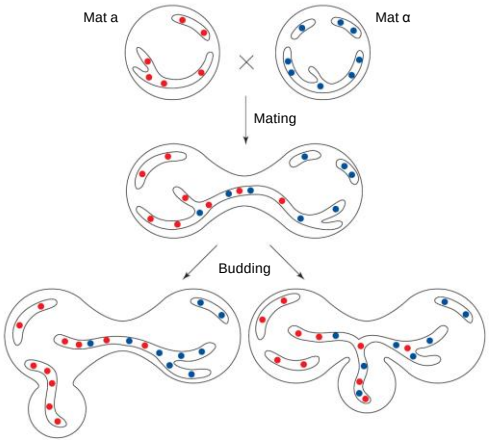


Figure 9. mtDNA inheritance during yeast mating and zygotic budding (adapted from Berger & Yaffe (2000)).

Genetic comparison of mtDNA from medial-bud and end-bud derived *S. cerevisiae* cells revealed that cells formed from medial buds generally inherited mtDNA from both parents (in form of recombinant mtDNA), whereas cells from end buds typically inherited mtDNA only from the proximal half of the zygote (Nunnari *et al.,* 1997) (fig. 9).

In all natural hybrids studied, *S. cerevisiae* x *S. bayanus* or *S. cerevisiae* x *S. kudriavzevii*, the mtDNA seems to be inherited from the non-*S. cerevisiae* parental (Gonzalez *et al.,* 2008; Rainieri *et al.,* 2008) (fig. 10). In a recent study, nucleo-mitochondrial incompatibilities within hybrids between *S. cerevisiae* and *S. bayanus* have been reported. In these hybrids, the *S. bayanus* nuclear gene *AEP2* is incompatible with the *S. cerevisiae* mitochondrial gene *OLI1* (Lee *et al.,* 2008), supporting the Dobzhansky-Muller mechanism of postzygotic barrier (Dobzhansky 1937).

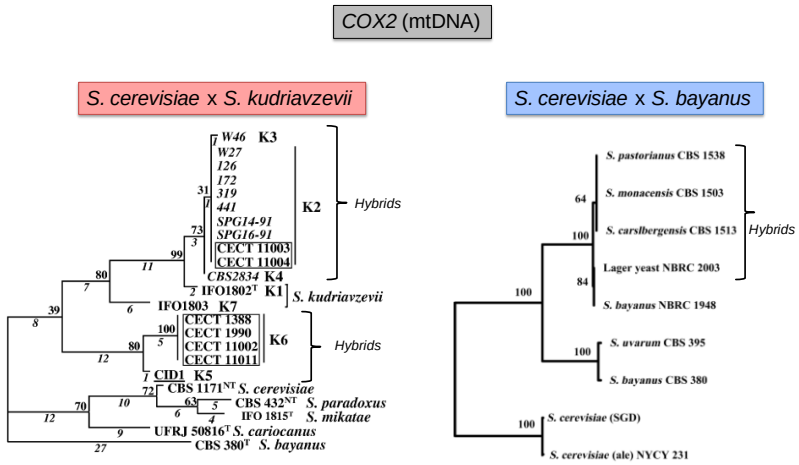


Figure 10. *COXII* phylogenetic tree (adapted from (Gonzalez *et al.,* 2008; Rainieri *et al.,* 2008).

Further cytonuclear incompatibilities have been reported between *MRS1* and *AIM22* genes in hybrids between *S. cerevisiae, S. paradoxus* and *S. bayanus* (Chou *et al.,* 2010). In those studies, low spore viability and high frequency of non-*S. cerevisiae* mtDNA inheritance were also observed. In the case of *S. cerevisiae* x *S. kudriavzevii* hybrids, the genetic incompatibilities that could generate the postzygotic barrier, involved in the low viability of spores, has not been explored.

3. IMPORTANCE OF THE HYBRIDS FOR BIOTECHNOLOGY

3.1 History of winemaking and brewing

Molecular evidence for the production of fermented beverages dates back to 7000 BC from the Neolithic village of Jiahu in China (McGovern *et al.,* 2004).

The earliest reports about grapevine domestication date from 7000-4000 BC from a region between the Black Sea and Iran. The first evidence of winemaking is associated with Mesopotamia 5400-5000 BC and from

there vineyards and wine production expanded around the world (Chambers & Pretorius 2010; Sicard & Legras 2011) (fig 11).

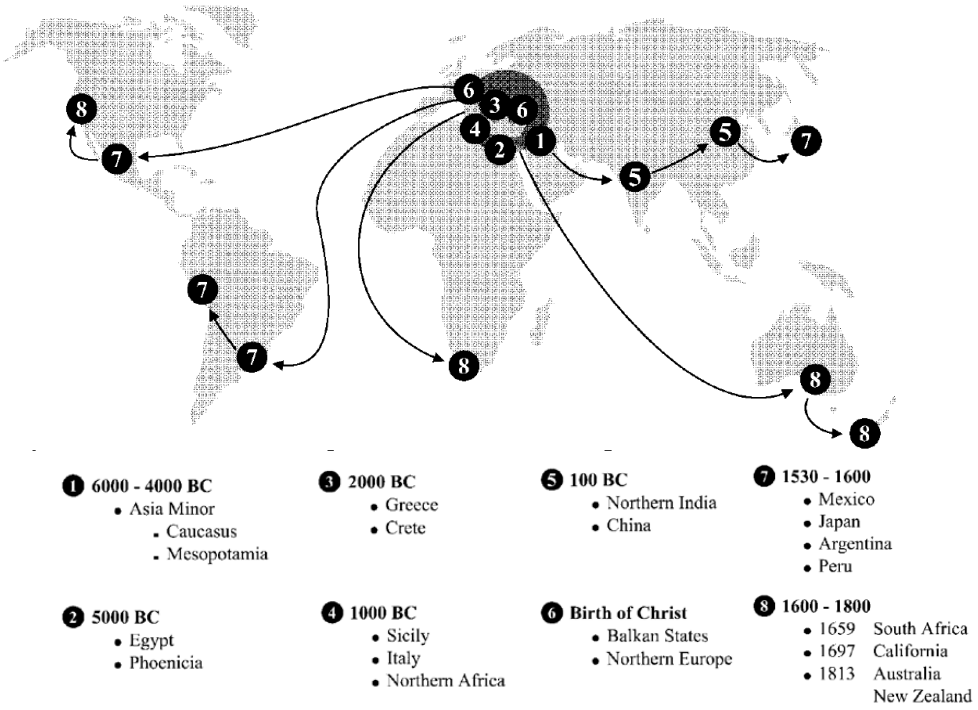


Figure 11. Vine and vineyard technology expansion around the world (adapted from Pretorius (2000)).

Beer elaboration was first mentioned in the Mesopotamian, recorded in a Sumerian tablet. Analysis of 1500-1300 BC old beer jars suggested that beer was made from cooked and uncooked malt (Samuel 1996). Later on brewing would diverge into two processes, ale and lager, differentiated by the fermentation temperature. Ale beer acquired from the Middle East by Germanic and Celtic tribes around the 1st century AD and lager appeared during the Middle Ages in Europe (Sicard & Legras 2011).

3.2 Alcoholic fermentation

Alcoholic fermentation (AF) is the anaerobic transformation of sugars into ethanol and carbon dioxide by yeasts (Zamora 2009). In addition to ethanol and carbon dioxide, several important flavor compounds are produced throughout wine and beer fermentation.

3.2.1 Winemaking process

Winemaking process, which starts in the vineyard, is an ancient art influenced by several factors such as viticulture practices, soil quality, and the cultivar of *Vitis vinifera*. All of these factors are of crucial importance for wine quality (Vivier & Pretorius 2002) (fig 12).

3.2.2 Brewing process

A brief diagram of brewing process is depicted in figure 12.

There are two types of brewing processes:

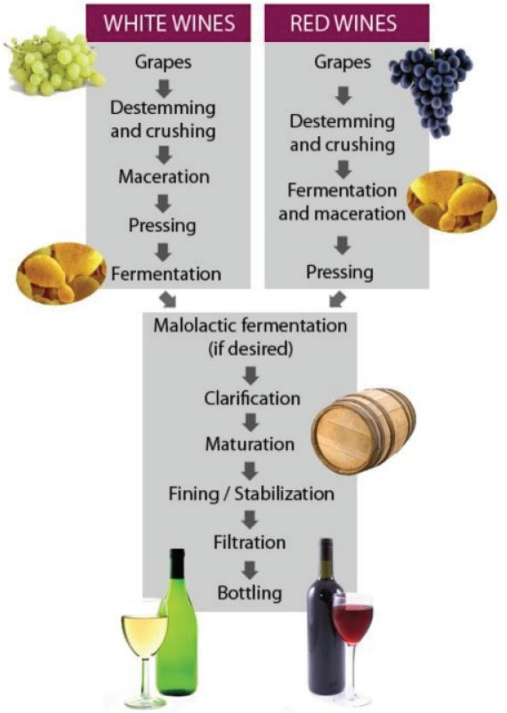
i) Ale fermentation, which usually uses a *S. cerevisiae,* top-cropping yeast at a temperature of 14 to 17ºC. The fermentation is fast and cooling is applied to maintain a constant temperature.

ii) Lager fermentation is done at lower temperature, typically 8 to 13ºC, using bottom-cropping *S. pastorianus* hybrid yeast. The traditional lagering process involves a primary fermentation using flocculant yeasts, which is followed by a secondary fermentation using nonflocculant yeasts at lower temperatures, around 8ºC. Finally, yeasts are removed and beer is put under an aging process where it is stabilized and matured at low temperature (Priest & Stewart 2006).

In certain types of traditional beers, as Belgian lambic and gueuze, fermentation is conducted by spontaneously growing yeasts. In these processes, fruits as cherry (‘Kriek’) or raspberry (‘Framboise’) are added to the beer. Late fermentation steps are conducted by *Brettanomyces* (Van Oevelen *et al.,* 1977).

OUTLINE OF

Winemaking process



Brewing process

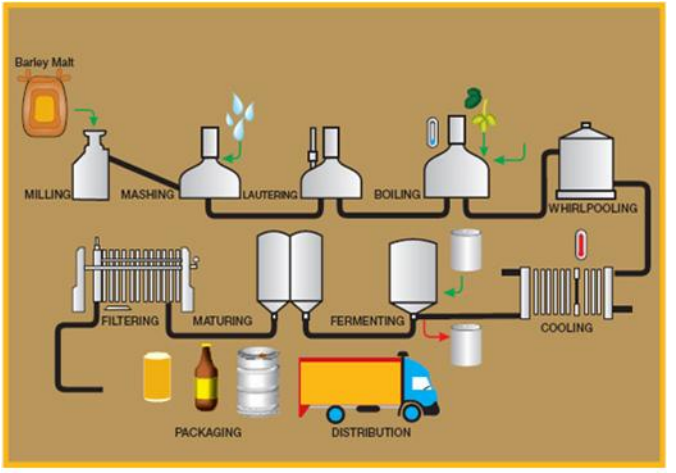


Figure 12. Winemaking and brewing processes (adapted from Pretorius (2000) and "http://www.monarch-beverage.com").

3.2.3 Stress conditions associated with alcoholic fermentation

During AF yeast cells are subjected to several stress conditions (Ivorra *et al.,* 1999; Carrasco *et al.,* 2001) (figure 13).

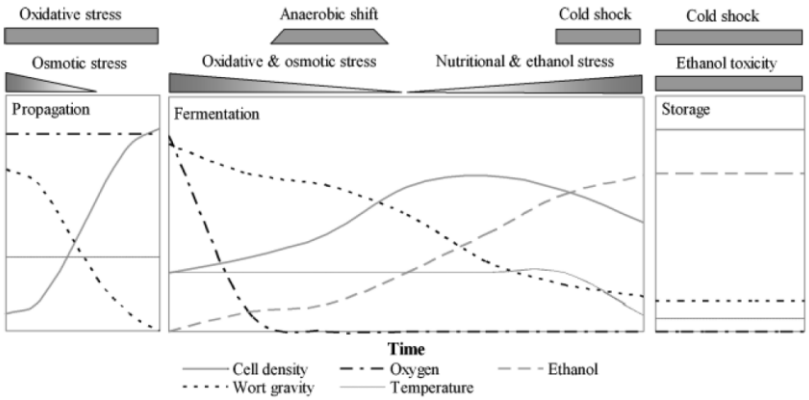


Figure 13. Schematic representation of the temporal and sequential nature of potential stress encountered by yeast during AF (adapted from (Gibson *et al.,* 2007)).

The most important are:

i) heat-shock stress, nowadays eliminated from the fermentation process by modern temperature control systems.

ii) oxidative stress, that occurs during biomass production and yeast drying (Erasmus *et al.,* 2003).

iii) hyperosmolarity, which is an ever present stress condition for wine yeasts. A typical grape must usually contains 125-260 g/L of an equimolar mixture of glucose and fructose selectively influencing the species and strains of yeasts responsible for the fermentation (Belloch *et al.,* 2008). In some situations, such as the production of dessert wines the sugar concentration may be as high as 500 g/L. Osmotic stress can also occur during yeast biomass production, downstream processing and drying (Ivorra *et al.,* 1999).

iv) nitrogen limiting conditions that can increase the H_2S levels. Excessive amount of ethanol inhibits the uptake of solutes (sugars and

amino acids) and also inhibits yeast growth rate, viability and fermentation capacity

v) ethanol concentration of the must that can adversely affect nitrogen uptake .

vi) temperature, which is one of the most important parameters influencing AF can affect the duration and rate of fermentation, but more significantly, the final quality of the wine (Torija *et al.,* 2003). Low temperature fermentations are becoming more frequent due to the consumer's demand of wines with more pronounced aromatic profiles. The high risk of stuck and sluggish fermentations is the biggest drawback to these fermentations, as low temperatures (10-15ºC) restrict yeast growth and lengthen fermentations.

Similarly, during beer production yeasts are exposed to fluctuations in oxygen concentration, osmotic potential, pH, ethanol concentration, nutrient availability and temperature (Briggs *et al.,* 2004).

3.3 Biotechnological yeasts

The *Saccharomyces* genus contains species that are industrially important. While *S. cerevisiae* is the predominant species responsible for AF (wine, ale-brewing, sake and different traditional fermented beverages), other species such as *S. uvarum* has been described as adapted to low-temperature fermentations during wine-making (Naumov *et al.,* 2000; Naumov *et al.,* 2002) and cider production (Naumov *et al.,* 2001; Coton *et al.,* 2006). *S. paradoxus* is being used for fermentation of Croatian wines (Redzepović *et al.,* 2002).

During AF, yeasts are gown in different stress conditions which might compel special genome features present in most of biotechnological *Saccharomyces* strains when compared with non-biotechnological strains (Querol & Bond 2009; Barrio *et al.,* 2006). Carreto *et al.,* 2008, using array Comparative Genome Hybridization (aCGH), described several traits of

deleted/duplicated genes that are common to wine commercial *S. cerevisiae* strains compared with natural and clinical isolates. Genome renewal has been observed in population of *S. cerevisiae* isolated from spontaneous wine fermentation, due to the high frequency of homozygous strains (Mortimer *et al.,* 1994).

One of the most interesting mechanisms observed in the adaptation of these yeasts to industrial process is the formation of interspecific hybrids.

3.4. Biotechnological hybrid strains

In the last years, an increasing demand to produce wines and beers with different organoleptic properties has occurred. Climatic change affects negatively wine quality by generating grape musts with higher fermentable sugar content and higher pH deriving in wines with high ethanol concentration (Jones *et al.,* 2005). Attempts to decrease alcohol content in wine start by harvesting grapes at an early maturation stage thus containing less fermentable sugars; however, these wines present an astringent character due to high tannins concentration that are not consumer desirable.

Nowadays, wine and beer companies are looking for new fermenting strains that generate low alcohol amount while increasing glycerol concentration thus solving the astringency problem (Arroyo-López *et al.,* 2009). Moreover, new yeast strains are also required to provide more aromatic wines. Wines produced at low temperatures are known to preserve better the aromatic varietal and fermentative compounds, therefore yeast strains able to perform superbly at low temperatures are also desirable (Torija *et al.,* 2003).

Previous studies had shown that *S. kudriavzevii* was a worthy glycerol producer while showing a good growth profile at low temperatures. However, fermentation using *S. kudriavzevii* often leads to stuck fermentations, as this yeast species is not able to perform till end of

fermentation due to its low ethanol resistance (Gonzalez, *et al.,* 2006; Belloch *et al.,* 2008; Arroyo-López *et al.,* 2010). Fermentation at low temperatures using hybrids between *S. cerevisiae* and *S. kudriavzevii* seems to be a good alternative, as natural hybrids are not considered genetically modified organisms (GMOs) (Gonzalez *et al.,* 2008; González *et al.,* 2006; Masneuf *et al.,* 1998; Sipiczki, 2008). Moreover, natural hybrid strains between *S. cerevisiae* x *S. kudriavzevii*, appear well adapted to stress conditions occurring during alcoholic fermentations while showing intermediate temperature and ethanol tolerances when compared with their parental *S. cerevisiae* and *S. kudriavzevii* strains (Arroyo-López *et al.,* 2009; Tronchoni *et al.,* 2009; Arroyo-López *et al.,* 2010). Unfortunately, how these natural hybrids between *S. cerevisiae* x *S. kudriavzevii* were originated in nature remains unclear (Gonzalez *et al.,* 2008) (fig. 14).

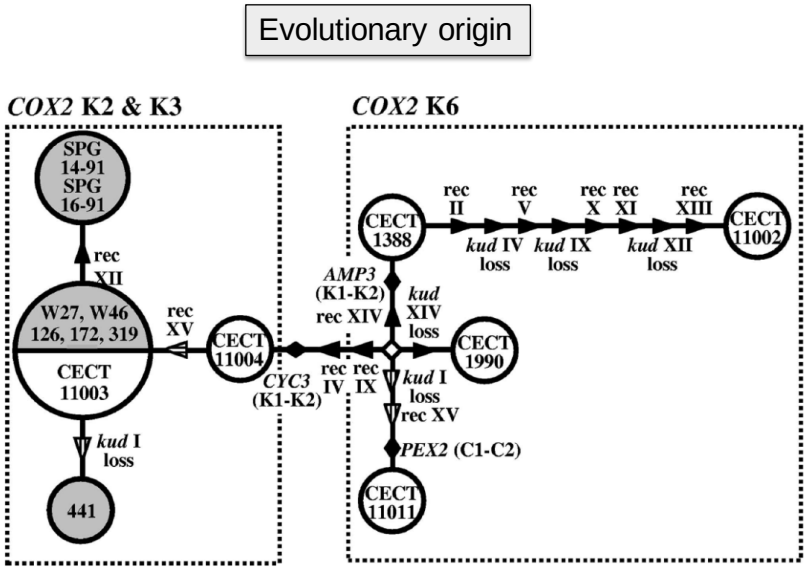


Figure 14. Evolutionary reconstruction of the *S. cerevisiae* x *S. kudriavzevii* origin (adapted from Gonzalez *et al.,* 2008).

BACKGROUND & JUSTIFICATION

In the last decade, natural *Saccharomyces* hybrids have been found responsible of diverse fermentative processes carried out at low temperatures (Masneuf *et al.,* 1998; Nguyen *et al.,* 2000; Rainieri *et al.,* 2006; González *et al.,* 2006; Gonzalez *et al.,* 2008; Sipiczki 2008). *Saccharomyces* hybrids, constituted by mating of the parental species *S. cerevisiae* and *S. bayanus* or *S. kudriavzevii*, or all of them, seem to contain a composite genome containing portions contributed by the different parental species (Dunn & Sherlock 2008; Rainieri *et al.,* 2006; Belloch *et al.,* 2009; Borneman *et al.,* 2011).

Extensive efforts have been done for the elucidation of the genomic and biotechnological particularities in case of the hybrids constituted by the species *S. cerevisiae* and *S. bayanus*, the species *S. pastorianus*, mostly conducting beer fermentation (Boulton & Quain 2001; Dunn & Sherlock 2008; Nakao *et al.,* 2009, Libkind 2011). However, recently described hybrids between *S. cerevisiae* and *S. kudriavzevii* have been barely investigated, albeit they have been found in wine and beer (Gonzalez *et al.,* 2006 and 2008).

Competitive genome hybridization of a wine *S. cerevisiae* x *S. kudriavzevii* hybrid revealed the chimeric nature of the nuclear genome whereas mitochondrial DNA appeared to be exclusively *S. kudriavzevii* (Belloch *et al.,* 2009).

Recent research evaluating the resistance of *S. cerevisiae* x *S. kudriavzevii* hybrids to diverse physical and chemical agents, considered as stress factors in winemaking, revealed that hybrids were better suited to grow at low temperature and high ethanol concentration than *S. cerevisiae* and *S. kudriavzevii* respectively (Belloch *et al.,* 2008; Arroyo-Lopez *et al.,* 2010). Physiological studies predicting winemaking performance of a wine hybrid *S. cerevisiae* x *S. kudriavzevii* compared to a commercial *S.*

cerevisiae wine strain and *S. kudriavzevii* revealed mixed traits coming from both parental species, where ethanol resistance seems to be inherited from the *S. cerevisiae* parental and the ability to ferment at low temperature and increased glycerol production would come from the *S. kudriavzevii* parental (Arroyo-Lopez *et al.,* 2009; Tronchoni *et al.,* 2009). Moreover, the study of production and release of aromas during winemaking revealed an increase in desirable chemical aromatic compounds at low temperature by *Saccharomyces* hybrids respect to cold temperature adapted *S. cerevisiae* strains (Gangl *et al.,* 2009; Gamero *et al.,* 2011a, b).

This doctoral thesis explores the genome diversity present in a varied selection of *S. cerevisiae* x *S. kudriavzevii* hybrids isolated from different sources and geographical locations. The different techniques applied have made possible the accurate identification of *S. cerevisiae* x *S. kudriavzevii* hybrids and the extensive characterization of their chimeric genomes procuring new hints on the origin of these hybrids and the most probable scenario for any ancestral hybridization events.

LIST OF TABLES, FIGURES & ADDITIONAL FILES

LIST OF TABLES, FIGURES & ADDITIONAL FILES

All these tables, figures and additional files are found in the CD-ROM and in my personal webpage:
(http://www.uv.es/~perisnav/Publications/DAVID_PERIS_Thesis.rar).
Descriptions of additional files are shown in pdf documents in each corresponding folder.

OBJECTIVES

1. Identification of new natural *Saccharomyces cerevisiae* x *Saccharomyces kudriavzevii* hybrids.
2. Genome characterization of natural *Saccharomyces cerevisiae* x *Saccharomyces kudriavzevii* hybrids.
3. Study of the origin of natural *Saccharomyces cerevisiae* x *Saccharomyces kudriavzevii* hybrids.
4. Reconstruction of ancestral hybridization events between *Saccharomyces* species.

<u>OBJECTIVE 1.</u> Identification of new natural *Saccharomyces cerevisiae* x *Saccharomyces kudriavzevii* hybrids.

Chapter 1

The molecular characterization of new types of *S. cerevisiae* x *S. kudriavzevii* hybrid yeasts unveils a high genetic diversity.

David Peris[1], Carmela Belloch[2], Ksenija Lopandić[3], José Manuel Álvarez-Pérez[4], Amparo Querol[2] and Eladio Barrio[1,*]

[1]Institute Cavanilles of Biodiversity and Evolutionary Biology, University of Valencia, Valencia, Spain.

[2]Department of Biotecnology, Institute of Agrochemistry and Food Technology (CSIC), Valencia, Spain.

[3]Austrian Center of Biological Resources and Applied Mycology, Institute of Applied Microbiology, University of Natural Resources and Applied Life Sciences, Muthgasse 18, 1190 Vienna, Austria.

[4]Vine and Wine Research Institute, University of León, 24400-Ponferrada, Spain

***Corresponding author:**

Dr. Eladio Barrio
Institut 'Cavanilles' de Biodiversitat i Biologia Evolutiva
Universitat de València
P.O. Box 22085
E- 46071 València
SPAIN
Tel.: +34 963 543 667
Fax: +34 963 543 670
E-mail: eladio.barrio@uv.es

2012 - Yeast 29(2): 81-91

Abstract

New double and triple hybrid *Saccharomyces* yeasts were characterized by using PCR-restriction fragment length polymorphism of 35 nuclear genes, located at different chromosome arms, and the sequencing of one nuclear and one mitochondrial genes. Most of these new hybrids were originally isolated from fermentations, however, two of them correspond to clinical and dietary supplement isolates. This is the first time that the presence of double hybrids *S. cerevisiae* x *S. kudriavzevii* in non-fermentative substrates is reported and investigated.

The phylogenetic analysis of the *MET6* nuclear gene confirmed the double or triple parental origin of the new hybrids. The restriction analysis of gene regions in these hybrids revealed a high diversity of genome types. From these molecular characterizations, a reduction of the *S. kudriavzevii* fraction of the hybrid genomes is observed in most hybrids.

Mitochondrial inheritance in hybrids was deduced from the analysis of the mitochondrial *COX2* gene sequences, which showed that most hybrids received the mitochondrial genome from the *S. kudriavzevii* parent. However, two strains inherited a *S. cerevisiae COX2*, being the first report of *S. cerevisiae* x *S. kudriavzevii* hybrids with *S. cerevisiae* mitochondrial genomes. These two strains are those showing a higher *S. kudriavzevii* nuclear genome reduction, especially in the wine hybrid AMH. This may be due to the release of selective pressures acting on the other hybrids to maintain *kudriavzevii* mitochondria-interacting genes.

Keywords: *Saccharomyces* hybrids*, S. cerevisiae, S. kudriavzevii*, wine, dietary, clinical yeasts.

1. Introduction

The genus *Saccharomyces* consists of eight species, three of them associated with industrial fermentation processes (*S. bayanus*, *S. cerevisiae*, and *S*. *pastorianus*), and five isolated from natural habitats (*S. arboricolus*, *S. cariocanus*, *S. kudriavzevii*, *S. mikatae* and *S. paradoxus*) (Kurtzman & Robnett , 2003; Wang & Bai , 2008). *S. cerevisiae*, the predominant species responsible for the alcohol fermentation, has been found associated to diverse fermentation processes including baking, brewing, distilling, wine making, cider production, etc. and also in different traditional fermented beverages and foods around the world. The species *S. bayanus* includes two recognized varieties, *bayanus* and *uvarum* (Vaughan-Martini & Martini , 2011). *S. bayanus* var. *uvarum* is present in wine and cider fermentations from cold regions of Europe (as examples see (Demuyter *et al.,* 2004; Naumov *et al.,* 2001). The *S. pastorianus* taxon includes hybrid strains between *S. bayanus* and *S. cerevisiae*, which are responsible for the production of lager beer (Kodama *et al.,* 2005). The rest of the species are only associated with natural habitats, with the exception of some *S. paradoxus* strains isolated from Croatian vineyards (Redzepovic *et al.,* 2002), that show a good winemaking performance (Orlic *et al.,* 2010).

During their evolution, yeasts have suffered diverse selective processes to become adapted to the fermentation conditions (Querol *et al.,* 2003). Diverse molecular mechanisms were involved in the generation of the evolutionary novelties that allowed the adaptation of yeasts to the fermentation processes (for review, see (Barrio *et al.,* 2006). In the case of the genus *Saccharomyces*, one of the most interesting mechanisms involved in their adaptation to industrial processes, is the generation of interspecific hybrids (Querol & Bond , 2009). Hybrids between *S. cerevisiae* and *S. bayanus* were already identified several decades ago (for review, see (Kodama *et al.,* 2005). In the last years, a new type of hybrids, between

S. cerevisiae x *S. kudriavzevii*, have been found both in winemaking and brewing (Bradbury *et al.,* 2006; Gonzalez *et al.,* 2006; Lopandic *et al.,* 2007).

In the present study we characterize the genome composition of new *S. cerevisiae* x *S. kudriavzevii* hybrids. These new hybrids include two strains isolated from wine regions located in the southernmost limits of the Oceanic and Continental Europe, and two hybrids isolated for the first time from non-fermentative sources, such as a human respiratory tract isolate (de Llanos *et al.,* 2004) and a strain employed as dietary supplement. Other hybrids, molecularly characterized for the first time in this study, are some commercial wine strains described as such by (Bradbury *et al.,* 2006) and some of the Austrian wine hybrids (Lopandic *et al.,* 2007), as well as two triple hybrids *S. bayanus* x *S. cerevisiae* x *S. kudriavzevii* CID1 (Groth *et al.,* 1999) and CBS2834 (Gonzalez *et al.,* 2006). The genetic characterization was performed by restriction analysis of 35 nuclear genes located in different chromosomes, and by sequencing the nuclear gene *MET6* and the mitochondrial *COX2* genes. Accordingly, these new hybrids were compared to those characterized in our previous study (González *et al.,* 2008).

2. Materials and methods

2.1 Yeast strains and culture media

The natural yeast hybrids *S. cerevisiae* x *S. kudriavzevii* used in this study were originally isolated from different sources and locations as described in Table 1. Yeast strains were grown at 28ºC in GPY medium (2% glucose, 0.5% peptone, 0.5% yeast extract).

Table 1. List of strains used in this study. Double hybrids correspond to *S. cerevisiae* x *S. kudriavzevii* hybrids and triple hybrids to *S. bayanus* x *S. cerevisiae* x *S. kudriavzevii* hybrids. Accession numbers of new gene sequences are indicated.

Strain type	Strain reference	Isolation source	*COX2*	*MET6*-C	*MET6*-K
Double hybrids	AMH	Commercial strain, Pinot noir wine, Assmanshausen, Germany	HQ414035	HQ414054	
	HA1835	Weißer Burgunder (Pinot blanc) grapes, Perchtoldsdorf, Austria	HQ414039	HQ414049	HQ414059
	HA1837	Weißer Burgunder grapes, Perchtoldsdorf, Austria	HQ414040	HQ414050	HQ414060
	HA1841	Weißer Burgunder grapes, Perchtoldsdorf, Austria	HQ414041	HQ414051	
	HA1842	Weißer Burgunder grapes, Perchtoldsdorf, Austria	HQ414042	HQ414052	HQ414061
	IF6	Brewer's yeast dietary supplement, Barcelona, Spain	HQ414034	HQ414057	
	MR25	Human respiratory tract isolate, Barcelona, Spain	HQ414033	HQ414058	HQ414065
	PB7	Pietro Picudo wine, Los Oteros Winery, León, Spain	HQ414036	HQ414056	HQ414064
	SOY3	Graševina (Welschriesling) must fermentation, Daruvar, Croatia	HQ414032	HQ414055	HQ414063
	VIN7	Commercial strain of unknown origin, Anchor, South Africa	HQ414031	HQ414053	HQ414062
Triple hybrids	CBS 2834	Wine, Wädenswil, Switzerland			
	CID1	Home-made cider, Brittany, France			
S. kudriavzevii	ZP542	Oak bark, Adagoi, Portugal	HQ414038		
	ZP591	Oak bark, Castelo de Vide, Portugal	HQ414037		

Double hybrids correspond to *S. cerevisiae* x *S. kudriavzevii* hybrids and triple hybrids to *S. bayanus* x *S. cerevisiae* x *S. kudriavzevii.* Accession nos of new gene sequences are indicated.

2.2 PCR amplification and restriction analysis of 35 nuclear gene regions

Characterization of the hybrids was performed by PCR amplification and restriction of 35 gene regions located in different chromosome arms (Fig. 2). DNA was extracted following the procedure described by Querol *et al.,* (1992). Amplification and digestion of the nuclear genes was performed by using the methodology described in González *et al.,* (2008) except for the subtelomeric *MNT2* gene, that failed to amplify the *S. kudriavzevii* gene and, hence, it was replaced by *GCN1*. Primers used for amplification of *GCN1* gene were *GCN1*-5 (GGTTTRGTKAAAGGTTAYGG) and *GCN1*-3' (CACCAGCYAAAATRGTTGG) and PCR conditions were as in González *et al.,* (2008), but using an annealing temperature of 55.5 ºC.

2.3 Amplification, sequencing and phylogenetic analysis of COX2 *and* MET6 *genes*

The genes *COX2* and *MET6* were amplified by PCR using the primers and conditions described in Belloch *et al.,* (Belloch *et al.,* 2000) and González *et al.,* (2006), respectively. PCR products were cleaned with the Perfectprep Gel Cleanup kit (Eppendorf, Hamburg, Germany) and both strands of the DNA were directly sequenced using the BigDyeTM Terminator V3.0 Cycle Sequencing Kit (Applied Biosystems, Warrington, UK), following the manufacturer's instructions, in an Applied Biosystems automatic DNA sequencer Model ABI 3730l (Life Technologies Coorporation, Carlsbad, California).

COX2 and *MET6* sequences obtained for the present study are listed in Table 1 with their accession numbers. Other sequences from hybrids were retrieved from sequence databases (accession numbers for *MET6* sequences AJ973280-AJ973295 and AJ973305-AJ973322; and for *COX2* sequences AJ938037-AJ93844, AJ938047, AJ938048 and AJ966727-AJ966733).

Finally, *MET6* sequences from reference or type strains of *S. bayanus* var. *uvarum* (MCYC 623), *S. cerevisiae* (S288C), *S. kudriavzevii* (IFO 1802^T), *S. mikatae* (IFO 1815^T) and *S. paradoxus* (CECT 1939^{NT}) were retrieved from the fungal alignment viewer of the *Saccharomyces* Genome Database (http://db.yeastgenome.org/cgi-bin/FUNGI/showAlign).

Each set of homologous sequences was aligned in MEGA 4 (Tamura *et al.,* 2007). The sequence evolution model that fits our sequence data best was optimized using the corrected Akaike Information Criterion (AICc) with a BioNJ tree as the initial tree, implemented in jModelTest program (Posada , 2008). The best fitting model of evolution for *MET6* sequences was TIM1 model (Posada , 2003) with a gamma distribution (G) of substitution rates with a shape parameter α= 0.35; and for *COX2* gene sequences the TVM model (Posada , 2003) with a gamma distribution (G) of substitution rates with a shape parameter α= 0.123 and 46.2% of invariable sites (I). The parameters of each model, estimated in the previous analysis, were used to obtain the best trees under optimality criterion of maximum-likelihood (ML). Tree reliability was assessed using non-parametric bootstrap re-sampling of 1000 pseudo-replicates. Phylogenetic analyses were performed using PhyML 3.0 program (Guindon *et al.,* 2010).

In the case of COX2 sequences, due to evidences of recombination obtained from sequence comparisons, a Neighbor-net network analysis was also performed with SPLITSTREE4 program (Huson & Bryant , 2006).

3. Results

3.1 Analysis of the hybrid nature of the strains by the phylogenetic analysis of MET6 *gene sequences.*

To confirm the hybrid nature of the strains under study and their genealogical relationships, we performed phylogenetic analyses of partial

sequences of a nuclear (*MET6*) and a mitochondrial (*COX2*) genes, because such sequences are also available for other hybrids (González *et al.,* 2006, 2008).

Three different *MET6* sequence types were found in hybrids that correspond to those of the reference strains of the parental species *S. bayanus*, *S. cerevisiae* and *S. kudriavzevii* (Fig. 1). Thus, the average number of nucleotide substitutions among *S. cerevisiae* alleles is 0.97 ± 0.88 (from 0 to 3 differences), among *S. kudriavzevii* alleles is 0.23 ± 0.59 (from 0 to 3) and among *S. bayanus* var. *uvarum* is 0 ± 0. In contrast, average numbers of All double and triple hybrids included in the analysis contain two or three *MET6* alleles coming from their parental species (B, C and K), except double hybrids HA1841, IF6 and AMH that lost the *S. kudriavzevii MET6* allele. These results confirm the hybrid nature of the new strains.

3.2 Nuclear genome characterization of Saccharomyces *hybrids*

The restriction patterns of the 35 genes for the differentiation of the *S. cerevisiae* and *S. kudriavzevii* alleles were described in González *et al.,* (2008) with the exception of *MNT2* which was replaced in the present study by *GCN1*. The restriction analysis of the *GCN1* gene region yielded the following fragments: *Hae*III, *S. cerevisiae* 462 + 302 + 144 + 114 bp, and *S. kudriavzevii* 450 + 366 + 206 bp; *Msp*I, *S. cerevisiae* 514 + 508 bp, and *S. kudriavzevii* 1022 bp; and *Cfo*I, *S. cerevisiae* 766 + 256 bp, and *S. kudriavzevii* 634 + 388 bp. Hybrids characterized in a previous study González *et al.,* (2008) were also assayed for this gene, resulting in the presence of both parental copies in all of them.

The PCR-RFLP patterns of the 10 newly characterized *S. cerevisiae* x *S. kudriavzevi* hybrids, and the 2 triple hybrids are depicted in Figures 2

and 3, respectively. The specific alleles present in each hybrid strain are given in the Table S1 and the new restriction patterns in Table S2.

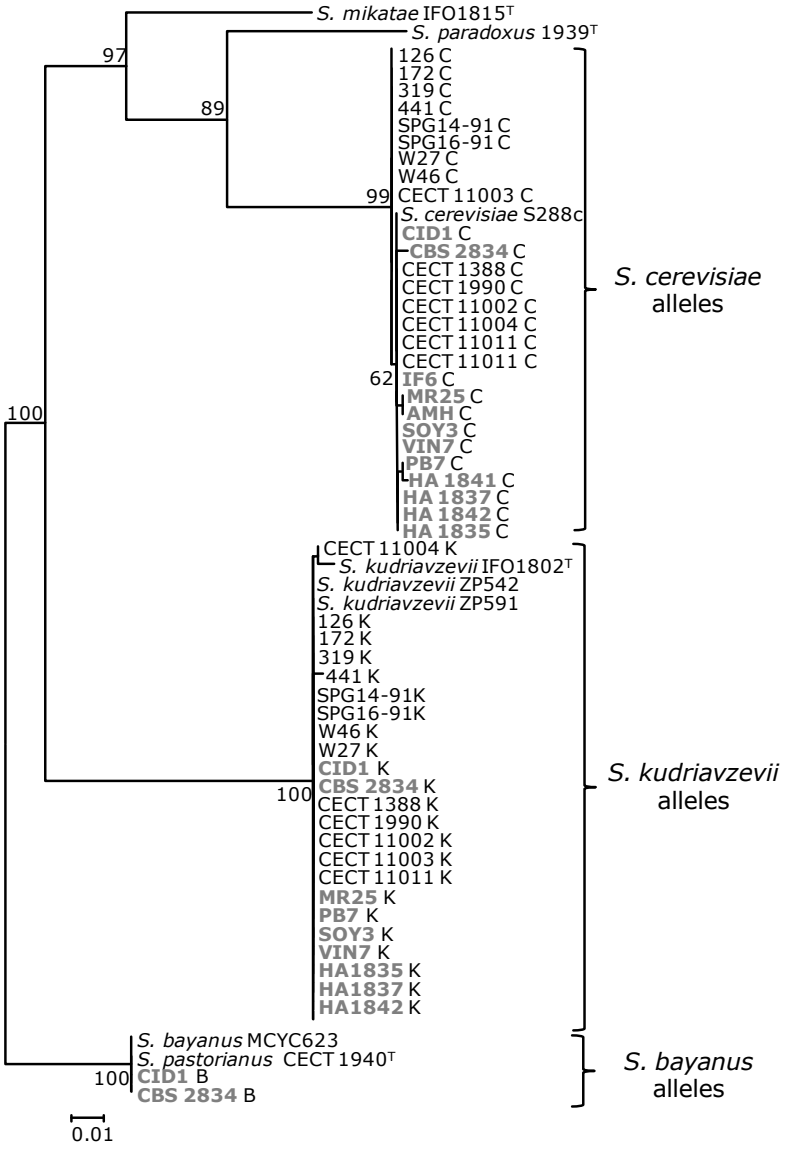


Figure 1. Phylogenetic tree obtained with partial sequences of the nuclear *MET6* gene from hybrid strains and reference strains of *Saccharomyces*. The new hybrids are indicated in bold gray characters. Hybrid strains contain one, two or three different *MET6* alleles named C (*S. cerevisiae*), B (*S. bayanus* var. *uvarum*) or K (*S. kudriavzevii*) according to the closest parental relative. Numbers at the nodes correspond to bootstrap values based on 1000 pseudo-replicates. The scale is given in nucleotide substitutions per site.

Since the *S. cerevisiae* and *S. kudriavzevii* genomes are colineal (Kellis *et al.,* 2003), the locations of the gene regions under analysis were chosen to obtain information about the presence of possible chromosomal rearrangements in the hybrid genomes, as described in other hybrids (Gonzalez *et al.,* 2008; Belloch *et al.,* 2009). This way, the absence in the hybrids of *S. kudriavzevii* alleles for genes located in the same chromosome likely resulted from the loss of the whole chromosome. However, the loss of one gene located in a chromosome but not the other genes of the same chromosome can be postulated as a result of recombination between homeologous chromosomes, as demonstrated for some hybrids (Belloch *et al.,* 2009). This resulted in the replacement of the missing segment by the homologous segment from the other chromosome of different parental origin (see Figures 2 and 3).

This way, chromosomal rearrangements can be postulated as occurred in chromosomes IV (AMH), V (IF6), VII (AMH, VIN7, IF6 and MR25), IX (IF6, MR25), X (IF6, MR25), XI (PB7 and IF6), XIII (IF6, MR25), XIV (MR25), XV (AMH) and XVI (IF6). In four wine hybrids (SOY3, from Croatia, and HA 1835, HA 1837 and HA 1842 from Austria) no rearrangement can be deduced because they contain both parental alleles for all genes.

In general, the *S. cerevisiae* genome fraction is maintained in all these double hybrids whereas a progressive loss of the *S. kudriavzevii* genes is observed. This reduction is more evident in the case of hybrid AMH, which has lost most of the *S. kudriavzevii* chromosomes.

In the case of the triple hybrids (Fig. 3), the typical restriction pattern of *S. bayanus var. uvarum* was found in addition to those of *S. cerevisiae* and *S. kudriavzevii* alleles, indicating that they contain chromosomes from the three parental species. The *S. cerevisiae* and *S. kudriavzevii*

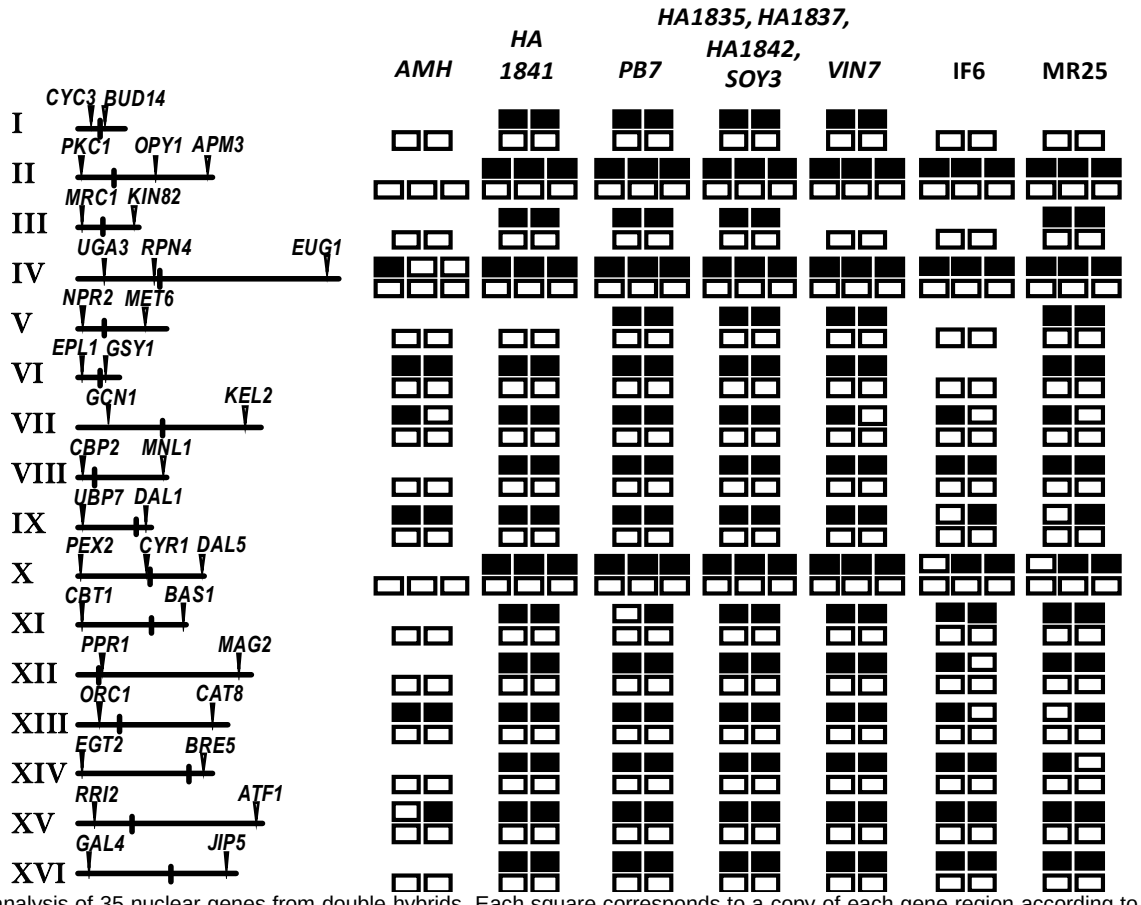


Figure 2. RFLPs analysis of 35 nuclear genes from double hybrids. Each square corresponds to a copy of each gene region according to its chromosome location, indicated on the left map. Alleles of *S. cerevisiae* are indicated as white squares and *S. kudriavzevii* alleles are represented as black squares.

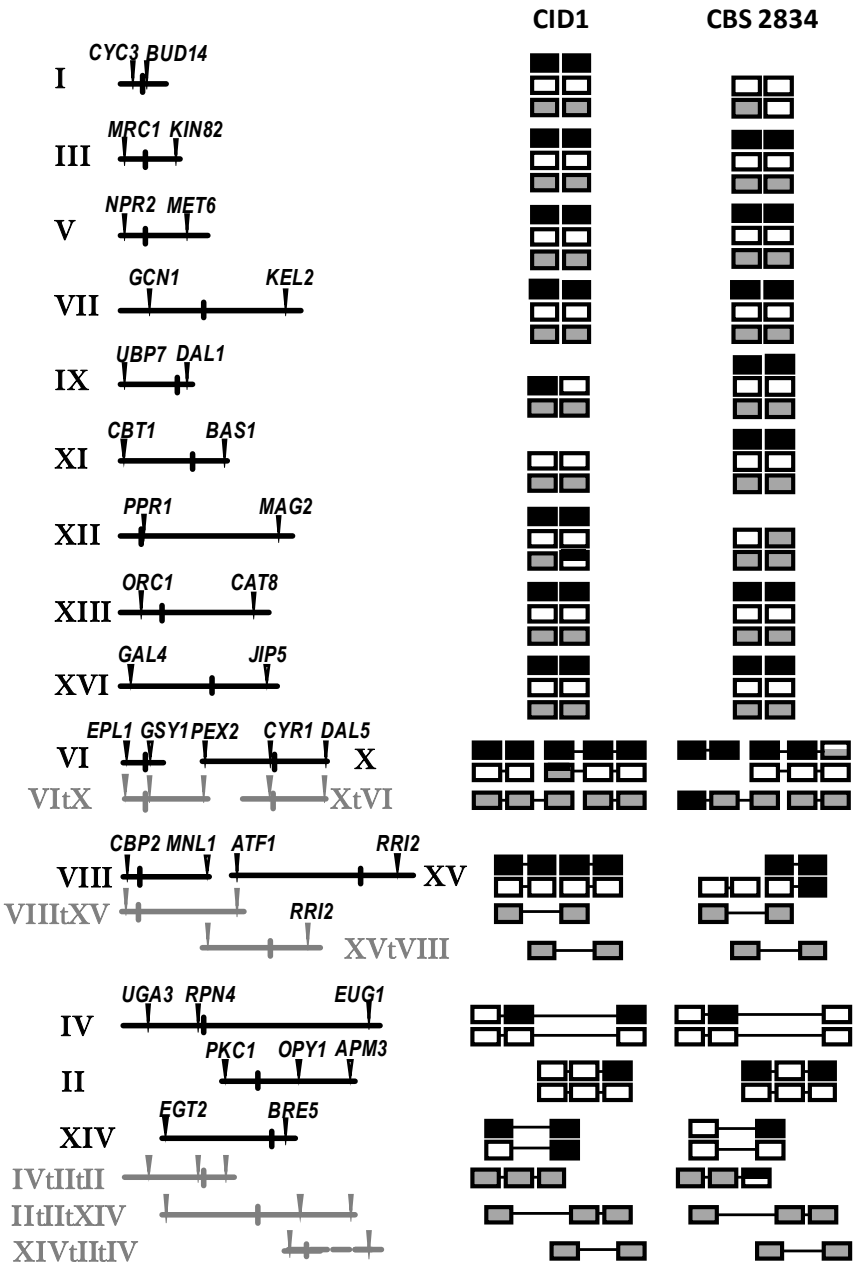


Figure 3. RFLPs of 35 nuclear genes from triple hybrids. Each square corresponds to a copy of each gene region according to its chromosome location, indicated on the left map. Alleles of *S. cerevisiae* are indicated as white squares, *S. kudriavzevii* alleles are represented as black squares and *S. bayanus* var. *uvarum* alleles are depicted in grey squares. Squares filled with two colors indicate that the presence of any of these alleles is possible. Gene orders are the same for *S. cerevisiae* and *S. kudriavzevii* because their genomes are colineal, however, gene orders differ for *S. bayanus* var. *uvarum* because this species exhibits a series of reciprocal translocations as depicted.

chromosomes are co-lineal (synthenic), however, the chromosomes of *S. bayanus* var. *uvarum* contain 4 differential reciprocal translocations (Kellis *et al.,* 2003), as depicted in Figure 3. In the case of triple hybrids, a higher preservation of the *S. bayanus* var. *uvarum* fraction is observed.

The comparison of the RFLP patterns obtained in this study for the new hybrids and those described by González *et al.,* (2008, see their figure 3) reveals a considerable diversity in the genome structure of *S. cerevisiae* x *S. kudriavzevii* hybrids, although certain similarities among strains are observed as well. Accordingly, double hybrid strains can be classified in three groups according to the parental genome rearrangements. The first group includes hybrids that maintain the complete genome from both parents (most HA strains and SOY3) or have independently lost from 1-2 chromosomes or chromosome regions from *S. kudriavzevii* (wine strains PB7, VIN7 and most brewing hybrids), the second group comprises strains with a moderate loss (3-4) of *S. kudriavzevii* chromosomes or chromosome regions, including 3 shared events (Swiss wine hybrids and the brewing strain CECT 11003), and the third group includes strains with moderate (MR25, 6 losses) to large *S. kudriavzevii* gene losses (CECT 11002, IF6 and AMH, with 9, 11 and 13, respectively).

3.3 Mitochondrial inheritance in hybrids

The analysis of mitochondrial *COX2* gene sequences has been shown as useful to decipher which parental species contributed with their mitochondria to the hybrid strains (González *et al.,* 2006).

The comparative analysis of *COX2* sequences with those previously described (González *et al.,* 2006), showed the presence of new haplotypes in hybrids PB7, AMH and IF6 (Fig. 2 and 3). The wine hybrids AMH and IF6 contain COX2 sequences more related to *S. cerevisiae* (1 and 14 differences, respectively being the first description of *S. cerevisiae* x *S.*

kudriavzevii hybrids that received their mitochondrial genomes from a *S. cerevisiae* parent.

The other new hybrids contain *COX2* sequences that correspond to previously described haplotypes. Thus, with the exception of PB7, all new wine hybrids contain haplotype K4, already described in the triple hybrid CBS 2834. This haplotype is closely related to haplotypes K2 and K3 from Swiss wine hybrids (1 and 2 nucleotide differences, respectively) and haplotypes exhibited by the Japanese type (haplotype K1, 5 differences) and European strains from *S. kudriavzevii* (haplotypes K8 and 9, with 1 and 3 differences, respectively). The clinical isolate MR25 exhibits the same haplotype K6 described in brewing hybrids, which is related to haplotype K10 present in the wine hybrid PB7 (6 nucleotide differences).

However, a detailed analysis of the *COX2* sequence alignment suggested the possibility of reticulate evolution due to recombination (Table 2). This way, haplotypes K5 (triple hybrid CID1), K6 (brewing hybrids CECT 1388, 1990, 11002, 11011 and the clinical strain MR25) and K10 (wine hybrid PB7) appear as putative recombinant sequences with similarities to *S. kudriavzevii*, *S. cerevisiae* and *S. paradoxus* sequences in their 5'-end, central and 3'-end regions, respectively (see Table 2).

In the case of reticulate evolution due to recombination, a better representation of the phylogenetic relationships is obtained by a Neighbor-net network analysis (Figure 4). Most wine hybrids (except PB7 and AMH) and two Trappist beer hybrids (CECT 11003 and 11004) inherited their mitochondrial genomes (haplotypes K2, K3 and K4) from *S. kudriavzevii*, AMH and IF6 received their mitochondrial genomes from *S. cerevisiae*, although IF6 *COX2* appears in a striking intermediate position between *S. cerevisiae* and *S. paradoxus*-*S.mikatae* clades, likely due to its highly divergent 3' end. Finally, most brewing hybrids and the clinical isolate

Table 2. Comparison of COX2 haplotype sequences from hybrid and type and reference strains of *Saccharomyces* species. A dot indicates nucleotides identical to that from the type strain of *S. cerevisiae* CECT 1942^{T}. *COX2* regions in hybrids that exhibit a higher similarity to *S. cerevisiae*, *S. kudriavzevii* and *S. paradoxus COX2* sequences are indicated in squared white, black and grey backgrounds, respectively.

			COX2 variable nucleotide positions (in vertical)
			1111111123333333444444455555555555555555555555555
		COX2	4467902333557501279992335688001111112222223333444556 6
Species	Strains	haplotype	0643424046475343801474037347280147890345692589014036 58
S. cerevisiae	CECT1942^{T}	C1	ATTAATTTATTTTATATTCTATTATTTTACTCTAGCATTCTGGTGACATATGGC
Hybrids	AMH	C2	.....G.....
	IF6	C3	.....A.....CA....AAA.ACT.ATC.AT
	CID1	K5	C.A.T..CC.....T....CA.C..A..C...TACCAAT
	PB7	K10	C.A.T..CC.....T....CA.C..A..C.CTGACCAAT
	MR25 & brewing	K6	.T..CC.....T...GCA.C..A..C.CTGACCAAT
	Swiss & 11003-4	K2	
	W46	K3	
	HAs, SOY3, VIN7	K4	
S. kudriavzevii	IFO1802^{T}	K1	
	ZP542	K8	
	ZP591	K9	
S. paradoxus	CECT1939^{NT}	K2	.....C.A.T..CC.....T...G...CAG.A.C.CTGACCAAT
S. mikatae	IFO1815^{T}	K2	....G.....TA.T...C...CA.T...G...CAG...CACTGACC.AT

A dot indicates nucleotides identical to that from the type strain of *S. cerevisiae* CECT 1942^{T}. *COX2* regions in hybrids that exhibit a higher similarity to *S. cerevisiae, S. kudriavzevii* and *S. paradoxus COX2* sequences are indicated in squared white, black and grey backgrounds, respectively.

(haplotype K6), the cider CID1 (K5) and the wine PB7 (K10) hybrids appear in an intermediate position due to their chimerical *COX2* sequences.

3.4 Different groups of hybrids according to their nuclear and mitochondrial genome constitutions

The combined analysis of the nuclear and mitochondrial genome compositions of *S. kudriavzevii* double and triple hybrids indicates a higher genetic diversity. Strains that differed in a few chromosomal rearrangements contain different mitochondrial haplotypes (e.g. PB7 and the Austrian and Croatian hybrids) and others showing important chromosomal differences share the same mitochondrial sequences (e.g. MR25 and brewing hybrids).

In other cases, there is a certain association between the nuclear and mitochondrial diversities. This way, the two hybrids with a *S. cerevisiae* mitochondrial DNA are those that lost a higher fraction of *S. kudriavzevii* nuclear genome. As well, with the mentioned exception of PB7, wine hybrids appear in two closely related clusters, the Austrian-Croatian cluster (also including VIN7) with low number of chromosomal rearrangements and the sharing the same *S. kudriavzevii*-like mitochondrial haplotype K4, and the Swiss cluster (also including Trappist hybrids CECT11003 and 11004), which share several fixed rearrangements (Belloch *et al.,* 2009) and the *S. kudriavzevii*-like mitochondrial haplotype K2 (including the derived K3).

In the case of the two triple hybrids known so far, they also show important differences both in their mitochondrial and nuclear genomes. Thus, these strains do not share any common chromosomal rearrangements indicating independent losses in the three fractions of their hybrid genomes. Moreover, the wine triple hybrid inherited a *S. kudriavzevii* mitochondrial genome similar to that present in wine double hybrids, whilst the cider hybrid contains a mitochondrial *COX2* closely related to that

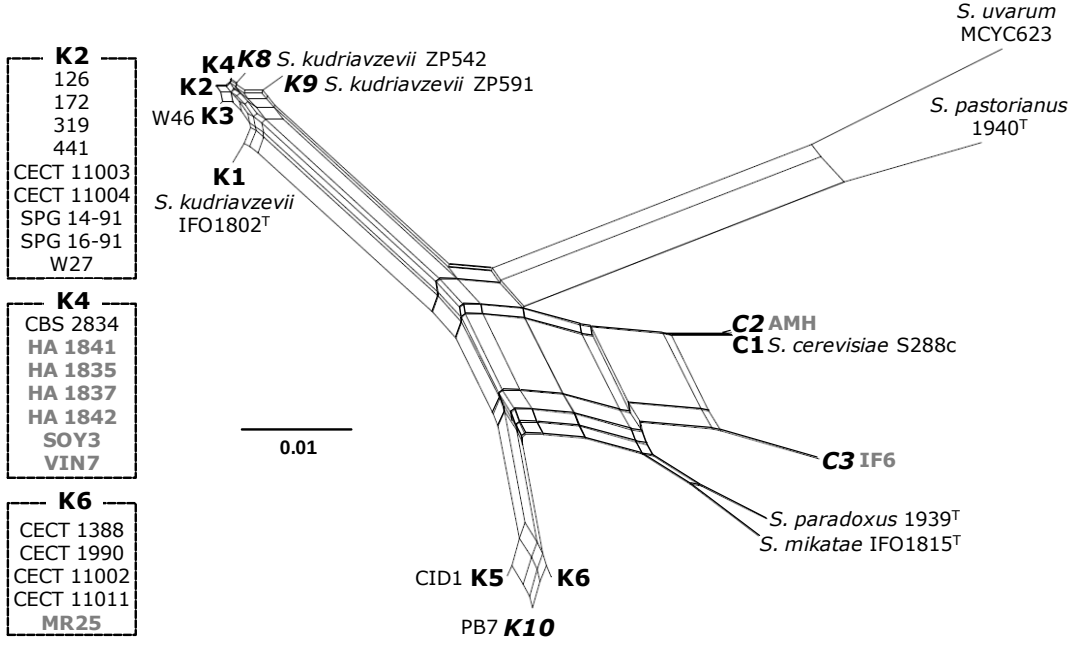


Figure 4. Phylogenetic Neighbor-net network obtained with partial sequences of the mitochondrial *COX2* gene from hybrid strains and reference *Saccharomyces* strains. The new hybrids are indicated in bold gray characters. The different *COX2* sequence haplotypes are named by the initial of the species name of the closest parental (C, for *S. cerevisiae*; and K, for *S. kudriavzevii*) followed by a number, according to González *et al.,* (, 2008). The new *COX2* haplotypes described in the present study are indicated in italics. Strains sharing the same haplotype are given at the left.

present in most brewing, the clinical and a wine hybrid with similarities intermediate between *S. cerevisiae* and *S. kudriavzevii*.

4. Discussion

4.1 New strains expanding the distribution range of Saccharomyces kudriavzevii *hybrids*

It is more than a decade since an unusual *S. bayanus* x *S. cerevisiae* hybrid, CID1, isolated from home-made Breton cider, was identified as bearing a mitochondrial genome coming from *S. kudriavzevii* (Masneuf *et al.,* 1998; Groth *et al.,* 1999). Later, a *S. kudriavzevii* contribution to a fraction of the chimerical nuclear genome of this strain was demonstrated (Naumova *et al.,* 2005; González *et al.,* 2006).

Some years later, a new type of natural hybrid strains between *S. cerevisiae* x *S. kudriavzevii* was described in wine fermentations (Bradbury *et al.,* 2006; González *et al.,* 2006; Lopandic *et al.,* 2007). and brewing environments (Gonzalez *et al.,* 2008).

In the present study, new *S. cerevisae* x *S. kudriavzevii* hybrid yeasts are described and molecularly characterized. These hybrids contribute to expand the geographical distribution range of this type of hybrids as well as the sources whence they can be isolated.

This way, the new wine hybrids (PB7 and SO3) were isolated from wine fermentation in the southernmost locations where this kind of hybrids has been isolated so far (Pajares de los Oteros, in Northwestern Spain, and Daruvar, in Central Croatia, respectively). These new descriptions extend the distribution limits of *S. cerevisiae* x *S. kudriavzevii* hybrids to the Southern limit of the European wine regions of Oceanic and Continental climate, where these hybrids have been found so far associated to fermentation processes. In these wine regions, hybrids can be predominant (Schütz & J Gafner 1994; González *et al.,* 2006; Lopandic *et al.,* 2007)

likely due to a better adaptation to lower temperatures compared to *S. cerevisiae* (González *et al.,* 2007).

The molecular characterization of PB7 showed that, although its nuclear genome composition is similar to other wine hybrids, exhibits a recombinant mitochondrial genome different but closely related to brewing hybrids. Its marginal distribution and its peculiar genome characteristics are indicative of a putative independent origin from other wine hybrids. However, the genome composition of the Croatian SOY3 hybrid was identical to Austrian hybrids, predominant in another wine region of the same Pannonian basin (Lopandic *et al.,* 2007), with similar climatologic characteristics as well as historical links in the development of viticulture and enology.

In these Southern locations where the new wine hybrids were isolated, hybrids did not appear as predominant. In both cases, these wine hybrids were found at low frequencies and coexisting with the dominant *S. cerevisiae* strains during the first stages of the wine fermentations. Perhaps the milder temperatures at which spontaneous fermentations occur in these Southern regions still allow *S. cerevisiae* to outcompete these hybrids.

The present study also describes for the first time *S. cerevisiae* x *S. kudriavzevii* hybrids isolated from non-fermentative environments. Strain MR25 is a human respiratory isolate from 'Hospital del Vall d'Hebron', Barcelona, Spain; and IF6, is commercialized as a dietary supplement. These hybrids are quite different at the genome level, particularly in their mitochondrial genomes. The clinical isolate MR25 shares a *COX2* sequence identical to that present in 4 brewing hybrids, indicating that beer could likely be the source of infection, and the dietary supplement IF6 exhibits a S*. cerevisiae* mitochondrial DNA.

4.2 The high genetic diversity among Saccharomyces kudriavzevii *hybrids suggests independent hybridization origins*

The analysis of the nuclear and mitochondrial genome compositions of *S. kudriavzevii* double and triple hybrids unveiled a high diversity, which likely is indicative of independent primary, as well as secondary, hybridization events.

The fact that hybrids inherited 3 types of mitochondrial genomes (*S. cerevisiae*-like, *S. kudriavzevii*-like and recombinant) from their parental ancestors is indicative of at least 3 different origins. Moreover, the important differences in their nuclear genome compositions could also be taken as evidences of independent primary hybridization events.

The presence of recombinant mitochondrial genomes in hybrids can be explained by recombination events occurring after the fusion of mitochondria observed in conjugating *Saccharomyces* spores or cells. This kind of recombination events were already described in *S. cerevisiae* at the within-species level (Berger & Yaffe 2000), but this is the first time that is described in hybrids at the between-species level. However, we suspect that these recombination events are limited to this *COX2* region because sequences from the next downstream gene, *COX3*, correspond to *S. kudriavzevii* (data not shown).

In addition, the existence of natural triple *S. bayanus* var. *uvarum* x *S. cerevisiae* x *S. kudriavzevii* hybrids can be explained by secondary hybridization between either *S. cerevisiae* x *S. kudriavzevii* hybrids with *S. bayanus* var. *uvarum* strains or *S. bayanus* var. *uvarum* x *S. cerevisiae* hybrids with *S. kudriavzevii* strains. Although both types of double hybrids have been found associated to fermentation environments, the first type of secondary hybridization event could be more probable because *S. kudriavzevii* seems to be present only in natural environments (Sampaio & Gonçalves 2008; Lopes *et al.,* 2010) and is outcompeted by *S. cerevisae* in

experimental wine fermentations (Arroyo-López *et al.,* 2011), whilst *S. bayanus* var. *uvarum* coexists with, or even replaces, *S. cerevisiae* in wine fermentations from cold regions of Europe (Torriani *et al.,* 1999; Naumov *et al.,* 2000, 2002; Rementeria 2003; Demuyter *et al.,* 2004). However, a secondary hybridization event in natural environments, involving a *S. kudriavzevii* and a *S. bayanus* x *S. cerevisiae*, cannot be totally discarded.

After hybridization, the hybrid genome suffers random genomic rearrangements mediated by crossing-over between homeologous chromosomes (Belloch *et al.,* 2009). If these rearrangements were randomly fixed, hybrids with a higher number of rearrangements should derive from older hybridization events, and hybrids with no rearrangements should be very recent. However, double hybrids showed a trend to maintain the *S*. *cerevisiae* genome and to reduce the *S. kudriavzevii* that can only be explained by selection acting under the strong restrictive conditions prevailing during fermentation (nutrient depletion, osmotic stress, fermenting temperature, increasing levels of ethanol, etc.). The better adaptation of *S. cerevisiae* to these prevailing conditions constrains the loss of the *S. cerevisiae* fraction of the hybrid genome, and only the *S. kudriavzevii* genome fraction of selective importance for the hybrid (e.g. involved in adaptation to low fermentation temperatures) would be maintained. The fact that hybrids with a *S. kudriavzevii* mitochondrial genome maintain a larger fraction of the *S.kudriavzevii* genome than hybrids with a *S. cerevisiae* mitochondrial DNA, such as AMH and IF6, is also indicative that the inheritance of a *S. kudriavzevii* mitochondrial genome constrains to maintain those *S. kudriavzevii* genes involved in the proper function and maintenance of the mitochondria. Incompatibility between nuclear and mitochondrial genes has been reported for artificial *S. cerevisiae* x *S. bayanus* hybrids (Lee *et al.,* 2008). Accordingly, strains possessing *S. cerevisiae*-inherited mitochondria overcome this restriction

and may lose these *S. kudriavzevii* mitochondrial-related genes from their nuclear genome.

Author's contribution

This study is the result of the collaboration between AQ and EB laboratories. CB, AQ and EB conceived and supervised this study. DP, CB, AQ and EB designed the experiments. DP performed the experimental work and data analyses. KL and JMA provided the potential hybrids from Austria and León (Spain), respectively. DP and CB wrote the first version of the manuscript. DP, CL, AQ and EB participated in the final manuscript revision.

Acknowledgements

We thank Helmut Gangl, Rosa de Llanos, Silvia LLopis, Sandi Orlić, Lallemand Bio and Anchor Wine Yeasts for providing yeast strains. This work was supported by Spanish Government projects AGL2009-12673-CO2-01 and AGL2009-12673-CO2-02 to AQ and EB, respectively, and Generalitat Valenciana (project PROMETEO/2009/019) to AQ, EB and CB. DP and JMA-P acknowledge to the Spanish Government for their FPI (Ministerio de Ciencia e Innovación) and FPU (Ministerio de Educación) fellowships, respectively.

<u>OBJECTIVE 2.</u> Genome characterization of natural *Saccharomyces cerevisiae* x *Saccharomyces kudriavzevii* hybrids.

Chapter 1

Comparative genomics among *Saccharomyces cerevisiae* x *Saccharomyces kudriavzevii* natural hybrid strains isolated from wine and beer reveals different origins

David Peris[1], Christian A. Lopes[2,3], Carmela Belloch[2], Amparo Querol[2], and Eladio Barrio[1,*]

[1]'Cavanilles' Institute of Biodiversity and Evolutionary Biology, University of Valencia, Parc Científic, P.O. Box 22085. E-46071 Valencia, Spain.

[2]Department of Biotechnology, Institute of Agrochemistry and Food Technology (CSIC), Valencia, Spain.

[3]Yeast Biodiversity & Biotechnology Group. Instituto Multidisciplinario de Investigación y Desarrollo de la Patagonia Norte (IDEPA), CONICET, Universidad Nacional del Comahue, Neuquén, Argentina.

***Corresponding author:**

Dr. Eladio Barrio
Institut 'Cavanilles' de Biodiversitat i Biologia Evolutiva
Universitat de València
P.O. Box 22085
E- 46071 València
SPAIN
Tel.: +34 963 543 667
Fax: +34 963 543 670
E-mail: eladio.barrio@uv.es

BMC Genomics (*in press*)

Abstract

Background: Interspecific hybrids between *S. cerevisiae* × *S. kudriavzevii* have frequently been detected in wine and beer fermentations. Significant physiological differences among parental and hybrid strains under different stress conditions have been evidenced. In this study, we used comparative genome hybridization analysis to evaluate the genome composition of different *S. cerevisiae* × *S. kudriavzevii* natural hybrids isolated from wine and beer fermentations to infer their evolutionary origins and to figure out the potential role of common *S. kudriavzevii* gene fraction present in these hybrids.

Results: Comparative genomic hybridization (CGH) and ploidy analyses carried out in this study confirmed the presence of individual and differential chromosomal composition patterns for most *S. cerevisiae* × *S. kudriavzevii* hybrids from beer and wine. All hybrids share a common set of depleted *S. cerevisiae* genes, which also are depleted or absent in the wine strains studied so far, and the presence a common set of *S. kudriavzevii* genes, which may be associated with their capability to grow at low temperatures. Finally, a maximum parsimony analysis of chromosomal rearrangement events, occurred in the hybrid genomes, indicated the presence of two main groups of wine hybrids and different divergent lineages of brewing strains.

Conclusion: Our data suggest that wine and beer *S. cerevisiae* × *S. kudriavzevii* hybrids have been originated by different rare-mating events involving a diploid wine *S. cerevisiae* and a haploid or diploid European *S. kudriavzevii* strains. Hybrids maintain several *S. kudriavzevii* genes involved in cold adaptation as well as those related to *S. kudriavzevii* mitochondrial functions.

Keywords: *Saccharomyces* hybrids*, S. cerevisiae, S. kudriavzevii*, wine, beer, comparative genomics.

1. Introduction

The development of molecular methods of yeast characterization has demonstrated that some wine and brewing *Saccharomyces* strains possess complex genomes composed by genetic elements from two or more species (Barros Lopes *et al.,* 2002; Gonzalez *et al.,* 2006; González *et al.,* 2008; Groth *et al.,* 1999; Liti *et al.,* 2005; Lopandic *et al.,* 2007; Masneuf *et al.,* 1998). These strains are widely known as interspecific hybrids.

The best characterized industrial interspecific hybrid is the lager yeast *S. pastorianus*, originated from hybridization between *S. cerevisiae* and a *S. bayanus*-related yeast, which recently has been suggested to belong to the new species *S. eubayanus* (Libkind *et al.,* 2011). The hybridization between *S. cerevisiae* and the cryotolerant *S. eubayanus* have been suggested as the result of selective pressures derived from brewing at low temperatures (Libkind *et al.,* 2011).

Other kind of natural *Saccharomyces* hybrids are those originated from hybridization between *S. cerevisiae* and *S. kudriavzevii*. These hybrids have mainly been isolated from wine and brewing environments (Gonzalez *et al.,* 2006; Lopandic *et al.,* 2007; González *et al.,* 2008).

The role of the *S. kudriavzevii* genome in these hybrids is unclear, since the known strains of this species have been found in decaying leaves from Japan and oak trees from Portugal and Spain (Sampaio *et al.*, 2008; Lopes *et al.*, 2010), but not in fermentative industrial environments yet. The physiological evaluation of some of these *S. kudriavzevii* isolates showed that this species is characterized by a higher cryotolerance than *S. cerevisiae*, but a lower ethanol tolerance (Arroyo-López *et al.*, 2009; Salvadó *et al.*, 2011).

Albeit differences between *S. cerevisiae* × *S. eubayanus* and *S. cerevisiae* × *S. kudriavzevii* hybrids, the role of the *S. eubayanus* or *S.*

kudriavzevii genomes in the hybrid seems to be similar, that is, the maintenance of good fermentative performance at low temperatures.

The characterization of a particular group of Swiss wine hybrids by PCR-RFLP, DNA arrays, ploidy analysis and gene dose determination by quantitative real-time PCR, evidenced the existence of a single common hybridization event to explain the origin of these hybrids followed by extensive chromosomal rearrangements including chromosome losses and the generation of chimerical chromosomes (Belloch *et al.*, 2009).

In this work, genome composition by array-CGH of a more diverse set of wine and brewing *S. cerevisiae* × *S. kudriavzevii* natural hybrids from diverse origins was evaluated to decipher their origins and evolution. The examination of gene losses and gains as well as the maintenance of specific metabolic pathways from the *S. cerevisiae* or *S. kudriavzevii* parental genomes was also analyzed with the aim of elucidating the role of each parental genome in the fermentative performance of the hybrid strains.

2. Material and methods

2.1 *Yeast strains and culture media*

The natural yeast hybrids *S. cerevisiae* × *S. kudriavzevii* used in this study have been isolated from wine and brewing fermentations in different locations (Table 1). The haploid strain *S. cerevisiae* S288c was used as control for microarray DNA hybridizations. Yeast strains were grown at 28°C in GPY medium (2% glucose, 0.5% peptone, 0.5% yeast extract).

2.2 Ploidy estimations by flow cytometry

Ploidy estimates are very important to interpret aCGH data from hybrids because hybridization signals are commonly normalized with respect to those of the reference haploid strain S288c.

The DNA content of both hybrid and control strains was assessed by flow cytometry by two different procedures. The first ploidy estimates were obtained in a FACScan cytometer (Becton Dickinson Inmunocytometry Systems, California, United States) by using the propidium iodide dye method described in Belloch *et al.,* (2009). Due to discrepancies with the aCGH analysis, new estimates were later obtained in a Beckman Coulter FC 500 (Beckman Coulter Inc., California, USA) by using the SYTOX Green dye method described in Haase and Reed (Haase and Reed, 2002). In both cases, ploidy levels were scored on the basis of the fluorescence intensity compared with the haploid (S288c) and diploid (FY1679) reference *S. cerevisiae* strains. Ploidy reported for each strain is the result of three independent measures. Results were tested by one way ANOVA and Tukey's HSD tests.

Table 1. List of hybrid strains used in this study.

Strain name	**Isolation source**
HA1841	wine, Perchtoldsdorf, Austria
HA1842	wine, Perchtoldsdorf, Austria
PB7	wine Pietro Picudo, León, Spain
Assmanhausen (AMH)	wine, Geisenheim, Germany
Anchor VIN7	commercial strain, Anchor, South Africa
SOY3	wine, Daruvar, Croatia
CECT1388	ale beer, United Kingdom
CECT1990	beer, Göttinger Brauhaus AG, Germany
CECT11002	beer Chimay Trappist, Belgium
CECT11003	beer Orval Trappist, Belgium
CECT11004	beer, Westmalle Trappist, Belgium
CECT11011	brewery, New Zealand

2.3 DNA labeling and microarray competitive genome hybridization

Total DNA, extracted as described in Querol *et al.,* (1992), was resuspended in 50 µl of de-ionized water and digested with endonuclease *Hinf* I (Roche Applied Science, Germany), according to the manufacturer's instructions, to fragments of an average length of 0.25 to 8 kbp. Each sample was purified using High Pure PCR Product Purification Kit (Roche Applied Science, Germany) and 2 µg was labelled using BioPrime Array CGH Genomic Labelling System (Invitrogen, California, USA). Unincorporated label was removed using MinElute PCR Purification Kit (Qiagen, Germany). Equal amounts of labelled DNA from the corresponding hybrid strains and the control S288c strain were used as probes for microarray hybridization.

Array competitive genomic hybridization (aCGH) was performed using a double-spotted array containing 6,240 ORFs of *S. cerevisiae* plus control spots totaling 6.4 K (Microarray Centre, University Health Network, Toronto, Canada). New microarrays were pre-treated for one hour at 65°C with pre-hybridization solution (7.5 ml 20× SSC, 0.5 ml 10% SDS, 0.5 ml 10 mg/ml bovine serum albumin in 50 ml final volume). Pre-hybridization solution was washed during 15 s in mili-Q H_2O, 2 s in 2-propanol, 2 s in milli-Q H_2O and dried by centrifugation at 1200 rpm, 10 min. Microarrays were treated with hybridization solution (15 µl SSC, 0.6 µl 10% SDS, 6 µl 1 mg/ml salmon DNA and DNA labelled in 60 µl final volume) at 95°C for 1 min and at room temperature for 5 min before DNA hybridization. Hybridization was performed for 18 h in chamber at 65°C, thus allowing hybridization of the *S. cerevisiae* part of the hybrid genome. A negative control of microarray hybridization was done by using DNA from *S. kudriavzevii* IFO 1802 strain vs. S288c. After hybridization microarrays were washed at 65°C for 5 min in 2× SSC, 0.1% SDS, at room temperature

in 0.1× SSC− 0.1% SDS for 10 min and six times in 0.1× SSC 1 min and dried by centrifugation at 1200 rpm, 10 min.

Experiments were carried out in duplicates and Cy5-dCTP and Cy3-dCTP dye-swap assays were performed to reduce dye-specific bias. The aCGH was performed for all hybrid strains except for W27, W46, SPG16-91 and SPG441 previously analyzed by Belloch *et al.,* (2009).

2.4 Microarray scanning and data normalization

Microarray scanning was done by using a GenePix Personal 4100A scanner (Axon Instruments/Molecular Devices Corp., California, USA). Microarray images and raw data were produced with the GenePix Pro 6.1 software (Axon Instruments/Molecular Devices Corp., California, USA) and background was subtracted by applying the local feature background median option. M-A plots ($M = Log_2$ ratios; $A = \log_2$ of the product of the intensities) were represented to evaluate if ratio data were intensity-dependent. The normalization process and filtering were done using Acuity 4.0 (Axon Instruments/Molecular Devices Corp., California, USA). Raw hybridization signals from hybrids were normalized with respect to those of the reference haploid strain S228c by using the ratio-based option, in which average hybridization ratios are adjusted to 1 (and hence, the corresponding $\log_2$ values to 0).

Normalized data were filtered by regression correlations $635/532 > 0.6$, signal intensity in both channels more than 350 units, and signal to noise (SNR) > 2.5. Features with artifacts or flagged as bad were removed from the analysis. Replicates were averaged after filtering. It is worth to remark that strong normalization factors were applied to the negative control signal in each channel (2 to the red and 0.46 to the green one). Raw data and normalized microarray data are available in ArrayExpress (Brazma, 2003), under the ref. E-MEXP-3114.

2.5 Chromosome structure and recombination sites in the chimerical chromosomes

The $\log_2$ of normalized Cy5/Cy3 signal ratio obtained for each ORF was represented with respect to its corresponding chromosomal location using the completely sequenced reference *S. cerevisiae* strain S288c. These plots, called caryoscopes, were generated using ChARM v.1.1 (Myers *et al.*, 2004). Highly stringent hybridization conditions (65°C) were used to avoid the cross hybridization of *S. kudriavzevii* DNA present in the hybrids. The caryoscope of the negative control experiment showed that most *S. kudriavzevii* genes did not hybridize under these conditions and in the case of cross hybridization (red signal) this was due to the very strong normalization factors applied in these control, which increased the red signal and reduced the green one by factors not applied in the case of the experiments performed with DNA from hybrids (see Figure S1). Accordingly, differences in the $\log_2$ ratio values observed in the caryoscopes revealed variations in the relative copy number of *S. cerevisiae* genes present in the hybrid strains.

The identification of over- and underrepresented regions was confirmed due to the normalization procedure, the hybridization ratios derived from aCGH analysis show the relative proportions of each gene with respect to the average number of copies in the hybrid, allowing the identification of over- and underrepresented regions in the hybrid genome by a one-way ANOVA test to determine the different levels of hybridization observed in the aCGH analysis. The approximate locations of the recombination points in the mosaic chromosomes were determined from the up and down jump locations in the ORFs mapping by microarray analysis of the hybrid yeast genomes.

Finally, by considering the collinearity of *S. kudriavzevii* and *S. cerevisiae* genomes (Cliften *et al.*, 2003), the *S. kudriavzevii* gene content in the hybrid genomes can be deduced from the presence/absence of the chromosome regions coming from each parental species, obtained in a previous PCR-RFLP analysis of these hybrids (Peris *et al.*, 2012a).

2.6 Gene Ontology (GO) analysis of S. kudriavzevii genes

GenMAPP v2.1 software (Doniger *et al.*, 2003) was used to perform gene ontology analysis of the *S. kudriavzevii* fraction in the hybrid genomes. Four different GO analyses were carried out using *S. kudriavzevii* genes present in all hybrid strains, including those previously characterized (Belloch *et al.*, 2009), these analyses corresponded to: i) the complete set of wine and brewing hybrids, except strain AMH, showing the lowest *S. kudriavzevii* gene content, ii) only wine hybrids, except AMH, iii) only brewing hybrids and iv) only AMH. In all cases, statistically significant GO term enrichments were shown by computing a *p*-value using the hypergeometric distribution (the background set of genes was 6241, the number of ORFs measured in microarray experiments). GO terms showing significant values (z-score >2 and *p*-value <0.05) were sorted according to their corresponding GO category.

2.7 Maximum parsimony tree

A list of minimal number of chromosomal rearrangements, chromosomal losses and restriction site changes were used to reconstruct the maximum parsimony tree. Data obtained from a previous study (Belloch *et al.*, 2009) were again included in this analysis. A binary matrix was constructed to codify each particular event (Table S1). Parsimony trees were constructed by PHYLIP 3.66 package using the Mix program (Felsenstein, 2005), taking chromosomal rearrangements and gain/losses as irreversible events (Camin-Sokal model) and the RFLP changes as

reversible events (Wagner model). The consensus tree was obtained with Consense program using the Majority rule.

3. Results

3.1 Hybrid genome structures

Caryoscopes, representing $\log_2$ hybridization ratios for each gene mapped onto its corresponding chromosome position, of six hybrid strains from wine and 6 hybrids from brewing were obtained by array comparative genomic hybridization (aCGH) (Figure S2). Due to the normalization procedure, the hybridization ratios derived from aCGH analysis show the relative proportions of each gene with respect to the average number of copies in the hybrid, allowing the identification of over- and underrepresented regions in the hybrid genome. However, aCGH analysis in combination with ploidy estimates and with information on the presence/absence of the chromosome regions coming from each parental species, obtained in a previous PCR-RFLP analysis of these hybrids (González *et al.*, 2008; Peris *et al.*, 2012a), allowed us to decipher the genome composition of hybrids.

This way, ploidy estimate for these hybrids were obtained by flow cytometry. The initial estimates with the propidium iodide method suggested that most hybrids were diploids or close to diploidy (relative C-values of 2.0 to 2.6). However, these ploidy values were not congruent with the caryoscope and PCR-RFLP data. The ratio-based normalization of hybridization signals adjusts the average signal ratios (problem strain/reference strain) to 1, and hence the $\log_2$ values to 0. In the analysis of hybrids, ploidy estimates were 2n-2.6n, corresponding on average to a subgenome coming from each parental species, i.e. for each gene there are on average a copy coming from *S. cerevisiae* and another from *S. kudriavzevii*. Due to the high astringent hybridization conditions used in the

aCGH analysis of hybrids, only the *S. cerevisiae* subgenome is hybridizing, as confirmed by the negative control performed with *S. kudriavzevii* DNA. Therefore, in the normalization of hybridization signals, these ratios correspond to the adjustment of average signals coming from 1 *S. cerevisiae* gene copy from the hybrid to 1 gene copy form the reference haploid *S. cerevisiae* strain. In the case of an increase of copy numbers in specific genes or chromosomal regions, $\log_2$ values should be higher than 0 (1, 2, etc. depending on the number of copies), but in the case of loss of *S. cerevisiae* gene copies in the hybrid, a ratio of 0 ($\log_2$ of $-\infty$) should be observed. However, 3–4 levels of $\log_2$ values, including negative but not infinite, are observed for some hybrids (Figure S2), which made difficult the interpretation of the aCGH results and suggested that ploidy estimates with propidium iodide were wrong.

Therefore, new ploidy estimates of hybrids were obtained by using SYTOX Green as the DNA-binding dye, because Haase and Reed (2002) demonstrated that improves linearity between DNA content and fluorescence, and decreases peak drift associated with changes in dye concentration, growth conditions or cell size. In this new ploidy analysis, Swiss wine hybrids analyzed in our previous study (Belloch *et al.*, 2009) were also included.

The statistical analysis of the new estimates showed two significantly different groups of hybrids according to ploidy levels: most hybrids, including the Swiss wine strains, appear as allotriploids and hybrids AMH and PB7 as allotetraploid yeasts (Table 2). The new ploidy estimates are in agreement with the different levels of hybridization observed in the aCGH analyses and also with the previous PCR-RFLP analysis of hybrids (Peris *et al.*, 2012a).

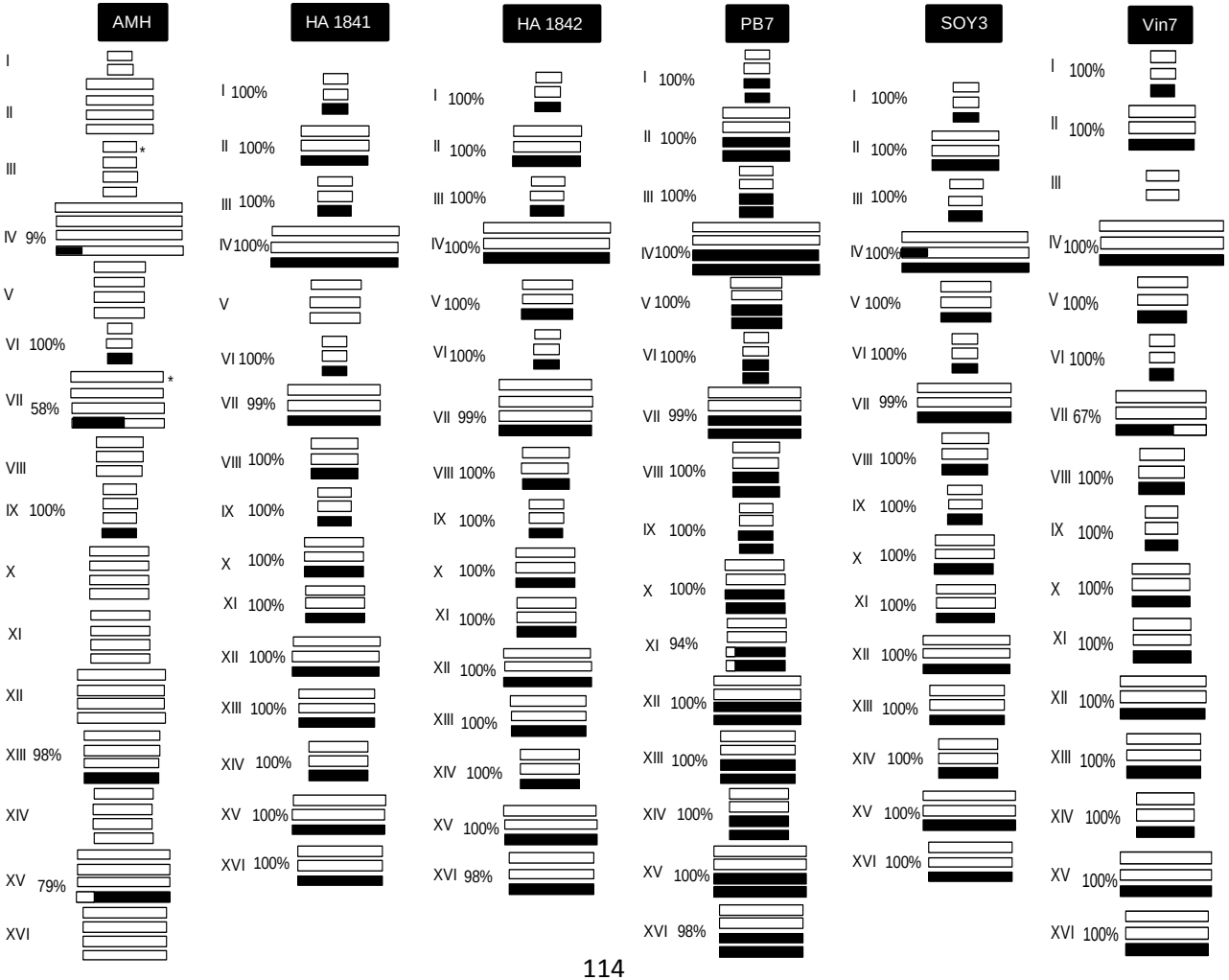
Table 2. DNA contents of natural hybrids, estimated by flow cytometry using the SYTOX green method with respect to the reference haploid and diploid *S. cerevisiae* strains, S288c and FY 1679, respectively. Results are the mean value of three replicates. Means with the same letters do not differ significantly by one way ANOVA and Tukey's HSD tests ($p < 0.05$).

Strain	DNA content relative to haploid strain S288c
FY1679	$2.00^{a} \pm 0.00$
HA 1841	$3.01^{b} \pm 0.08$
HA 1842	$3.07^{b} \pm 0.07$
VIN7	$3.04^{b} \pm 0.08$
SOY3	$2.89^{b} \pm 0.09$
CECT 1388	$3.25^{b} \pm 0.09$
CECT 1990	$2.86^{b} \pm 0.07$
CECT 11002	$3.02^{b} \pm 0.14$
CECT 11003	$3.21^{b} \pm 0.09$
CECT 11004	$3.13^{b} \pm 0.07$
CECT 11011	$2.99^{b} \pm 0.05$
W27	$3.18^{b} \pm 0.08$
W46	$3.20^{b} \pm 0.07$
441	$3.10^{b} \pm 0.09$
SPG16-91	$3.14^{b} \pm 0.08$
PB7	$3.96^{c} \pm 0.08$
AMH	$3.85^{c} \pm 0.18$

According to this combined analysis, 11 different patterns were differentiated in the 12 hybrids under analysis. As a general rule, different degrees of loss of *S. kudriavzevii* gene content in most hybrids were observed. Only the allotetraploid hybrid PB7 maintains a complete diploid set of chromosomes from each parental species, with the exception of small segment located in the left arm of chromosome XI of the *S.*

kudriavzevii subgenome. On the contrary, the largest reduction of the *S. kudriavzevii* gene content is observed in the partial allotetraploid hybrid AMH, which lost 72% of the *S. kudriavzevii* genes. The rest of hybrids, all of them allotriploid, showed intermediate situations derived from ancestors containing a diploid set of *S. cerevisiae* chromosomes and an haploid set of *S. kudriavzevii* chromosomes.

These combined analyses also allowed us to detect different types of chromosome rearrangements present in hybrids: i) the complete loss of a *S. kudriavzevii* parental chromosome compensated by an extra copy of the *S. cerevisiae* chromosome (chr. II,,III, V, X, XI, XII, XIV and XVI in AMH; chr. V in HA1841; chr. IV, IX and XII in CECT 11002; chr. I in CECT 11011); ii) aneuploidies (chr. I, VI and VIII in AMH; chr. IX in CECT 1388; chr. XIV in CECT 1990; chr. IX in CECT 11002; chr. III and V in CECT 11003 and CECT 11004; chr. III in VIN7), and iii) the presence of chimerical chromosomes (chr. IV, VII and XV in AMH; chr. XI in PB7; chr. IV in SOY3; chr. VII in VIN7; chr. VII and XIV in CECT 1388; chr. IV and XVI CECT 1990; chr. II, V, VII, X, XI, XIII and XIV in CECT 11002; chr. IV, V, VII, IX, XIV and XV in CECT 11003 and CECT 11004; and chr. VII in CECT 11011); (see Figure 1).


AMH
I
II
III
IV 9%
V
VI 100%
VII 58%
VIII
IX 100%
X
XI
XII
XIII 98%
XIV
XV 79%
XVI
HA 1841
I 100%
II 100%
III 100%
IV 100%
V
VI 100%
VII 99%
VIII 100%
IX 100%
X 100%
XI 100%
XII 100%
XIII 100%
XIV 100%
XV 100%
XVI 100%
HA 1842
I 100%
II 100%
III 100%
IV 100%
V 100%
VI 100%
VII 99%
VIII 100%
IX 100%
X 100%
XI 100%
XII 100%
XIII 100%
XIV 100%
XV 100%
XVI 98%
PB7
I 100%
II 100%
III 100%
IV 100%
V 100%
VI 100%
VII 99%
VIII 100%
IX 100%
X 100%
XI 94%
XII 100%
XIII 100%
XIV 100%
XV 100%
XVI 98%
SOY3
I 100%
II 100%
III 100%
IV 100%
V 100%
VI 100%
VII 99%
VIII 100%
IX 100%
X 100%
XI 100%
XII 100%
XIII 100%
XIV 100%
XV 100%
XVI 100%
Vin7
I 100%
II 100%
III
IV 100%
V 100%
VI 100%
VII 67%
VIII 100%
IX 100%
X 100%
XI 100%
XII 100%
XIII 100%
XIV 100%
XV 100%
XVI 100%

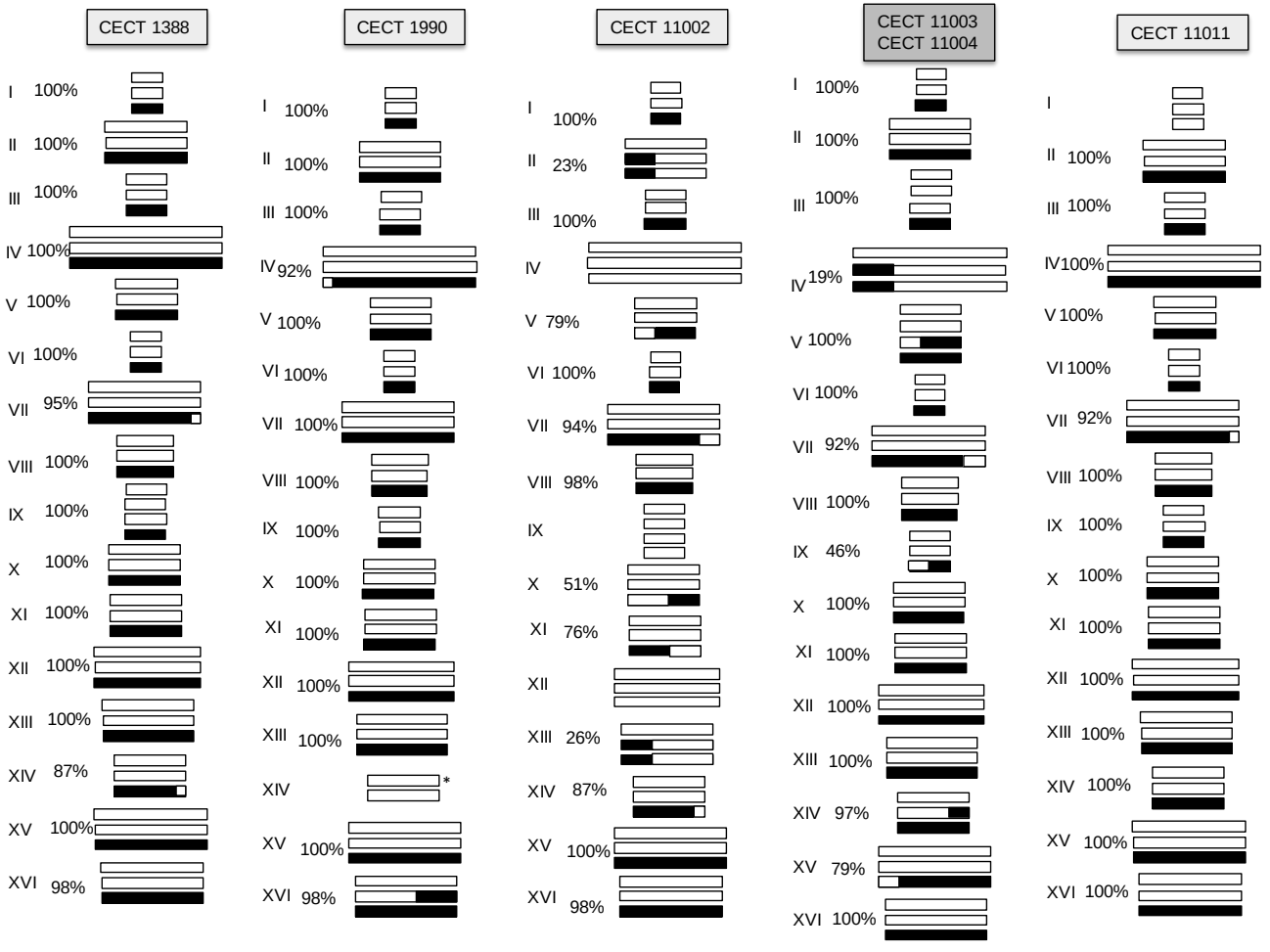

CECT 1388
I 100%
II 100%
III 100%
IV 100%
V 100%
VI 100%
VII 95%
VIII 100%
IX 100%
X 100%
XI 100%
XII 100%
XIII 100%
XIV 87%
XV 100%
XVI 98%
CECT 1990
I 100%
II 100%
III 100%
IV 92%
V 100%
VI 100%
VII 100%
VIII 100%
IX 100%
X 100%
XI 100%
XII 100%
XIII 100%
XIV *
XV 100%
XVI 98%
CECT 11002
I 100%
II 23%
III 100%
IV
V 79%
VI 100%
VII 94%
VIII 98%
IX
X 51%
XI 76%
XII
XIII 26%
XIV 87%
XV 100%
XVI 98%
CECT 11003
CECT 11004
I 100%
II 100%
III 100%
IV 19%
V 100%
VI 100%
VII 92%
VIII 100%
IX 46%
X 100%
XI 100%
XII 100%
XIII 100%
XIV 97%
XV 79%
XVI 100%
CECT 11011
I
II 100%
III 100%
IV 100%
V 100%
VI 100%
VII 92%
VIII 100%
IX 100%
X 100%
XI 100%
XII 100%
XIII 100%
XIV 100%
XV 100%
XVI 100%

Figure 1 Genome composition of hybrids deduced from aCGH analysis, ploidy estimates and a previous analysis of absence/presence of parental genes by RFLP analysis (González *et al.*, 2008; Peris *et al.*, 2012a). White and black bars are used to represent the *S. cerevisiae* and *S. kudriavzevii* genome fractions, respectively. Chromosomes showing black and white regions correspond to chimerical chromosomes. The percentages of *S. kudriavzevii* genes maintained in each chromosome are shown for each chormosome. Strains names are depicted on a black or a gray background corresponding to wine or brewing strains, respectively. Asterisks in AMH Chr. III and VII indicate regions where non-reciprocal translocations or segmental duplications can be present

These chimerical chromosomes are characterized by over- and underrepresented regions evidenced as up and down jumps in the $\log_2$ ratio in the caryoscopes, which is indicative of probable non-reciprocal recombination events between homeologous chromosomes (homologous from different species) (Table 3). The recombination sites in the chimerical chromosomes were mapped according to the genome browser from Saccharomyces genome database (SGD). Using a windows size of 15–20 Kb (four genes in the left and right of the most plausible recombination point) we found Ty elements, ARS sequences, clusters of homologous regions (CHRs) and tRNA elements that may have facilitated the recombination of the two homologous parental chromosomes (Table 3). In several cases, a common recombination site was observed in chromosomes belonging to two or more hybrids, indicative of common ancestry. This is the case of chromosomes IV, V, IX, XIV and XV in brewing hybrids CECT 11003 and 11004; chromosome XIV in CECT 1388 and 11002; chromosome XV in CECT 11003, 11004 and AMH and chromosome VII in hybrids CECT 11003, CECT 11004, CECT 11002, CECT 11011 and CECT 1388 (Table 3 and Figure S2).

Table 3. List of chimerical chromosome (CC) types found in the different *S. cerevisiae* × *S. kudriavzevii* hybrids. Chr., chromosome number; CHR, cluster of homology region. Strain names in italics correspond to wine hybrids and in bold to brewing hybrids. Some recombination sites were described elsewhere (Belloch *et al.*, 2009), as indicated

Chr.	CC type	Strains	Breakpoint mapping interval	Putative recombining sequences
II	type 1	**CECT 11002**	YBL018C-YBL011W	Ty1 LTR, Ty3 LTR, tRNA-Ile, tRNA-Gly, ARS
IV	type 1	*W27, W46, 441, SPG16-91*, **CECT 11003, CECT 11004**	YDL095W	*PMT1* (Belloch et al, 2009)
	type 2	*AMH, SOY3*	YDL185W-YDL179W	CHR 12
	type 3	**CECT 1990**	YDL185W-YDL179W	CHR 12
V	type 1	*W27, W46, 441, SPG16-91*, **CECT 11003, CECT 11004**	YER006W	*NUG1* (Belloch et al, 2009)
	type 2	**CECT 11002**	YEL018C-YEL011W	Ty1 LTR, Ty4 LTR, tRNA-Gln
VII	type 1	*W46,* **CECT 11003, CECT 11004, CECT 11002, CECT 11011, CECT 1388**	YGR249W-YGR244C	ARS, CHR 29
	type 2	*AMH*	YGR062C-YGR058W	CHR 30
	type 3	*VIN7*	YGR106C-YGR112C	tRNA-Leu, tRNA-Lys, Ty1 LTR, tRNA-Cys, Ty3 LTR, ARS
IX	type 1	*W27, W46, 441, SPG16-91*, **CECT 11003, CECT 11004**	YIL053W	*RHR2-RPL34B* (Belloch et al, 2009)
X	type 1	**CECT 11002**	YJL039C-YJL036C	tRNA-Asp, tRNA-Arg, Ty1 LTR, ARS, tRNA-Val
XI	type 1	**CECT 11002**	YKR025C-YKR028W	Ty1 LTR
	type 2	*PB7*	YKL203C-YKL204W	ARS
XIII	type 1	**CECT 11002**	YML012C-YML009W-B	CEN13, ARS
XIV	type 1	*W27, W46, 441, SPG16-91*, **CECT 11003, CECT 11004**	YNR001C	CEN14 (Belloch et al, 2009)
	type 2	**CECT 1388, CECT 11002**	YNR029C-YNR032W	ARS
XV	type 1	*W27, W46, 441, SPG16-91***, CECT 11003, CECT 11004,** *AMH*	YOL053W	*THI20-PSH1* (Belloch et al, 2009)
XVI	type 1	**CECT 1990**	YPR007C-YPR011C	Ty1LTR, tRNA-Gly, tRNA-Lys,

3.2 S. cerevisiae *gene depletions in hybrids*

Although hybrids maintain in their genomes at least a complete set of *S. cerevisiae* chromosomes, aCGH data from all hybrids analyzed in this work, as well as from those previously analyzed (Belloch *et al.*, 2009), can be used to determine the common fraction of *S. cerevisiae* genes showing gene copy variations in hybrids compared to the reference strain S288c. A common set of genes showing the same copy number variations in hybrids may be indicative of common origins.

The analysis of the *S. cerevisiae* gene content from all hybrids revealed the presence of less copies of a common set of genes. Among them, the most interesting were *CUP1*, *ASP3,* and *ENA* gene families, as well as Ty elements and 13 ORFs of unknown function (Table S2). In general, copy variations in the *S. cerevisiae* genome fraction of the hybrids were found in genes located in subtelomeric regions (Figure S2), although in some cases involve genes located in intrachromosomal regions, such as *CUP1*.

Short segment amplifications were also detected in the aCGH analysis. This was the case of hybrid AMH that showed three short region amplifications in chr. III, VII and XIII. The higher hybridization signals of genes located in the two first regions could be postulated as indicative of the presence of chimerical chromosomes, however according to the previous PCR-RFLP analysis *S. kudriavzevii* genes were absent. Other amplifications of *S. cerevisiae* segments located in chromosome XVI are observed in hybrids CECT 1388 (between genes YPL159C and YPL126W) and CECT 11002 (between YPL141C and YPL126W). Finally, a deleted region was found in one of the two copies of *S. cerevisiae* chromosome XIV from strain CECT 1990 (between loci YNR013C and YNR031C) (Figure S2).

3.3. S. kudriavzevii *gene content and Gene Ontology (GO) analyses*

Data obtained from all hybrids analyzed in this work as well as from those previously analyzed (Belloch *et al.*, 2009) were also used to evaluate the presence of common *S. kudriavzevii* genes (Table S3). These common set of genes could be interesting to unveil potentially genes of adaptive value in hybrids.

As a general rule, most hybrids maintained around 90% of the *S. kudriavzevii* genome, with the exception of the brewing strain CECT 11002 and the wine strain AMH which only maintain 56.9% and 30.5% respectively.

To determine if a group of *S. kudriavzevii* genes associated particular cellular components, molecular functions or biological processes may have been maintained in all hybrids due to potential adaptive value, four different gene ontology (GO) term enrichment analyses were performed (Table S4). The first analysis included all wine and brewing hybrids. Due to the low representation of the *S. kudriavzevii* genome fraction in AMH, this strain was removed from this first analysis. Gene ontology analysis was also separately performed according to the source of isolation of hybrids, wine and brewing fermentations. GO terms showing significant values were sorted according to their corresponding GO categories (Table S4). Table 4 shows only those significantly represented GO terms of putative importance for wine or brewing fermentations.

Significantly represented GO terms common to both wine and brewing hybrids mainly corresponded to genes related to fatty acid metabolism (particularly transport), sulfur metabolism and the NAD^+ salvage pathway. Genes associated with amino acid metabolism (N-linked glycosylation and glutamate metabolism) were also represented (Table 4).

Table 4. Summary of the most relevant metabolic pathways and biological processes obtained after Gene Ontology analysis using the *S. kudriavzevii* genes retained in each group of hybrids. Due to the massive S. kudriavzevii gene losses in AMH, this strain was not included in any grouping, and hence, analyzed alone.

Group of hybrids	GO ID	GO Name	$N_{present}/N_{measured}$	%	p-value
WINE	6487	Protein amino acid N-linked glycosylation	36/42	85.7	0.013
	6839	Mitochondrial transport	10/10	100	0.033
		Ergosterol Biosynthesis	17/19	89.5	0.049
BREWING	6487	Protein amino acid N-linked glycosylation	28/42	66.7	0.017
		Fatty acid elongation saturated	4/4	100	0.039
		Glycine serine and threonine metabolism	27/42	64.3	0.03
		Arginine_and_proline_metabolism	16/23	69.6	0.049
		Sulfur_Degradation	4/4	100	0.048
ALL	6487	Protein amino acid N-linked glycosylation	25/42	59.5	0.003
	15908	Fatty acid transport	4/4	100	0.025
		Glutamate metabolism	15/27	55.6	0.046
		Sulfur metabolism	8/11	72.7	0.021
		NAD salvage pathway	5/6	83.3	0.027
		Sulfate assimilation pathway II	5/6	83.3	0.019
AMH	6972	Hyperosmotic response	5/7	71.4	0.036
	9331	Glycerol 3 phosphate dehydrogenase complex	3/3	100	0.033
		Histidine biosynthesis	5/7	71.4	0.039
		Fatty acid metabolism	11/17	64.7	0.010

GO terms related to amino acid N-linked glycosilation were also significantly present in hybrids from wine and brewing analyzed independently. Moreover, GO terms associated with ergosterol biosynthesis and mitochondrial transport were also significantly detected in wine hybrids; while those related to metabolism of amino acids such as glycine, threonine, arginine and proline, sulfur metabolism, as well as fatty acid elongation were significant present in brewing strains (Table 4). Finally, an independent analysis of significant GO terms for AMH hybrid revealed the presence of genes involved in hyperosmotic response, glycerol-3-phosphate dehydrogenase complex, histidine biosynthesis and fatty acid metabolism (Table 4).

3.4 Phylogenetic relationships among hybrids

A maximum parsimony tree was constructed based in presence/absence of chromosomes and chromosome regions data obtained for each particular genetic event in all analyzed hybrids. The tree topology revealed the presence of two main groups containing most allotriploid hybrids, particularly those from wine (Figure 2).

Group I was constituted by Swiss wine strains W46, 441, W27 and SPG 16-91 as well as the brewing strains CECT 11003 and CECT 11004. This group is supported by the presence of five shared chimerical chromosomes as well as the *CYC3* K2 allele (González *et al.*, 2008).

Group II includes the remaining allotriploid wine hybrids HA1841, HA 1842, VIN7 and SOY3. This group is only supported by the common presence of *S. kudriavzevii* K2 alleles for genes *EUG1* and *APM3* (Peris *et al.*, 2012a), and the possession of a higher fraction of *S. kudriavzevii* genome.

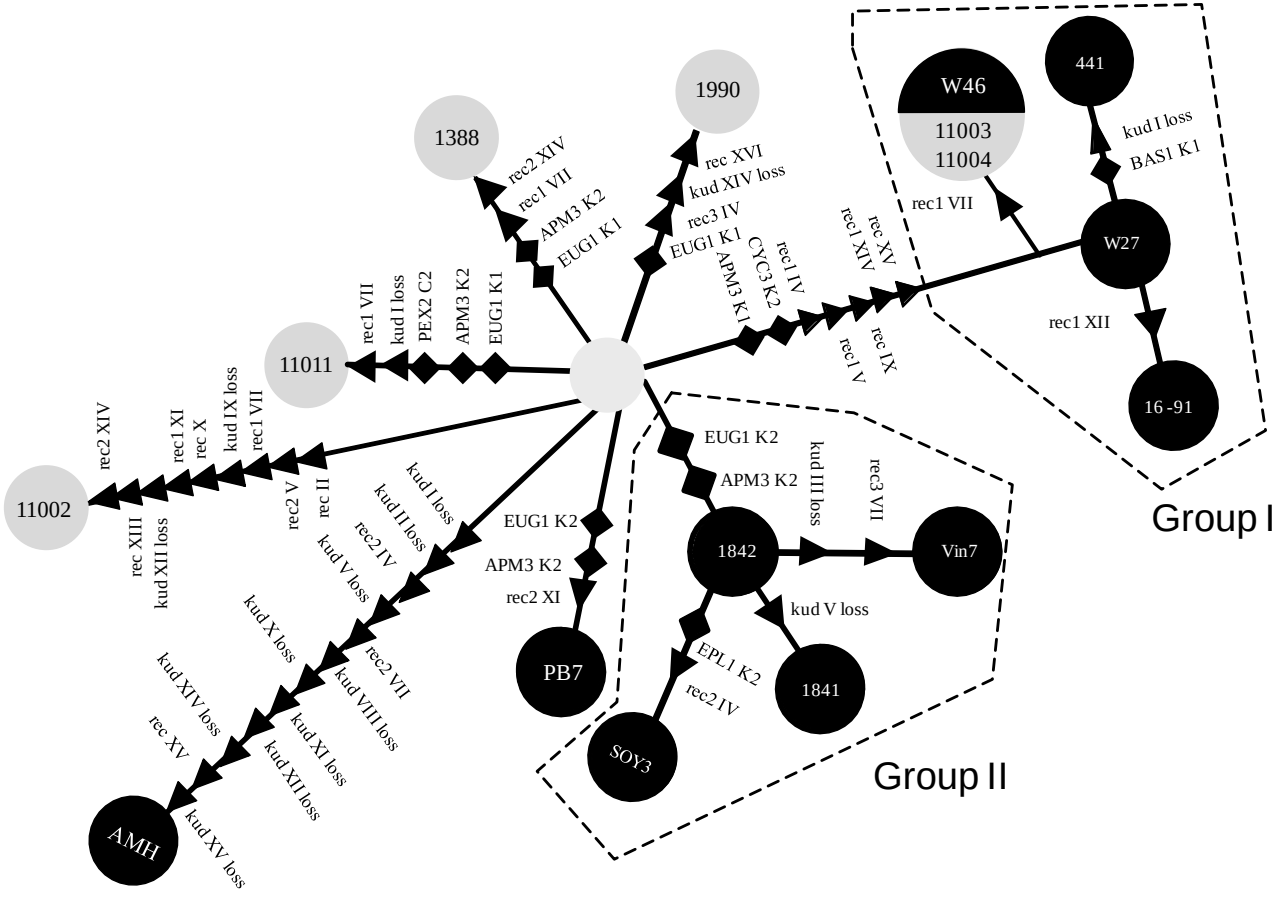

1388
1990
11011
11002
W46
11003
11004
441
W27
16 -91
1842
Vin7
1841
PB7
SOY3
AMH
Group I
Group II
rec2 XIV
rec1 VII
APM3 K2
EUG1 K1
rec XVI
kud XIV loss
rec3 IV
CYC3 K2
rec1 IV
rec XV
rec1 XIV
rec1 V
rec IX
kud I loss
BAS1 K1
rec1 XII
rec1 XI
rec X
kud IX loss
rec2 V
rec II
kud II loss
rec2 IV
kud V loss
rec XIII
kud XII loss
PEX2 C2
EUG1 K2
kud III loss
rec3 VII
rec2 XI
kud X loss
rec2 VII
kud VIII loss
kud XIV loss
rec XV
kud XI loss
kud XII loss
kud XV loss
EPL1 K2

Figure 2 Maximum parsimony tree indicating the minimum number of chromosomal rearrangements and restriction site changes (presence/absence matrix is given in Table S1) necessary to connect the different genotypes exhibited by the *S. cerevisiae* × *S. kudriavzevii* hybrids to a putative hybrid ancestor. This putative ancestor is not necessarily the same for all lineages, it just corresponds to an ancestral state containing the complete *S. cerevisiae* and *S. kudriavzevii* genomes, but it could be generated several times from different parental strains, as discussed in the main text. Genotypes are represented by white and gray circles for wine and brewing hybrids, respectively. Rearrangements are indicated by arrows giving the direction of the irreversible change and were treated under the Camin-Sokal criterion. Rearrangements were assumed to be caused by nonreciprocal recombination (rec) among homoeologous chromosomes (roman numbers) and whole chromosome losses (loss) of one of the parental chromosomes (*kud*, *S. kudriavzevii*). Restriction site changes can be reversible (gains/losses represented by diamonds) and were treated under the Wagner criterion. The gene region and the restriction patterns involved are also indicated (for a description see references González *et al.*, 2008 and Peris *et al.*, 2012a)

The rest of the allotriploid hybrids, isolated from brewing, and the wine allotetraploid PB7 and AMH strains, appeared in separated branches with strain-specific chromosomal rearrangements. The only exception is the shared loss of *S. kudriavzevii* chr. XII between the partial allotetraploid AMH and the allotriploid CECT 11002, which can be considered a convergent event. PB7 also shared similar restriction alleles with Group II but this strain is also allotetraploid (Table 2).

This most parsimonious tree shows several convergent events, such as chromosomal losses, chromosomal rearrangements and restriction site changes (evidencing different allelic variants). *S. kudriavzevii* chr. I seems to have been lost independently in hybrids SPG 441, CECT 11011, and AMH. In a similar way, the lack of chr. V in hybrids HA 1841 and AMH, and chromosome XIV in CECT 1990 and AMH seem to be independent events according to this parsimony analysis.

Convergent events involving recombinant chromosomes were also found. This is the case of the type 2 recombination in chr. IV (shared by AMH and SOY3), type 1 recombination in chr. VII (shared by CECT 11002, CECT 11011, 1388, W46, CECT 11003 and CECT 11004), type 2 recombination in chr. XIV (CECT 1388 and CECT 11002) and the recombinant chr. XV (AMH and Group I hybrids). This could be indicative of the presence of recombination hotspots in the *Saccharomyces* genomes.

4. Discussion

4.1 The genome diversity in S. cerevisiae *x* S. kudriavzevii *hybrids*

The genome composition of 11 new wine and brewing *S. cerevisiae* x *S. kudriavzevii* hybrid strains was described in this work by means of aCGH analysis. Additionally, a comparison between them and other four wine hybrids already described by (Belloch *et al.*, 2009) was also performed. Individual and differential chromosomal composition patterns were found for each particular strain, except for brewing strains CECT 11003 and CECT 11004 which appear closely related to the previously described Swiss wine hybrids (Belloch *et al.*, 2009). The close relationships between wine hybrid strains from Switzerland and the brewing strains CECT 11003 and 11004 was already observed in a previous study based on PCR-RFLP analysis of hybrids as well as in the phylogenetic reconstruction based on COX2 sequences (Peris *et al.*, 2012a). In that work, a recombination in chromosome XV was proposed as the unique difference between strains 11003 and 11004; however, aCGH analysis carried out in this study demonstrated that this recombination is present in both strains (Figure 2). These Swiss wine hybrids were previously described as diploids (Belloch *et al.*, 2009) on the basis of ploidy estimations with propidium iodide. However, in the reanalysis of ploidy with SYTOX Green, they also resulted to be allotriploids as CECT 11003 and CECT 11004.

Flow cytometry results with SYTOX Green were in accordance with genome structure deduced from aCGH analysis carried out in this work and with the presence/absence of parental genes deduced from a previous PCR-RFLP analysis of hybrids (Peris *et al.*, 2012a). Most *S. cerevisiae* × *S. kudriavzevii* hybrid strains were allotriploids, with the exception of AMH and PB7 which were allotetraploids. Some aneuploidies were also found in several hybrids. Aneuploidies seem to be common in *Saccharomyces*

hybrids since this phenomenon have also been observed in *S. cerevisiae* × *S. bayanus* hybrids (Dunn and Sherlock, 2008; Nakao *et al.*, 2009). The role of aneuploidies in the hybrid genomes is not clear, but their presence in *S. cerevisiae* affected both the transcriptome and proteome, generating significant phenotypic variation and bringing fitness gains under diverse conditions (Pavelka *et al.*, 2010).

Recently, the hybrid genome of VIN7, one the hybrids analyzed in the present study, has completely been sequenced (Borneman *et al.*, 2012), concluding that this strain is an almost perfect allotriploid hybrid that contains a heterozygous diploid *S. cerevisiae* genome and a haploid *S. kudriavzevii* genome. The genome constitution of VIN7 deduced from the sequencing analysis is basically similar to the one inferred by aCGH in the present study, but there are some differences. The genome sequence analysis detected a homeologous recombination generating a chimerical chromosome VII, a genomic substitution of a region of 15 kb, of *S. kudriavzevii* genomic DNA from chromosome IV by the orthologous sequences from *S. cerevisiae* and a genomic substitution of a 13 kb region of *S. cerevisiae* genomic DNA from chromosome IV by *S. kudriavzevii* sequences combined with homeologous recombination between the *S. kudriavzevii* and *S. cerevisiae* alleles. The first rearrangement involving a chimerical chromosome VII was clearly detected in the aCGH analysis, but not the two genomic substitutions. Both genomic substitution involve short segmental replacements of a few genes (7 and 8), and the second an almost reciprocal recombination between homeologous chromosomes that cannot be observed by aCGH analysis. However, the presence/absence analysis of parental genes in hybrids (Peris *et al.*, 2012a) detected the loss of *S. kudriavzevii* chromosome III in our VIN7. As an ongoing project, our group is also sequencing the whole genome of several *S. cerevisae* × *S. kudriavzevii* hybrids, including the commercial VIN7 yeast. We checked in

the preliminary sequencing of our VIN7 strain for the presence of *S. kudriavzevii* chromosome III sequences and the result was negative, confirming our aCGH results and indicating that our VIN7 strain is different. These differences may be due to the fact that our VIN7 strain was isolated from a commercial dry yeast sample provided by Anchor Yeast but Borneman *et al.,* (2012) sequenced the original mother culture of VIN7, as they mention in their acknowledgements. Therefore, the continuous propagation of this yeast in molasses under aerobic conditions to obtain commercial dry yeasts may have promoted a new chromosomal rearrangement, the loss of the *S. kudriavzevii* chromosome III.

Taking into consideration the ploidy data as well the fact that most hybrids possess either trisomic (2 *S. cerevisiae* chromosomes: 1 *S. kudriavzevii* chromosome) or tetrasomic chromosomes (2 *S. cerevisiae* chromosomes: 2 *S.kudriavzevii* chromosomes), two scenarios on the hybridization process are plausible. In the case of allotriploid hybrids, the simplest explanations for their origins are hybridization events by rare-mating between a diploid cell of *S. cerevisiae* and a haploid cell or spore of *S. kudriavzevii*. This is also supported by the genome sequencing of VIN7, one of the allotriploid strains, which resulted to contain heterozygous diploid genome from *S. cerevisiae* and a haploid genome from *S. kudriavzevii* (Borneman *et al.*, 2012).

On the other hand, diploid and diploid cell rare-mating between *S. cerevisiae* and *S. kudriavzevii* should be invoked to explain the origin of allotetraploid hybrids. In the case of PB7 it was observed high spore viability (95%) due to the presence of the two chromosomes copies of each parental strain.

Rare-mating between diploid cells was already proposed as a probable mechanism for hybrids generation (Sipiczki, 2008; Belloch *et al.*, 2009). However, haploid cell or spore mating between *S. cerevisiae* and *S.*

kudriavzevii, followed by a whole genome duplications due to endoreplication or chromosome duplications due to non-disjunction, and subsequent chromosomal rearrangements, although less plausible, cannot totally be discarded.

4.2 Characterization of the S. kudriavzevii *subgenome from hybrids*

According to Sipiczki (2008), genomes from each parental species interact in the new hybrid genome. This interaction can be observed in the loss of large parts of one or both genomes as well as in the presence of chimerical chromosomes that make the hybrid genome as stable as possible to future genetic modifications. Additionally, adaptive evolution of these hybrid genomes under fermentative environmental conditions could make hybrid genome to conserve the chromosomes, or part of them, which grant a selective advantage (Barrio *et al.*, 2006). According to the results obtained in this work as well as in our previous studies (González *et al.*, 2008; Belloch *et al.*, 2009; Peris *et al.*, 2012a), *S. cerevisiae* × *S. kudriavzevii* hybrids seem to have the common trend to lose the *S. kudriavevii* parental chromosomes maintaining the *S. cerevisiae* ones. The reduction of the non-*S. cerevisiae* genome observed in both wine and brewing *S. cerevisiae* × *S. kudriavzevii* hybrids was already reported for artificial *S. cerevisiae* × *S. uvarum* hybrids genetically stabilized by successive sporulation steps (Antunovics *et al.*, 2005). In contrast, *S. pastorianus* (*S. cerevisiae* × *S. eubayanus* hybrids) Group 1 strains obtained from different brewing processes and studied by aCGH analysis, showed a trend to lose the *S. cerevisiae* genome fraction (Dunn and Sherlock, 2008). The cause of the predominance of one or the other parental genome in the hybrids remains unclear yet. However, selective pressures acting under harsh environmental conditions and cytonuclear interactions have been suggested as the main factors affecting the genome

conformation of hybrids. In *S. cerevisiae* × *S. eubayanus* lager strains, supposed to be naturally selected after years of use in brewing, the predominance of a *S. eubayanus*-like genome has been related to the maintenance of the *S. eubayanus* mitochondria (Dunn and Sherlock, 2008; Rainieri *et al.*, 2008). However, artificial hybrids constructed from the same two parental species, but without selective pressures, inherited their mitochondrial genome from either one or the other parental species randomly (Rainieri *et al.*, 2008; Solieri *et al.*, 2008). The conservation of the mitochondrial genome from the parental species most represented in the nuclear genome was also observed in the stable artificial *S. cerevisiae* × *S. uvarum* hybrids, which maintained the mitochondrial genome of the *S. cerevisiae* parental strain (Antunovics *et al.*, 2005). All *S. cerevisiae* × *S. kudriavzevii* natural hybrids analyzed in this work, except for AMH, maintained a *S. kudriavzevii* mitochondrial genome (González *et al.*, 2008; Peris *et al.*, 2012a). However, *S. cerevisiae* × *S. kudriavzevii* artificial hybrids, randomly inherited the *S. cerevisiae* or the *S. kudriavzevii* mitochondrial DNA (Pérez-Través *et al.* personal communication). This discrepancy between the mtDNA inheritance in artificial vs. natural hybrids has been associated with the result of an unwitting human-driven selection of naturally generated hybrid strains for fermentations at low temperature (Rainieri *et al.*, 2008). A common origin for all hybrids could be another possible explanation, but the present analysis of the genome constitutions in hybrids suggests diverse origins.

Interestingly, the hybrid AMH, which maintained the *S. cerevisiae* mitochondria, has lost a 69% of the nuclear genes of *S. kudriavzevii* coding for proteins with functions associated to the mitochondria; while the rest of the analyzed hybrids with *S. kudriavzevii* mitochondria have lost only 0.67%–42.48% of the *S. kudriavzevii* genes related to mitochondrial functions. Due to the fact that a number of mitochondrial proteins encoded

in the nuclear genome play an important role in the mtDNA replication and transmission, both the type of mitochondrial DNAs and the functions of the mitochondria in a hybrid strain are clearly under the control of the nuclear genome (Vero *et al.*, 2003). One of the most interesting evidence about nuclear-mitochondrial genome interactions were described by Lee *et al.,* (Lee *et al.*, 2008), who demonstrated that the presence of the *S. bayanus* nuclear gene *AEP2* together with the *S. cerevisiae* mitochondrial gene *OLI1* cause a cytonuclear incompatibility. More recently, Chou *et al.,* (2010) identified other two genes, *MRS1* and *AIM22*, associated with cytonuclear incompatibility among *S. cerevisiae*, *S. paradoxus* and *S. bayanus*. A similar behavior involving the same or other different genes in *S. cerevisiae* × *S. kudriavzevii* hybrids was not yet demonstrated.

aCGH and GO analysis carried out with those *S. kudriavzevii* genes conserved in all *S. cerevisiae* × *S. kudriavzevii* hybrids with *S. kudriavzevii* mitochondria (excluding AMH) evidenced a significant enrichment in nuclear genes related to mitochondrial function (a total of 328 genes) supporting the hypothesis of a necessary interaction between the *S. kudriavzevii* nuclear-encoded proteins and the mitochondrial genomes or their products. Taking into consideration that a total of 751 proteins encoded by the nuclear genome are associated with the mitochondrial function in *S. cerevisiae* (Sickmann *et al.*, 2003), and considering a similar number in *S. kudriavzevii*, we can assume that the remaining genes up to 751 might be non-essential for the maintenance of the *S. kudriavzevii* mitochondria in hybrids. In particular the *S. kudriavzevii* gene *AEP2* reported by Lee *et al.,* (2008) was not common to all analyzed hybrids, indicating that different incompatible nuclear-mitochondrial pair of genes could be associated with each particular pair of *Saccharomyces* parental species involved in hybrid generation.

GO analysis was also very informative with regards to the conservation in hybrids of particular groups of genes, inherited from each parental species that may be potentially related to adaptive advantage for fermentation at low temperatures. A significant overrepresentation of *S. kudriavzevii* genes associated with the physiological adaptation of yeasts to grow at low temperatures, such as fatty acid transport and N-glycosilation of proteins in all hybrids, and ergosterol biosynthesis in the case of wine hybrids (Higgins *et al.*, 2003; Beltrán *et al.*, 2006; Aguilera *et al.*, 2007) was observed (Table 4). Changes in membrane fluidity are the primary signal triggering the cold shock response (Aguilera *et al.*, 2007). This response involves certain groups of genes: members of the *DAN*/*TIR* family of cell-wall mannoproteins, genes coding for temperature inducible protein (*TIP1*) and seripauperins (*PAU*), genes related to ergosterol and phospholipid synthesis (*ERG*, *INO1* and *OPI3*) and the gene coding for the only known desaturase in *S. cerevisiae* (*OLE1*), among others (Aguilera *et al.*, 2007). These sets of genes are present in the *S. kudriavzevii* subgenome of all hybrids analyzed in this work, with some exceptions mainly involving AMH (Table 4 and Table S4).

Our results are in agreement with results about stress tolerance, including adaptation to low temperatures, previously obtained in our laboratory using some of the *S. cerevisiae* × *S. kudriavzevii* hybrids analyzed in this work (Belloch *et al.*, 2008; Arroyo-López *et al.*, 2009). Physiological implications of possessing *S. kudriavzevii* genes in those particular functions or metabolic pathways must be elucidated in future studies involving both transcriptomic and metabolomic analyses.

4.3 Wine yeast signatures in the S. cerevisiae *subgenome from hybrids*

An interesting result obtained from aCGH analysis was the detection of a common set of *S. cerevisiae* genes that are in lower copies in the

genome of all *S. cerevisiae* × *S. kudriavzevii* hybrids (Table S2). This finding might indicate that the *S. cerevisiae* parental strains involved in the different hybridization events shared a similar genetic background and were closely related yeasts.

Using a similar methodology, a trend to loss some particular set of genes in *S. cerevisiae* wine strains, with regards to strains belonging to the same species but isolated from different sources, was previously demonstrated (Dunn *et al.*, 2005; Carreto *et al.*, 2008). Dunn *et al.,* (2005) proposed the term “commercial wine yeast signature” to refer to this set of genes. Most of these genes that are frequently depleted in wine strains are also depleted in the *S. cerevisiae* fraction of the hybrid genomes of all hybrids. This finding supports the hypothesis that these hybrids have likely been generated from wine *S. cerevisiae* parental strains.

4.4 On the origin of hybrids

The maximum parsimony analysis of the relationships between the wine and beer hybrids are congruent with diverse origins for the strains according to chromosomal rearrangement differences, mainly due to the presence of chimerical chromosomes, and *S. kudriavzevii* chromosome losses, in some cases compensated by the presence of an extra copy of the homeologous *S. cerevisiae* chromosome (Figure 2).

While the brewing strains seem to represent different and divergent lines (except strains CECT11003 and 11004), most wine hybrids clustered in two main groups of strains sharing common events, with the exception of AMH and PB7 that were independently originated. Brewing strains CECT 11003 and 11004 shared the same genome than wine hybrid W46 probably evidencing that either an original strain with this common genome structure was introduced in both fermentative processes, or colonize one fermentative process from the other. The parsimony tree obtained in this

study is congruent with previous phylogenetic reconstructions of hybrids based on COX2 sequences (Peris *et al.*, 2012a).

The occurrence of several chimerical chromosomes sharing similar—if not the same—recombination points, common to some *S. cerevisiae* × *S. kudriavzevii* hybrids located in different branches of the parsimony tree, indicates the presence of recombination hot spots. Recombination between homeologous chromosomes are probably mediated by highly recombining regions located in the recombination sites, such as ARS sequences (Di Rienzi *et al.*, 2009), Ty elements (Kim *et al.*, 1998), Y' elements, rRNA regions and conserved coding genes (Belloch *et al.*, 2009; Pérez-Ortín *et al.*, 2002). When recombination is initiated in a region with high homology, the mismatch repair system (MMR) stimulates the loss of one partner of the recombination event in the hybrids and the fixation of the other, thus generating a chimerical recombinant chromosome. With the exception of the almost perfect allotetraploid PB7, hybrids have low spore viability (<1%) indicating that they are maintained by mitotic budding. Therefore, mitotic homeologous recombination, although much less frequent than meiotic, may also explain the generation of chimerical chromosomes.

The genome composition of hybrids reveals that the ancestral hybrid strains were allotriploid or allotetraploid, resulting from rare mating between diploid *S. cerevisiae* and haploid or diploid *S. kudriavzevii* (Barros Lopes *et al.*, 2002; Borneman *et al.*, 2012). The presence of triple hybrids also supports this hypothesis (González *et al.*, 2006; Peris *et al.*, 2012a). Finally, the presence of *S. kudriavzevii* alleles shared between most hybrids and the European *S. kudriavzevii* population (Lopes *et al.*, 2010), as well as the presence of the gene *GAL4* from *S. kudriavzevii* (González *et al.*, 2008; Peris *et al.*, 2012a), which is a functional gene in the European populations of S. kudriavzevii but a pseudogene in the Japanese strains (Hittinger *et al.*,

2010), indicate that these hybrids were originated from a European *S. kudriavzevii* parental strain.

5. Conclusions

Hybridization between *S. cerevisiae* and *S. kudriavzevii* have occurred several times by rare-mating between different wine *S. cerevisiae* diploid and European *S. kudriavzevii* haploid or diploid progenitors. After hybridization, the hybrid genome suffered random genomic rearrangements mediated by crossing-over between homologous chromosomes and non-disjunction, promoting the loss of variable fractions of the parental subgenomes. Both the restrictions imposed by interactions between both parental genomes as well as between nuclear and mitochondrial genomes, together with the selective environmental conditions prevailing during fermentation modulated the final composition of the hybrid genomes, characterized by the maintaining of the *S. cerevisiae* genome and the progressive reduction of the *S. kudriavzevii* contribution.

Authors' contributions

This study is the result of the collaboration between AQ and EB laboratories. CB, AQ and EB conceived and supervised this study. DP, CB, AQ and EB designed the experiments. DP and CL performed the experimental work and data analyses. DP and CB wrote the first version of the manuscript. CL, AQ and EB participated in the final manuscript revision.

Aknowledgements and funding

This work was supported by Spanish Governments projects AGL2009-12673-CO2-01 and AGL2009-12673-CO2-02 to AQ and EB respectively, and by grant PROMETEO/2009/019 from Generalitat Valenciana to AQ, EB and CB. DP and CL acknowledges to the Spanish

Government for a FPI fellowship and a postdoctoral contract, respectively. We also thank Ksenija Lopandić, José Manuel Álvarez-Pérez, Sandi Orlić, Lallemand Bio and Anchor Wine Yeasts for providing yeast strains.

<u>OBJECTIVE 3.</u> Study of the origin of natural *Saccharomyces cerevisiae* x *Saccharomyces kudriavzevii* hybrids.

Chapter 1

Reconstruction of the evolutionary history of *Saccharomyces cerevisiae* x *S. kudriavzevii* hybrids based on multilocus sequence analysis.

David Peris[1], Christian A. Lopes[2,3], Armando Arias[1,4] and Eladio Barrio[2,5*]

[1]Institut "Cavanilles" de Biodiversitat i Biologia Evolutiva, Universitat de València, Spain

[2]Departamento de Biotecnología, Instituto de Agroquímica y Tecnología de Alimentos, CSIC, Valencia, Spain.

[3]Grupo de Biodiversidad y Biotecnología de Levaduras. Instituto Multidisciplinario de Investigación y Desarrollo de la Patagonia Norte (CONICET-UNCo). Facultades de Ingeniería y Ciencias Agrarias, Universidad Nacional del Comahue, Neuquén, Argentina.

[4]Laboratorio de Biotecnología, Unidad de Botánica y Zoología, CUCBA, Universidad de Guadalajara, Zapopan, Jalisco, Mexico.

[5]Departament de Genètica, Universitat de València, Spain.

*Corresponding author:

Dr. Eladio Barrio, Instituto de Agroquímica y Tecnología de Alimentos, CSIC. Lab 309. Avda. Agustín Escardino, 7. E-46980 Paterna, Valencia, Spain. Phone: +34 963 900 022 ext. 2309, Fax: +34 963 636 301.

E-mails: DP: david.peris-navarro@uv.es, CL: clopes@conicet.gov.ar, AA: aarias@cucba.udg.mx, EB: eladio.barrio@uv.es

Plos ONE (Accepted)

Abstract

In recent years, interspecific hybridization and introgression are increasingly recognized as significant events in the evolution of *Saccharomyces* yeasts. These mechanisms have probably been involved in the origin of novel yeast genotypes and phenotypes, which in due course were to colonize and predominate in the new fermentative environments created by human manipulation. The particular conditions in which hybrids arose are still unknown, as well as the number of possible hybridization events that generated the whole set of natural hybrids described in the literature during recent years.

In this study, we could infer at least six different hybridization events that originated a set of 26 *S. cerevisiae* x *S. kudriavzevii* hybrids isolated from both fermentative and non-fermentative environments. Different wine *S. cerevisiae* strains and European *S. kudriavzevii* strains were probably involved in the hybridization events according to gene sequence information, as well as from previous data on their genome composition and ploidy.

Finally, we postulate that these hybrids may have originated after the introduction of vine growing and winemaking practices by the Romans to the present Northern vine-growing limits and spread during the expansion of improved viticulture and enology practices occurred during the Late Middle Ages.

Keywords: *Saccharomyces*, interspecific hybridization, *S. cerevisiae*, *S. kudriavzevii,* phylogenetic supernetworks, multigene sequence analysis, hybridization origins.

1. Introduction

The first evidence of production of fermented beverages dates back to 7000 BC in the Neolithic village of Jiahu in China (McGovern *et al.,*, 2004), but the earliest evidence of winemaking is traced to Iran at the Hajji Firuz Tepe site (5400-5000 BC) (This *et al.,*, 2006). From these origins in the slopes of northern Zagros, eastern Taurus and Caucasus Mountains, vineyards and grape wine production gradually spread to adjacent regions of the Fertile Crescent such as Mesopotamia and the Jordan Valley, and beyond, to the Eastern Mediterranean regions of Egypt, Phoenicia, Crete and Greece (5000 BC). Colonization by the Phoenicians, Carthaginians and Greek spread winemaking far across the Western Mediterranean regions of Southern Europe and Northern Africa. By 500 BC, wine was being produced in Italy, Sicily, Southern France, the Iberian Peninsula and the Maghreb. Vine cultivation was later extended by the Romans to the Northern limits of their empire (100 BC-100 AD). The next important expansion of winemaking was during the European colonization of America (16th century), South Africa (17th century), and Australia and New Zealand (18-19th centuries) (Pretorius, 2000; McGovern, 2010).

On the other hand, beer elaboration is first recorded in the Mesopotamian region and in Egypt. Brewing diverged into two processes mainly differentiated by the prevailing fermentation temperature: ale, acquired from the Middle East by Germanic and Celtic tribes around the 1st century AD, and lager, which appeared during the Late Middle Ages in Europe (Corran, 1975; Sicard & Legras, 2011).

A fortuitous domestication that acted on the *S. cerevisiae* populations is associated with wine and beer elaboration: it occurred as a consequence of the expansion of these fermentation processes. The first genetic diversity characterization of *S. cerevisiae* strains, isolated from different sources, showed clear differences between wild and domesticated strains (Fay &

Benavides, 2005). Another study (Arias, 2008) evaluated the genetic variability of ~250 *S. cerevisiae* strains based on four nuclear gene sequences, and revealed for some genes the presence of two groups of alleles that differentiated wine strains from those isolated from other, non-wine, sources. Liti *et al.,* (2009) performed a genetic-population analysis based on whole genome sequences of 36 *S. cerevisiae* strains and reported the presence of five 'clean' (pure) lineages and different 'mosaic' (recombinant) strains. One of the 'clean' genotypic lineages comprises a number of wine strains from different geographic origins as well as European non-wine strains, and therefore, it was called wine/European population, the other lineages corresponded to strains isolated from other sources and origins (Liti *et al.,*, 2009).

In recent years, hybrids between *S. cerevisiae* and other *Saccharomyces* species such as the cryotolerant *S. uvarum* (Masneuf *et al.,*, 1998; Naumov *et al.,*, 2000; Le Jeune *et al.,*, 2007) and *S. kudriavzevii* (Bradbury *et al.,*, 2006; González, 2006; Gonzalez *et al.,*, 2008; Lopandic *et al.,*, 2007; Peris *et al.,*, 2012a) have been isolated from wine, cider and brewing fermentations, and other sources. These discoveries suggest that hybridization between different *Saccharomyces* species has been a frequent phenomenon in their evolution, particularly relevant during the adaption of *Saccharomyces* to fermentative conditions (Gonzalez *et al.,*, 2007; Belloch *et al.,*, 2008; Gangl *et al.,*, 2009). Some hybrids can be predominant even in the most Northern winemaking regions from Europe, very likely due to a better adaptation to growth at lower temperatures acquired from the non-*cerevisiae* parental, compared to *S. cerevisiae* (Gonzalez *et al.,*, 2007; Lopandic *et al.,*, 2007; Belloch *et al.,*, 2008; Erny *et al.,*, 2012).

Some reports carried out on a set of wine and beer *S. cerevisiae* x *S. kudriavzevii* hybrid strains suggested that those hybrids could be generated from hybridization between wine strains of *S. cerevisiae* and natural European

strains of *S. kudriavzevii*; however, those results were not completely conclusive (Sampaio & Gonçalves, 2008; Lopes *et al.,*, 2010; Peris *et al.,*, 2012b). The aim of this study was to evaluate, by means of a multigenic sequence approach, the potential origin of 24 *S. cerevisiae* x *S. kudriavzevii* and 2 *S. cerevisiae* x *S. kudriavzevii* x *S. uvarum* hybrid strains obtained from wine, beer and two other non-fermentative sources. The possible number of hybridization events that gave origin to the complete set of hybrids was also proposed based on the results obtained in this work and in previously reported data.

2. Material and methods

2.1 Saccharomyces *strains, culture media and nucleotide sequences*

Twenty-six *S. cerevisiae* x *S. kudriavzevii* hybrid strains from different origins (Table S1) and seven strains belonging to *S. kudriavzevii* species (Table S2) were used in this study. Yeasts were grown at 28ºC in GPY medium (2% glucose, 0.5% peptone, 0.5% yeast extract).

Nucleotide sequences corresponding to representative *S. cerevisiae* wine and non-wine alleles according to Arias (Arias, 2008) for genes *BRE5*, *CAT8*, *EGT2* and *GAL4* were also included in this study (Table S3 and Table S4).

Sequences for genes *BRE5*, *CAT8*, *CYC3*, *CYR1*, *EGT2*, *CAT8*, *GAL4* and *MET6* from *S. cerevisiae* strains (Table S2) representative of each pure population defined by Liti *et al.,* (2009) were obtained from SGRP (*Saccharomyces* Genome Resequencing Project, version 2 assemblies (20x coverage), except for strain RM11, which corresponded to version 1 (ftp://ftp.sanger.ac.uk/pub/dmc/yeast/SGRP2/assembly/). In addition, sequences from wine strain EC1118 (Novo *et al.,*, 2009) were retrieved from GenBank database. Finally, *S. kudriavzevii* ZP591 and IFO 1802 sequences

were downloaded from the *Saccharomyces sensu stricto* database (www.*Saccharomyces*SensuStricto.org).

2.2 PCR amplification and sequencing

DNA was extracted following the procedure described by Querol *et al.,* (1992). Genes *BRE5, CAT8, CYC3, CYR1*, *EGT2* and *GAL4* were amplified by PCR, using primers CAT8_3, CYR1_5, MET6_5, MET6_3, MET6_3K from González *et al.,* (2008) and newly designed primers (Table S5), obtained from the comparison among sequences from strains *S. cerevisiae* S288C and *S. kudriavzevii* IFO 1802 and ZP591.

Most primers were species-specific with the exception of those for genes *CAT8*, *EGT2* and *GAL4*. The analysis of these genes required a previous step of cloning, performed by using a TOPO XL PCR Cloning Kit (Invitrogen). To detect the *S. cerevisiae* alleles in clones, a screening was carried out by colony-PCR with the corresponding primers, and a subsequent digestion of the PCR fragments following the procedure described in González *et al.,* (2008).

PCR amplifications were performed by using conditions described in González *et al.,* (2008) in a G-Storm Thermocycler (G-Storm Ltd, UK). Amplification products were cleaned with a High Pure PCR Product Purification Kit (Roche Diagnostics, Mannheim, Germany) and both strands of the DNA were directly sequenced using the BigDyeTM Terminator V3.0 Cycle Sequencing Kit (Applied Biosystems, Warrington, UK), following the manufacturer's instructions in an Applied Biosystems automatic DNA sequencer Model ABI 3730l (Applied Biosystems). Sequences were edited and assembled with Staden Package v1.5 (Staden *et al.*,, 2000) to be deposited in GenBank under accession numbers JN709116 to JN709440.

2.3 Haplotype and haplogroup classification

Gene sequences were aligned in MEGA 5 (Tamura *et al.,*, 2011). Haplotype classification was done in DnaSP v5 (Librado & Rozas, 2009) using the previous haplotype number classification given by Arias (2008). New haplotypes were classified with consecutive Arabic numbers following the previous enumeration (Arias, 2008). Median joining (MJ) networks (Bandelt *et al.,*, 1999) for *BRE5*, *CAT8*, *EGT2, GAL4* were constructed using Network 4.5 (http://www.fluxus-engineering.com/).

2.4 Phylogenetic analysis and Supernetworks

The neighbor-joining (NJ) and maximum-parsimony (MP) methods of phylogenetic reconstruction were applied to *BRE5*, *CAT8, CYC3, CYR1, EGT2, GAL4* and *MET6* separate sequence alignments of *S. cerevisiae* and *S. kudriavzevii* alleles from hybrid and reference strains described in Table S2. NJ trees were obtained with nucleotide distances corrected using the Maximum Composite Likelihood method. MP trees were obtained using the Close-Neighbor-Interchange algorithm in which the initial trees were obtained with the random addition of sequences (10 replicates). In all cases, a bootstrap analysis based on 2,000 pseudo-replicates was performed. For each gene, two NJ and MP phylogenetic trees were obtained, a tree based on *S. cerevisiae* alleles and another based on *S. kudriavzevii* alleles. Phylogenetic analyses were performed with MEGA 5 (Tamura *et al.,*, 2011).

Two nexus files, with the collection of phylogenetic trees for *S. cerevisiae* and *S. kudriavzevii,* were created as an input of SPLITSTREE 4 package (Huson & Bryant, 2006). Two outputs corresponding to *S. cerevisiae* and *S. kudriavzevii* consensus super split networks (Supernetworks) were obtained, analyzing about 3.4 kb. For *S. cerevisiae* nexus file we reduced the number of splits setting maximum dimension parameter to 1, removing those splits in the network that are less supported. For the *S. kudriavzevii* nexus file

we reduced the number of splits to simplify the final Supernetwork. For this simplification we applied the filtered Z-Closure method (filtering = 2). A filter of 2 takes into account those splits that are compatible in at least 2 input trees in the nexus file. The result is a network that summarizes the relationships found in at least two trees simplifying the network (Whitfield *et al.,*, 2008).

2.5 Array competitive genomic hybridization (aCGH) and flow cytometry

Array competitive genomic hybridization (aCGH) experiments, scanning and data normalization were performed for IF6 and MR25 strains as previously described in Peris *et al.,* (2012b). A double-spotted array containing 6,240 ORFs of *S. cerevisiae* plus control spots totaling 6.4K (Microarray Centre, University Health Network, Toronto, Canada) was used in aCGH assays. Raw and normalized microarray data are available in ArrayExpress (Brazma *et al.,*, 2003), under accession number E-MEXP-3375.

Caryoscopes were obtained using ChARM v.1.1 (Myers *et al.,*, 2004). Genome composition of IF6 and MR25 was inferred by combining aCGH (present study) and previous PCR-RFLPs data (Peris*et al.,*, 2012a). aCGH was performed following the procedure described in Peris *et al.,* (2012b).

The approximate locations of the recombination points in the mosaic chromosomes were determined from the up and down jump locations in the ORFs mapping by microarray analysis of the hybrid yeast genomes. Collinearity between *S. kudriavzevii* and *S. cerevisiae* genomes (Cliften *et al.,*, 2003; Scannell *et al.,*, 2011) allowed us to deduce S*. kudriavzevii* gene content in the hybrid genomes.

The list of *S. kudriavzevii* genes, excluding those with unknown function, retained in the hybrid genomes of IF6 and MR25 were independently analyzed using YeastMine in SGD database (http://yeastmine.yeastgenome.org:8080/yeastmine/begin.do) to obtain those Gene Ontology terms enriched in them. GO terms enrichment with p-values <

0.05 were shown, after computing the Holm-Bonferroni for multiple hypothesis test correction. Significant GO terms were sorted according with their corresponding GO category.

The DNA content (C-value) of IF6 and MR25 was assessed by flow cytometry using a Beckman Coulter FC 500 (Beckman Coulter, USA) following the methodology described in Peris *et al.,* (2012b). Ploidy level was scored on the basis of the fluorescence intensity compared with the haploid *S. cerevisiae* S288c and diploid *S. cerevisiae* FY1679 reference strains.

2.6 Maximum parsimony tree of chromosomal rearrangements

A list of minimal number of chromosomal rearrangements, chromosomal losses and restriction site changes for IF6 and MR25 strains obtained in this work as well as data obtained from Belloch *et al.,* (2009) and Peris *et al.,* (2012b, b) were included in the maximum parsimony analysis. A binary matrix was constructed to codify each particular event and these data were used to generate parsimony trees using MIX program from Phylip 3.66 package (Felsenstein, 2005). For this analysis, both chromosomal rearrangements and chromosomal gain/losses were considered as irreversible events (Camin-Sokal criterion), but data obtained from PCR-RFLP or sequence analyses were considered reversible events (Wagner criterion). The consensus tree was obtained by using the majority rule in the Consense program.

This binary matrix was also used to reconstruct a Median Joining Network, using Networks 4.5 (http://www.fluxus-engineering.com/), and a NeighborNet Phylonetwork, using SPLITSTREE 4 package (Huson & Bryant, 2006).

3. Results

3.1 Phylogenetic analysis of S. cerevisiae *genes from hybrids*

Phylogenetic relationships between *S. cerevisiae* x *S. kudriavzevii* natural hybrids obtained from several origins and a set of pure strains of the two parental species were analyzed to decipher possible common origins of these hybrids.

Nucleotide sequence data for both *S. cerevisiae* and *S. kudriavzevii* alleles of seven nuclear genes (*BRE5*, *CAT8*, *CYC3, CYR1, EGT2*, *GAL4,* and *MET6*) were obtained from a total of 24 natural *S. cerevisiae* x *S. kudriavzevii* and 2 *S. cerevisiae* x *S. kudriavzevii* x *S. uvarum* hybrid strains from several origins (Table S1). In a first phylogenetic analysis, we compared the *S. cerevisiae* sequences obtained for genes *BRE5*, *CAT8*, *EGT2* and *GAL4* from hybrids and from a representative selection, at the genotypic level, of 65 wine and 19 non-wine *S. cerevisiae* strains previously analyzed in our laboratory (Table S3 and Table S4). These genes were selected because they had shown high variability among *S. cerevisiae* strains from different origins (Arias, 2008). Additionally, sequences from eight *S. cerevisiae* strains, five representative of the different “pure” lineages proposed by Liti *et al.,* (2009) and those from the completely sequenced genome of wine strain EC1118 (Novo *et al.,*, 2009) were also included in this study (Table S4).

Median-Joining networks (Figure 1) for all genes, except *GAL4*, showed two clearly differentiated groups of alleles or haplogroups. One haplogroup comprises those alleles present only in non-wine strains (so called non-wine alleles) and the second haplogroup includes alleles present in both wine and non-wine strains; however they are the only alleles exhibited by wine strains, and hence, they were called wine alleles. These wine alleles, when present in non-wine strains, are mainly found in heterozygosis with non-wine alleles. *GAL4* is the exception because non-wine alleles were clustered into two haplogroups. The first group is characterized by the presence of a common deleted region of 15 bp, and the second comprises different lineages and appears to be closer to the wine alleles than to haplogroup 1 (Figure 1D).

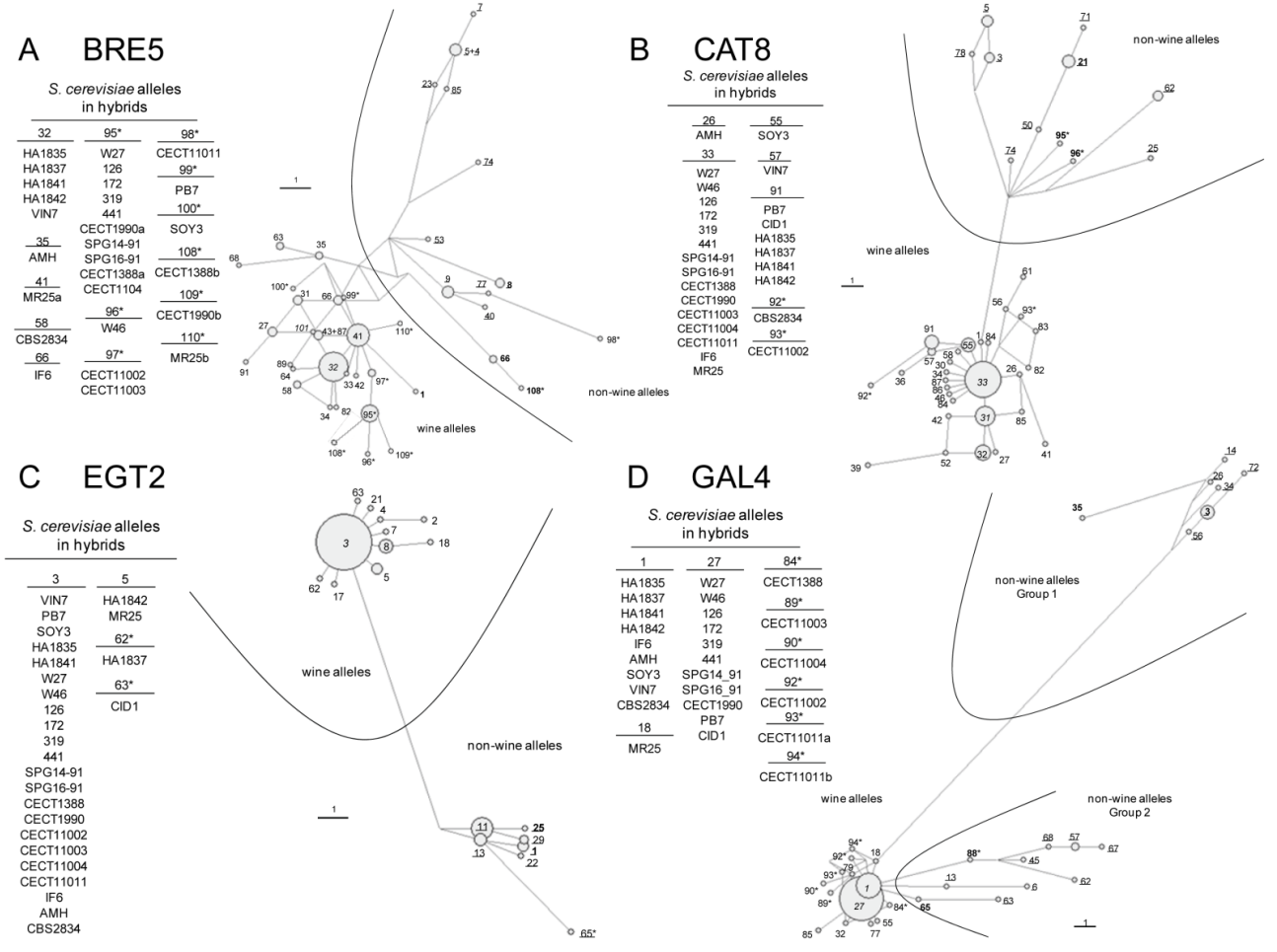

A BRE5
S. cerevisiae alleles in hybrids
32
HA1835
HA1837
HA1841
HA1842
VIN7
35
AMH
41
MR25a
58
CBS2834
66
IF6
95*
W27
126
172
319
441
CECT1990a
SPG14-91
SPG16-91
CECT1388a
CECT1104
96*
W46
97*
CECT11002
CECT11003
98*
CECT11011
99*
PB7
100*
SOY3
108*
CECT1388b
109*
CECT1990b
110*
MR25b
wine alleles
non-wine alleles
B CAT8
S. cerevisiae alleles in hybrids
26
AMH
33
W27
W46
126
172
319
441
SPG14-91
SPG16-91
CECT1388
CECT1990
CECT11003
CECT11004
CECT11011
IF6
MR25
55
SOY3
57
VIN7
91
PB7
CID1
HA1835
HA1837
HA1841
HA1842
92*
CBS2834
93*
CECT11002
non-wine alleles
wine alleles
C EGT2
S. cerevisiae alleles in hybrids
3
VIN7
PB7
SOY3
HA1835
HA1841
W27
W46
126
172
319
441
SPG14-91
SPG16-91
CECT1388
CECT1990
CECT11002
CECT11003
CECT11004
CECT11011
IF6
AMH
CBS2834
5
HA1842
MR25
62*
HA1837
63*
CID1
wine alleles
non-wine alleles
D GAL4
S. cerevisiae alleles in hybrids
1
HA1835
HA1837
HA1841
HA1842
IF6
AMH
SOY3
VIN7
CBS2834
18
MR25
27
W27
W46
126
172
319
441
SPG14_91
SPG16_91
CECT1990
PB7
CID1
84*
CECT1388
89*
CECT11003
90*
CECT11004
92*
CECT11002
93*
CECT11011a
94*
CECT11011b
non-wine alleles Group 1
wine alleles
non-wine alleles Group 2

Figure 1. Median Joining (MJ) networks obtained for genes *BRE5* (A), *CAT8* (B), *EGT2* (c) and *GAL4* (D) from hybrid strains and representative wine and non-wine allele sequences according to Arias (Arias, 2008). Strains representative of each different origin according to Liti *et al.,* (Liti *et al.,*, 2009) and the alleles from wine strain EC1118 (Novo *et al.,*, 2009) were also included. Asterisks indicate new alleles not reported by Arias (Arias, 2008). Numbers in italics indicate those alleles exhibited by wine strains from Liti *et al.,* (Liti *et al.,*, 2009) and Novo *et al.,* (Novo *et al.,*, 2009). Numbers in bold indicate alleles present in non-wine strains from Liti *et al.,* (Liti *et al.,*, 2009). Underlined numbers correspond to alleles classified as "non-wine" in Arias (Arias, 2008).

Fourteen *BRE5* alleles were present in hybrids (Figures 1 and S1), six are haplotypes already described in wine strains (32, 35, 41, 58, 66 and 95) and the other 8 were new alleles (96, 97, 98, 99, 100, 108, 109 and 110). MR25, CECT 1388 and CECT 1990 are heterozygous for this gene, exhibiting two wine *S. cerevisiae* alleles differing in one single nucleotide substitution (Figure S1). In the case of *CAT8*, 5 alleles from hybrids were present in wine yeasts (26, 33, 55, 57 and 91) and 2 were new (92 and 93). For *EGT2*, 2 alleles correspond to very common alleles in wine strains (3 and 5) and two were new (62 and 63), and finally *GAL4* showed a higher diversity in hybrids with 3 already known alleles (1, 18 and 27) and 6 new (84, 89, 90, 92, 93 and 94). These new alleles, found in hybrids for the first time, are indicated with asterisks in Figure 1. In general, alleles present in hybrids show few nucleotide differences (Figure S1) and are grouped together within the wine allele group for the four genes under analysis, with the exception of the *BRE5* new allele 98 from the brewing strain CECT11011 which is located within the non-wine haplogroup, probably due to the presence of 2 convergent nucleotide substitutions.

Strains DBVPG6044, Y12, YPS128 and UWOPS03-461.4 were selected as representative strains of the West African, Sake, North American and Malaysian pure populations of *S. cerevisiae*, respectively, as defined by Liti *et al.,* (2009). Sequences from these strains (indicated in bold in Figure 1) always clustered within the non-wine group for the four genes analyzed. To the contrary, L1528, EC1118 and RM11, three wine strains representative of the pure Wine/European genotypic lineage defined by Liti *et al.,* (2009),

always appear within the wine allele group (alleles indicated in italics in Figure 1). The laboratory strain S288c clustered within the wine (for *BRE5, CAT8* and *GAL4*) or non-wine groups (for *EGT2*) in accordance with its mosaic nature according to Liti *et al.,* (2009).

Because most *S. cerevisiae* alleles from hybrids are included within the wine allele group, the possible geographical origin of the hybrids was evaluated by analyzing the presence of these hybrid alleles in a set of 142 wine strains isolated from 8 different geographical areas, previously studied by Arias (2008). Table 1 shows the frequency of wine strains from each particular country sharing haplotypes with hybrids. The new alleles detected only in hybrids were not included in this analysis. As a general rule, the most frequent alleles in hybrids also corresponded to the most frequent alleles present in wine strains from several winemaking countries. For this reason, it is difficult to identify a specific geographic origin where hybridization processes may have occurred according to these comparisons (Table 1). Alleles 58 and 18 for *BRE5* and *GAL4* respectively were not found among the *S. cerevisiae* wine strains analyzed (Table 1), but they were detected in some non-wine strains (Table S4 and ref. (Arias, 2008)). However, these two alleles clustered within the wine allele groups (Figure 1 A and D).

To identify how many putative *S. cerevisiae* parental strains were potentially involved in the origin of *S. cerevisiae* x *S. kudriavzevii* hybrids, we increased the number of genes analyzed in a second phylogenetic analysis. For this new analysis we included sequence data previously reported by Liti *et al.,* (2009) and Novo *et al.,* (2009) for comparative purposes.

Initial phylogenetic analyses on yeast were based on single gene sequences (Kurtzman & Robnett, 1998), but several times they failed to establish the overall history of these organisms. As an improvement, multigene sequence approaches using a concatenation of genes were

Table 1: Frequency of wine strains isolated from different countries showing the same alleles found in hybrids.

Country	Total number of strains	Frequency (%) of each allele[a]														
		BRE5					*CAT8*				*EGT2*		*GAL4*			
		32*	35	41	58	66	26	33*	55	57	3*	5	1	18	27*	84
Argentina	37	49*	-	41	-	-	-	60*	3	-	81*	3	6	-	81*	3
Austria	30	30*	10	20	-	-	10	40*	-	-	60*	-	10	-	70*	-
Chile	23	23	-	41*	-	4	-	37*	4	-	81*	-	7	-	63*	-
France	13	23	-	46*	-	8	-	38*	23	-	92*	-	38	-	62*	-
Slovenia	5	20	-	-	-	-	-	20	-	-	80*	-	-	-	80*	-
South Africa	15	33*	-	-	-	-	-	33*	13	-	87*	-	33*	-	33*	-
Spain	14	21*	-	14	-	7	-	43*	14	21	93*	-	36	-	57*	-
Switzerland	5	40*	-	40*	-	-	20	-	-	-	80*	-	40*	-	40*	-

*The most frequent haplotype

[a] Only those alleles present in more than one strain were included.

proposed to construct the phylogenetic tree (Kurtzman & Robnett, 2003; Rokas *et al.,*, 2003); however, they would represent an oversimplified version of the genetic history (Huson & Bryant, 2006). As an alternative, the construction of consensus trees has also been proposed, but this method can be only used when each gene tree has the same taxa representation (Bull *et al.,*, 1993). In this work, because some hybrid strains have lost some particular *S. kudriavzevii* genes, both concatenated or consensus trees would oversimplify the results. Recently, a Z-closure method has been proposed to overcome this kind of problem (Huson *et al.,*, 2004; Huson & Bryant, 2006; Murphy *et al.,*, 2008; Whitfield *et al.,*, 2008). With this methodology, several gene trees with different taxa representation can be used as input files and a supernetwork with the complete set of taxa is obtained as (Huson *et al.,*, 2004). However, one of the limitations of the Supernetwork analysis is the absence of statistical support, for this reason we interpreted our results according to a complementary phylogenetic analysis of the individual genes based on both Maximum Parsimony and Neighbor Joining. Both methods gave very similar or identical phylogenetic reconstructions (Figures S2 and S3).

A supernetwork, containing the information of 7 *S. cerevisiae* nuclear genes (Figure 2A), showed two well defined groups of strains: a group comprising non-wine strains Y12, DBVPG6044, YPS128 and UWOPS03-461.4 and a group containing wine strains RM11, L1528, EC1118 and all hybrids (Figure 2A). The position of strain S288c in this supernetwork proved again ambiguous due to the mosaic nature of this strain.

According to this supernetwork analysis of *S. cerevisiae* gene sequences, hybrid strains appear clustered in two main subgroups (C1 and C2) and several independent lineages (Figure 2A). Subgroup C1 comprises Austrian (HA strains) and 3 other wine hybrids (PB7, SOY3 and Vin7), and the triple hybrids CID1 and CBS 2834, and subgroup C2 includes Swiss wine

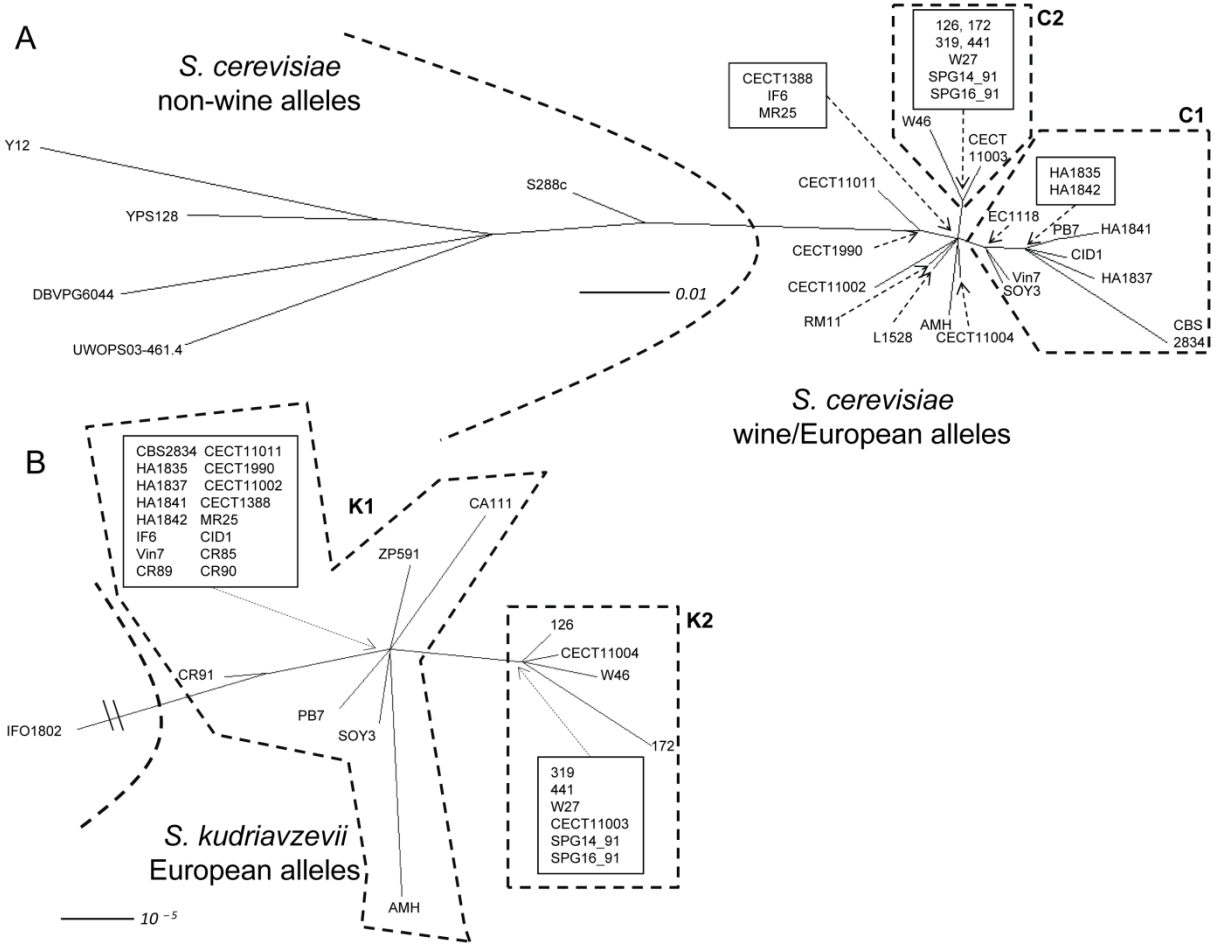

A
S. cerevisiae
non-wine alleles
Y12
YPS128
DBVPG6044
UWOPS03-461.4
S288c
0.01
C2
126, 172
319, 441
W46
SPG14_91
SPG16_91
CECT1388
IF6
MR25
C1
W46
CECT
11003
HA1835
HA1842
CECT11011
EC1118
PB7
HA1841
CECT1990
CID1
CECT11002
Vin7
SOY3
HA1837
RM11
AMH
L1528
CECT11004
CBS
2834
S. cerevisiae
wine/European alleles
B
CBS2834 CECT11011
HA1835 CECT1990
HA1837 CECT11002
HA1841 CECT1388
HA1842 MR25
IF6 CID1
Vin7 CR85
CR89 CR90
K1
CA111
ZP591
K2
126
CECT11004
W46
172
319
441
W27
CECT11003
SPG14_91
SPG16_91
CR91
IFO1802
PB7
SOY3
AMH
S. kudriavzevii
European alleles
10 −5

Figure 2. Supernetworks obtained using data from seven nuclear genes (*BRE5*, *CAT8*, *CYC3*, *CYR1*, *EGT2*, *GAL4* and *MET6*) for both *Saccharomyces cerevisiae* (A) and *Saccharomyces kudriavzevii* (B) alleles from hybrids, from reference *S. cerevisiae* (Liti *et al.,* 2009; Novo *et al.,* 2009) and *S. kudriavzevii* strains. Scale bar represents the edge's weights inferred using the tree size weighted means options, a measure similar to those from branches in a phylogenetic tree.

hybrids and Trappist beer strain CECT11003. The other hybrids appear in independent lineages (AMH, CECT 1990, 11002, 11004 and 11011) or in an ancestral position with respect to the two main subgroups (CECT1388, IF6 and MR25).

The supernetwork reconstruction method takes as input a set of complete or partial gene trees and produces a split network with the signals present in the gene trees, but it doesn't allow to test the reliability of the the phylogenetic relationships. Therefore, bootstrap analyses for each individual gene Maximum-Parsimony and Neighbor-Joining trees were performed to contrast the confidence of these groupings (Supplemental Figures S2 and S3). Three of the seven genes (*CYC3*, *CYR1* and *EGT2*) showed low variability among hybrids and were useless to differentiate hybrid subgroups, although *EGT2,* together with *CAT8,* were the best genes to discriminate among wine and non-wine alleles. The remaining genes (*BRE5, CAT8, GAL4 and MET6*) differentiate subgroups of hybrids, but due to the low variability and the presence of putative convergent nucleotide substitutions, bootstrap values were low and did not support significantly many of these groupings.

In these individual gene trees (Figures S2 and S3), strains comprised in the supernetwork subgroup C1 (Swiss double hybrids and CECT 11003) are always included in the same cluster (Figure S2, alleles in blue), however, in the case of subgroup C2 (Figure S2, alleles in yellow), only Austrian hybrids, VIN7 and CBS2834 always appeared in the same cluster. The positions of the remaining strains change from one subgroup to the other, or to independent or intermediate lineages (Figure S2, alleles in green) depending on the gene (summarized in Table S6). As examples, wine hybrid SOY3 always appears

within subgroup C1 group except for the *BRE*5 tree, where it is located in an intermediate position between wine and non-wine reference strains; W46 always appears within subgroup C2, except for *MET6* tree, in which it appears as part of subgroup C1; or CECT 1388 and 11002, which appear within subgroup 1 in two gene trees but within subgroup 2 in the other 2.

3.2 Phylogenetic analysis of S. kudriavzevii *alleles from hybrids*

Another composite supernetwork was also obtained for the sequences of the *S. kudriavzevii* alleles present in hybrids. It is important to remark that *S. cerevisiae* x *S. kudriavzevii* hybrids are characterized by a trend to lose parts of the *S. kudriavzevii* subgenome (Gonzalez *et al.,*, 2008; Belloch *et al.,*, 2009; Peris *et al.,*, 2012b, b), and hence, some of the genes under analysis are absent in some strains. The most extreme case is strain AMH, which lost ~72% of the S. kudriavzevii genome, and only maintains one of the seven genes under analysis (*CAT8*).

Homologous sequences from *S. kudriavzevii* pure strains isolated in Japan, Spain and Portugal were also included in the analysis (Table S1 and Table S2). This initial supernetwork was reconstructed without applying any filter (data not shown), however, a subsequent filtering was introduced to the analysis (see Methods section) to simplify the supernetwork analysis (Figure 2B). In this supernetwork, the European population represented by strains from Spain (CA111, CR85, CR89, CR90 and CR91) and Portugal (ZP591) forms a group far distant from the Japanese type strain IFO1802^{T} (Figure 2B). All *S. cerevisiae* x *S. kudriavzevii* hybrid strains were included within the European group. As in the case of the *S. cerevisiae* alleles, two main subgroups of hybrids are observed in this supernetwork. Subgroup K1 comprises most hybrids and occupies an ancestral position with respect to subgroup K2, including Swiss wine hybrids and Trappist beer hybrids CECT 11003 and 11004 (Figure 2B).

However, in the case of the *S. kudriavzevii* alleles, these groupings are better supported by the bootstrap analysis of Maximum-parsimony and Neighbor-Joining gene trees, even when nucleotide diversities are lower than in the case of *S. cerevisiae* alleles. In those trees based on variable genes *BRE5, CAT8, CYC3* and *CYR1*, Swiss wine hybrids and beer hybrids CECT 11003 and 11004 always appear within subgroup K2 (indicated in blue in Figure S2); and the wine hybrids from Austria (HA strains), VIN7 and SOY3 within subgroup K1 (indicated in yellow in Figure S2). In the case of hybrid IF6, this strain has lost two genes (*CAT8* and *CYC3*), but for the other genes it shares the same alleles than hybrids from subgroup K1 (Figure S1).

The positions of the remaining strains change from one subgroup to the other, or to independent positions (Figure S2, alleles in green) depending on the gene (summarized in Table S6). Thus, brewing hybrids CECT1388, 1990, and 11002, and the clinical isolate MR25 lost 1-2 genes (including the shared loss of *BRE5*). In the *CAT8* and *CYC3* trees, these strains appear within subgroup K1, but for *CYR1* they are included in a separate subgroup (indicated in green in Figure S2) due to the presence of allele 7, which differs from subgroup K1 allele 8 in a nucleotide substitution (Figure S1). Hybrid CECT11011 shares with the previous strains the *CYR1* allele 7 and their inclusion within subgroup K1 in the *CAT8* and *CYC3* trees, but within subgroup K2 in the *BRE5* tree, because maintains an allele identical to that from subgroup K2 strains. A similar situation is observed for triple hybrids CBS2834 and CID1, they appear within subgroup K2 in the *BRE5* tree but within subgroup K1 in the other gene trees, including *CYR1*. Finally, the Spanish wine hybrid PB7 appears within subgroup K1 in two gene trees (*CYC3* and *CYR1*), within subgroup K2 in other two (*BRE5* and *CAT8*), and it exhibits a different allele for *EGT2.*

3.3 Genotypes of the putative parents of hybrids based on the sequence analysis of seven nuclear genes

We tried to infer how many *S. cerevisiae* and *S. kudriavzevii* parents may have been involved in the generation of hybrids according to the phylogenetic analyses of the seven gene sequences. According to these sequences, the 24 double and 2 triple hybrids exhibit 20 different *S. cerevisiae* genotypes (allelic combinations) and 11 different *S. kudriavzevii* genotypes (Figure S1). These *S. cerevisiae* and *S. kudriavzevii* genotypes are found in 22 different combinations in hybrids. However, this does not mean that 22 different hybridization events occurred because hybrids are evolving after their origins. As seen before, the phylogenetic analysis of the sequences discriminate groups of alleles with putative common origins from an ancestral parental strain. In fact, the presence of rare alleles differing in few unique nucleotide substitutions (singletons) from the most common alleles in hybrids supports that these changes occurred after the hybridization process.

By considering the phylogenetic relationships among alleles and their combinations in hybrids (summarized in Table S6), we could infer 6 *S. cerevisiae* and 6 *S. kudriavzevii* putative ancestral genotypes (parental strains) that are arranged in 10 hybrid combinations (possible hybridization events). The first main hybrid combination is present in 6 wine hybrids, four from Austrian (HA strains), one from South Africa (VIN7, likely of European origin) and another from Croatia (SOY3). This SOY3 strain shares identical or closely related *S. cerevisiae* and *S. kudriavzevii* alleles with the other strains of this group for all genes except *BRE5*, which shows 4 nucleotide differences. This allele appears in the *BRE5* gene as closer to alleles from other hybrids (Figure 2S). These similarities could be explained by convergent substitutions, but we cannot rule out the possibility that the parental strain were originally heterozygous for *BRE5* and suffered a subsequent differential loss of heterozygosity in each derived hybrid lineage.

The second main combination is found in the 8 wine double hybrids from Switzerland and the Trappist beer hybrids CECT11003 and 11004 from Belgium. In this group, a slight discrepancy is also observed in strain CECT11004. This strain exhibits a *MET6* allele (allele 1) different to that present in other strains of this group (allele 2), but identical to that exhibited by strains from other groups (Figure S1). However, these *MET6* alleles 1 and 2 differ in one single synonymous substitution and a simple convergent change may explain this difference. An alternative explanation would be to consider allele *MET6*-1 as the ancestral one present in the *S. cerevisiae* parent of this group of hybrids later originating the derived allele *MET6*-2 shared by the Swiss and CECT11003 hybrids.

In the remaining hybrid combinations, both *S. cerevisiae* and *S. kudriavzevii* genotypes basically correspond to different arrangements of the alleles present in the first and second hybrid combinations described before. One explanation is that these recombining genotypes, generated by sexual mating at the within species level, were already present in the *S. cerevisiae* and *S. kudriavzevii* population before the hybridization events occurred. In this case, a minimum of 10 hybridization events would be necessary to explain the origin of these hybrids. However, another compatible explanation is that some hybrids may have originated by rare mating between diploid heterozygous cells, and a subsequent segregation of alleles due to chromosome loss (most hybrids are triploid (Erny *et al.,*, 2012; Peris *et al.,*, 2012b), or random loss of heterozygosity due to recombination and/or gene conversion would generate the different mosaic hybrids. In this case, the number of hybridization events would be smaller than ten. This could be the case of brewing strains CECT1388, 1990, 11002, 11011 and the clinical isolate MR25. These strains exhibit similar *S. kudriavzevii* genotypes (including the specific allele *CYR1*-7), but different *S. cerevisiae* allele combinations, including wine and non-wine alleles (*CYR1*-2 and -4 in strains CECT1990 and 11011).

3.4 The genome constitution of non-fermentative hybrids IF6 and MR25

In previous studies, we analyzed the genetic diversity of *S. cerevisiae x S. kudriavzevii* hybrids by RFLP analysis of 35 nuclear genes (Gonzalez *et al.,*, 2008; Peris *et al.,*, 2012a) combined with array comparative genome hybridization (aCGH) (Belloch *et al.,*, 2009; Peris *et al.,*, 2012b). These analyses provided us information on the genome rearrangements occurred in the hybrids after their origins. Most of these rearrangements are non-reversible events that can complement the information obtained with the phylogenetic analysis of gene sequences to unveil the origin and evolution of these *S. cerevisiae* x *S. kudriavzevii* hybrids.

However, the genome constitutions of hybrids IF6 and MR25 were not characterized in our previous studies, and therefore, they were subjected to aCGH and flow cytometry analyses to assess their genome compositions. Our results indicated that DNA content of IF6 and MR25 were 3.25 and 2.92 times that of the reference haploid strain S288c, respectively. These DNA content values, together with the aCGH analysis and PCR-RFLP data for 35 nuclear genes previously reported (Peris *et al.,*, 2012a), allowed us to detect the presence of three chimerical chromosomes in hybrid IF6 (chr. X, XII and XIII) and five in MR25 (chr. IV, VII, IX, XII and XIV) (Figure 3). The hypothetical recombination points were mapped according to the *Saccharomyces* genome described in the SGD database (http://db.yeastgenome.org) using a window size of 15-20Kb (four genes in the left and right of the most plausible recombination point). These recombination points were located in sequences corresponding to Ty LTRs, ARS and tRNAs (Table S7). RFLP analysis of genes located at the end of chromosomes (Peris *et al.,* 2012a) confirmed the presence of *S. kudriavzevii* segments in chromosomes VII and IX from IF6,

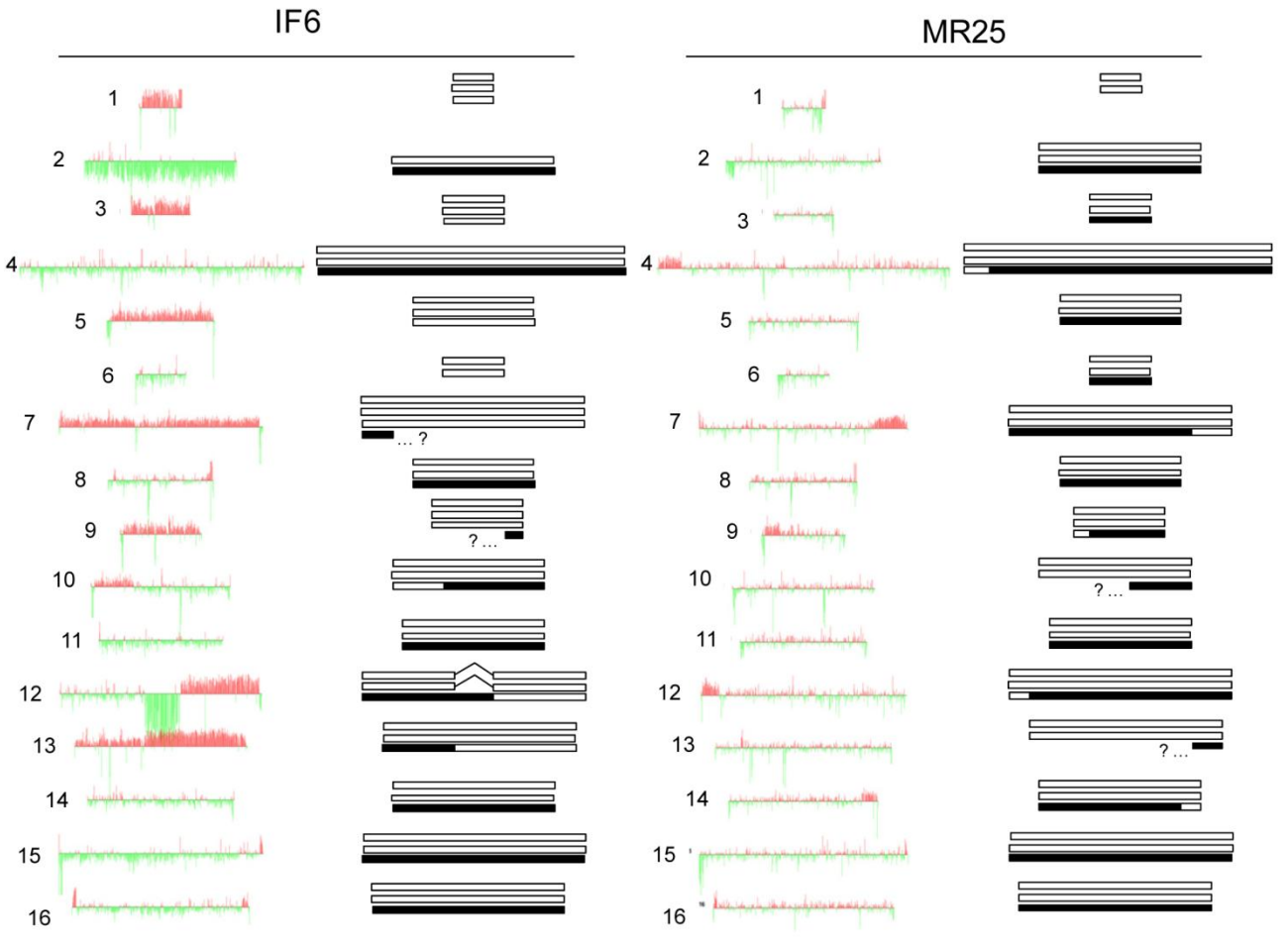

IF6
MR25
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
... ?
? ...
? ...
? ...

Figure 3. Genome composition of hybrid strains IF6 and MR25 obtained by combining aCGH (this work) and PCR-RFLP (Peris *et al.,* 2012a) analyses. Red and green signals correspond to the hybrid strain and the reference strain (S288c), respectively. White and black bars are used to represent *S. cerevisiae* and *S. kudriavzevii* fractions, respectively. Chromosomes showing black and white sections correspond to chimerical chromosomes. As an example, chromosome XIV in MR25 displayed a double RFLP pattern for *EGT2*, corresponding to the *S. cerevisiae* and *S. kudriavzevii* alleles, and one single pattern for *BRE5*, matching the *S. cerevisiae* allele restriction pattern (Peris *et al.,* 2012a). The chimerical nature of this chromosome is confirmed by the caryoscope diagram where two different $\log_2$ ratios are observed, indicating a different *S. cerevisiae* chromosome content. By combining both sources of information, we can deduce that most chromosome XIV corresponds to two copies of *S. cerevisiae* (according to the *EGT2* RFLP pattern and aCGH data) and one of *S. kudriavzevii* (according to *EGT2* RFLP pattern), but chromosome XIV right end corresponds to three copies of *S. cerevisiae* (according to *BRE5* RFLP pattern and aCGH data). The recombination site in the chimerical chromosome can be located according to the $\log_2$ ratio jump observed in the caryoscope diagram.

and chromosomes X and XIII from MR25, however, their putative chimerical nature could not be detected by the aCGH analysis (Figure 3).

Following the same methodology used in our previous study (Peris *et al.,*, 2012b), we obtained a list of *S. cerevisiae* genes lost in both hybrids IF6 and MR25. Both IF6 and MR25 have depleted a similar number of genes classified as retrotransposons as well as genes belonging to the *ASP3*, *CUP1* and *ENA* clusters (Table S8). In particular, hybrid IF6, obtained from a dietary supplement, exhibited a deleted region (YLR155C-YLR256w) in its *S. cerevisiae* chromosome XII (Figure 3). This region is adjacent to the rDNA repeat region located between YLR154C and YLR155C, which is not included in the microarray platform. A PCR amplification of the 5.8S-ITS region and the subsequent restriction analysis (Gonzalez *et al.,*, 2006), revealed the absence of *S. cerevisiae* rDNA genes in this region (data not shown).

With respect to their *S. kudriavzevii* subgenome, IF6 and MR25 hybrids lost ~33% and ~18% of the total *S. kudriavzevii* genes, respectively. Gene Ontology (GO) enrichment analysis applied to the common set of *S. kudriavzevii* genes maintained by the two hybrids, demonstrated a high frequency of stress response genes among those *S. kudriavzevii* genes conserved in both hybrids (Table S9). Some of the significant GO terms shared by MR25 and IF6 are “response to stimulus”” with p-values < 0.05. In

the case of MR25 is also important to note the significant GO term “cellular lipid metabolic process” and “response to stress” (p-value < 0.05).

3.5 Analysis of the number of hybridization events

Genome composition data obtained for the 26 *S. cerevisiae* x *S. kudriavzevii* hybrids from this study as well as from previous studies (Belloch *et al.,*, 2009; Peris *et al.,*, 2012b) were used to reconstruct a parsimony tree based on the presence of chimerical chromosomes, on the absence of chromosomes from one or another parental strain and the presence of specific allelic variants. Using the information from this parsimony tree together with the putative genetic constitution of the hypothetical parental strains obtained from the phylogenetic analysis of nuclear gene sequences, as well as from *COX2* sequences also obtained in our previous studies (Peris *et al.,*, 2012a), allowed us to reduced the number of hybridization events to a minimum of six for the *S. cerevisiae* x *S. kudriavzevii* hybrids under analysis, and two additional events for the origin of the *S. cerevisiae* x *S. kudriavzevii* x *S. uvarum* triple hybrids. The putative ploidies of the parental cells involved in hybridization were also estimated by analyzing the genomic constitution of the hybrids derived from each event.

Figure 4 shows five out of the six different origins for double hybrids proposed according to this study. AMH is not included due to its complex genome structure, because it is a tetraploid hybrid that lost most of the *S. kudriavzevii* subgenome (Peris *et al.,*, 2012a, b). Independent origin for AMH is clearly supported by the different sets of data used in this analysis.

Wine hybrid strains from Switzerland (W27, SPG14-91, SPG16-91,126, 172, 319 and 441), and the Trappist brewing strains CECT11003 and CECT11004 share a common origin. Their nuclear genomes derive from a hybridization event between the hypothetical *S. cerevisiae* CG2 and *S. kudriavzevii* KG2 parents (Figure S1). They inherited their mtDNA type K2

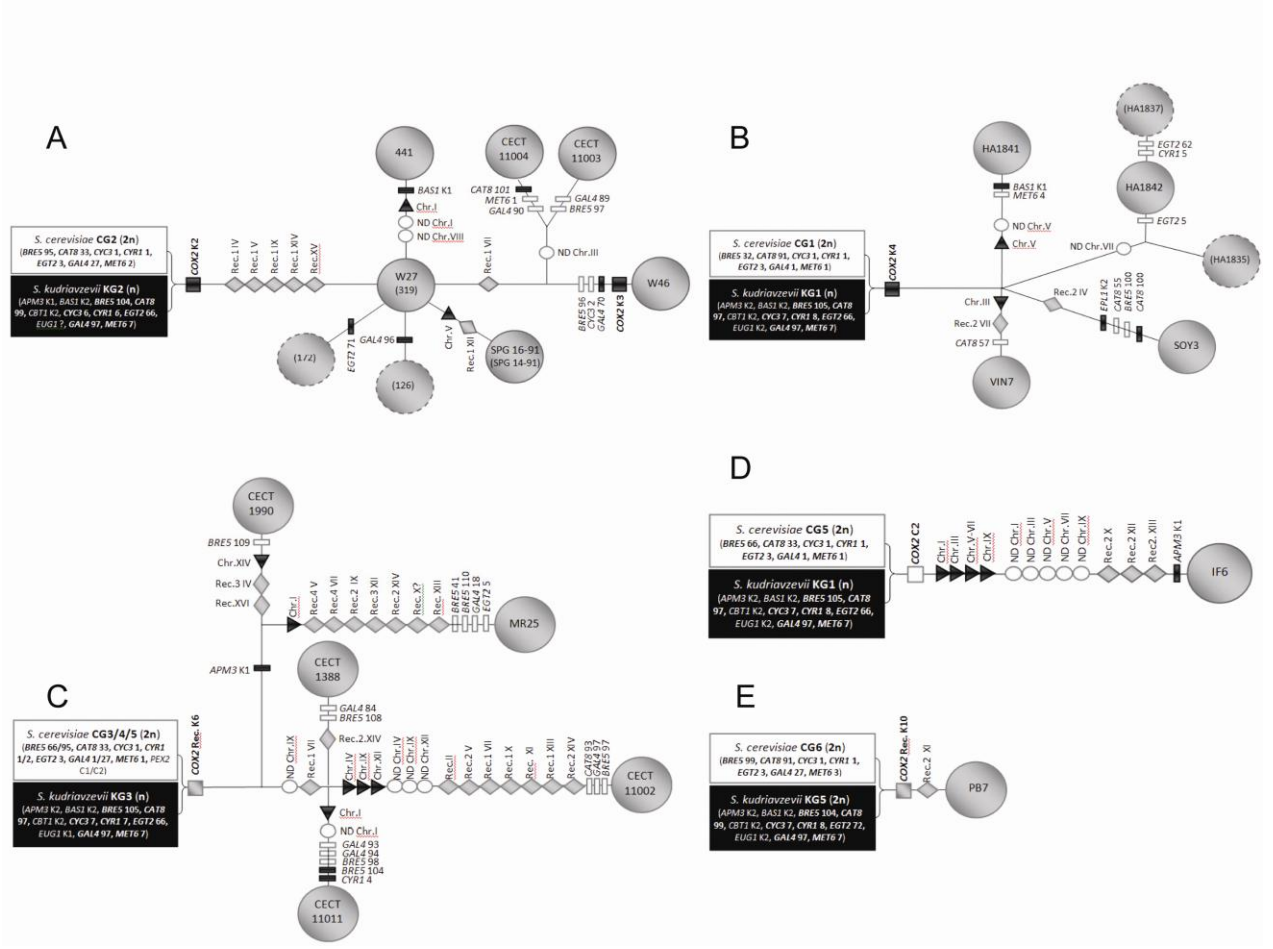

A
441
CECT 11004
CECT 11003
S. cerevisiae CG2 (2n)
(BRE5 95, CAT8 33, CYC3 1, CYR1 1, EGT2 3, GAL4 27, MET6 2)
S. kudriavzevii KG2 (n)
(APM3 K1, BAS1 K2, BRE5 104, CAT8 99, CBT1 K2, CYC3 6, CYR1 6, EGT2 66, EUG1 ?, GAL4 97, MET6 7)
W27 (319)
W46
(1/2)
(126)
SPG 16-91 (SPG 14-91)
B
HA1841
(HA1837)
HA1842
(HA1835)
S. cerevisiae CG1 (2n)
(BRE5 32, CAT8 91, CYC3 1, CYR1 1, EGT2 3, GAL4 1, MET6 1)
S. kudriavzevii KG1 (n)
(APM3 K2, BAS1 K2, BRE5 105, CAT8 97, CBT1 K2, CYC3 7, CYR1 8, EGT2 66, EUG1 K2, GAL4 97, MET6 7)
SOY3
VIN7
C
CECT 1990
CECT 1388
CECT 11002
CECT 11011
MR25
S. cerevisiae CG3/4/5 (2n)
(BRE5 66/95, CAT8 33, CYC3 1, CYR1 1/2, EGT2 3, GAL4 1/27, MET6 1, PEX2 C1/C2)
S. kudriavzevii KG3 (n)
(APM3 K2, BAS1 K2, BRE5 105, CAT8 97, CBT1 K2, CYC3 7, CYR1 7, EGT2 66, EUG1 K1, GAL4 97, MET6 7)
D
IF6
S. cerevisiae CG5 (2n)
(BRE5 66, CAT8 33, CYC3 1, CYR1 1, EGT2 3, GAL4 1, MET6 1)
S. kudriavzevii KG1 (n)
(APM3 K2, BAS1 K2, BRE5 105, CAT8 97, CBT1 K2, CYC3 7, CYR1 8, EGT2 66, EUG1 K2, GAL4 97, MET6 7)
E
PB7
S. cerevisiae CG6 (2n)
(BRE5 99, CAT8 91, CYC3 1, CYR1 1, EGT2 3, GAL4 27, MET6 3)
S. kudriavzevii KG5 (2n)
(APM3 K2, BAS1 K2, BRE5 104, CAT8 99, CBT1 K2, CYC3 7, CYR1 8, EGT2 72, EUG1 K2, GAL4 97, MET6 7)

Figure 4. Possible multiple origins for hybrid strains based on Supernetworks, Polymorphic sites (Figure S1), Parsimony (Figure S2) and Neighbor-Joining (Figure S3) gene trees, PCR-RFLP data (González *et al.,* 2008; Peris *et al.,* 2012a), *COX2* sequence data (Peris *et al.,* 2012a) and maximum parsimony analysis of chromosome rearrangements (Peris *et al.,* 2012b). Five out of six hybridization events are depicted in this figure, AMH and tripe hybrid origins have not been depicted due because they involved secondary hybridization events, in the case of AMH with another *S. cerevisiae* strain. The putative genetic backgrounds of the parental strains involved are indicated by squares on the left of each network. Symbols: triangles correspond to chromosome loss; squares to mitochondrial COX2 haplotypes; diamonds to chromosome recombination events; rectangles to mutations generating new allele variants; circles to chromosome non-disjunctions. Those depicted in white are referring to events occurring in the *S. cerevisiae* subgenome of hybrids; in black, in the *S. kudriavzevii* subgenome; and in grey, those events involving both subgenomes (recombination events).

from *S. kudriavzevii* (Peris *et al.,*, 2012a). Hybrid W46 was also included in this group although it exhibits a mitochondrial type K3 (Figure 4), derived from K2 by a single nucleotide difference (Gonzalez *et al.,*, 2008).

The group of Austrian hybrids HA as well as wine hybrids VIN7 and SOY3 have also a common origin in a hybridization event involving hypothetical parents *S. cerevisiae* CG1 and *S. kudriavzevii* KG1 (Figures 2, 4 and S1), and sharing the mitochondrial type K4 from *S. kudriavzevii* (Peris *et al.,*, 2012a).

A third group includes the brewing triploid hybrids CECT1388, CECT1990, CECT11002, CECT11011 and the clinical isolate MR25, sharing several genome rearrangements and restriction patterns as well as a recombinant mtDNA type K6 (Peris *et al.,*, 2012a). According to the seven gene sequence analysis, these strains seem to have independent hybridization origins from crosses between different *S. cerevisiae* parents (CG3, CG4 and CG5) but the same *S. kudriavzevii* strain KG3 characterized by an specific *CYR1* allele (Figure S1). These contradictory results may be explained by considering a heterozygous *S. cerevisiae* diploid cell containing wine and non-wine alleles as the parental strain, as mentioned above.

Wine strain PB7 from Leon, Spain, was included in the same subgroup than the Austrian wine strains according to the supernetwork analyses, due to network simplification (Figure 2). However, this strain likely originated in an independent hybridization event because it derives from different parents, the

mosaic *S. cerevisiae* CG6 and *S. kudriavzevii* KG5 genotypes (Figure S1), exhibits a recombinant mtDNA K10 (Peris *et al.,*, 2012a), and finally, it possesses a tetraploid genome (Peris *et al.,*, 2012b).

In the case of IF6, although it shares the same *S. kudriavzevii* KG1 ancestor with Austrian hybrids, its *S. cerevisiae* parental strain is clearly different: a mosaic CG5 genotype closer to the *S. cerevisiae* parent of the brewing hybrids. The different hybrid combination of parental genotypes supports an independent origin for this strain.

4. Discussion

By analyzing the sequences of four nuclear genes from a total of more than 250 *S. cerevisiae* yeast isolates from wine (Europe, South America and South Africa) and non-wine origins (wild, brewing, cider, sake and traditional beverage fermentations mainly from Latin America, but also from Africa and Asia), Arias (2008) demonstrated the existence of two groups of alleles, those present only in strains isolated from non-wine sources, called non-wine alleles, and another group of alleles that, while they also appear in non-wine alleles, they are the only alleles present in wine strains (wine alleles). These wine alleles are much less frequent in non-wine strains, and they mainly appear in heterozygosis with non-wine alleles. Liti *et al.,* (2009) obtained the complete genome sequences of 37 *S. cerevisiae* strains from different sources of isolation and geographic origins. The phylogenetic analysis of nucleotide polymorphisms showed a complex *S. cerevisiae* population structure. Liti *et al.,* (2009) observed five genotypic lineages, called according to their origins or source of isolation as Malaysia, West Africa, sake, North America and 'Wine/European', which exhibited the same phylogenetic relationships across their entire genomes. The strains from these five lineages were considered as 'clean', pure strains, representative of diverged populations. The other strains evidenced variable phylogenetic relationships depending on the genome

region analyzed, and were considered as ‘mosaics’ with a mixed genome architecture that could be due to human traffic in yeast strains and subsequent recombination between them. The analysis of the sequences of the same four gene regions used by Arias (2008) indicated that alleles present in the four non-wine lineages fell within the group of non-wine alleles; alleles present in strains of the wine/European lineage were included within the ‘wine allele’ group, and the locations of the ‘mosaic’ alleles were variable depending on the gene. Because Liti *et al.,* (2009) sequenced derivative monosporic cultures, some of the ‘mosaic’ parental strains could be heterozygous for wine and non wine alleles for many genes, as observed by Arias (2008). High levels of heterozygosity for non-wine yeast were also observed by Fay and Benavides (2005), and for ale strains by Dunn and Sherlock (2008).

The accessibility to such a collection of sequences (including genome sequences) from *S. cerevisiae* strains from different sources of isolation and geographic origins was an excellent opportunity to decipher the nature of the *S. cerevisiae* parents involved in the origin of hybrids. This way, for all genes under analysis, *S. cerevisiae* alleles from hybrids were always clustered within the wine allele group, with the exception of the *BRE5* allele from the brewing strain CECT11011, which clustered in the non-wine group, and *CYR1* allele from the brewing CECT1388 and CECT11011, which clustered with non-wine strains from Liti *et al.,* (2009) in the individual gene trees. Moreover, the phylogenetic supernetwork analysis of *S. cerevisiae* alleles from hybrids identified two main subgroups of *S. cerevisiae* parental strains, and due to its simplification it failed to detect mosaic *S. cerevisiae* genotypes. It followed that the *S. cerevisiae* CG2 parental strain was involved in the hybridization event that originated the complete group of wine Swiss hybrids and *S. cerevisiae* CG1 was involved in the origin of the Austrian wine hybrids, SOY3 and Vin7.

The aCGH analyses of hybrid genome composition (present study and (Peris *et al.,*, 2012b)) showed the depletion or underrepresentation of certain

S. cerevisiae genes (Ty retrotransposons and *ENA* and *ASP* gene families), which were proposed as genomic signatures for wine *S. cerevisiae* yeasts (Carreto *et al.,*, 2008; Dunn *et al.,*, 2005), which is in agreement with the postulated wine origin of the *S. cerevisiae* parental strains involved in the generation of these hybrids. The maintenance of *S. kudriavzevii* genes related to stress response, in MR25 and IF6, and lipid metabolism, in MR25, also confirms the importance of *S. kudriavzevii* subgenome in cold stress resistance, postulated in previous studies (Peris *et al.,* 2012b).

In the case of IF6, aCGH and PCR confirmation of 5.8S-ITS regions support the loss of heterozygosity (LOH) of the rDNA region in chromosome XII, maintaining only the *S. kudriavzevii* sequences for this region. This region has been characterized, in plants and animals, to be under concerted evolution (Joly *et al.,*, 2004; Wendel *et al.,*, 1995; Gromicho *et al.,*, 2006). This has been also observed in a natural hybrid *S. pastorianus* (CBS 1538 strain), where the *S. cerevisiae* rDNA region of chromosome XII has been lost (Kodama *et al.,*, 2005).

The wine origin of the *S. cerevisiae* parent of most *S. cerevisiae* x *S. kudriavzevii* hybrids has already been postulated in previous works based on genomic composition data inferred by aCGH and PCR-RFLP analysis (Peris *et al.,* 2012b), as well as by microsatellite analysis (Erny *et al.,*, 2012). The use of a multilocus sequence analysis approach certainly confirms the wine origin of the *S. cerevisiae* strains involved in the generation of most *S. cerevisiae* x *S. kudriavzevii* hybrids.

The exceptions are the brewing hybrid CECT11011, in which a possible recombinant *BRE5* allele is present, and CECT1990 and CECT11011, which contain *CYR1* non-wine alleles, and hence, a heterozygous non-wine *S. cerevisiae* strain, with both wine and non-wine alleles, could be involved in their origin. Dunn and Sherlock (2008) demonstrated that *S. pastorianus* hybrids, responsible of lager beer fermentations, very likely derived from a

cross between a haploid *S. bayanus*-like strain, later identified as belonging to the new species *S. eubayanus* (Libkind *et al.,*, 2011), and a diploid *S. cerevisiae* strain, related to ale brewing strains, which are characterized by a high heterozygosity. Arias (2008) also included in his study several ale strains that showed as heterozygous, for wine and non-wine alleles. Therefore, the parental *S. cerevisiae* involved in the origin of brewing hybrids CECT1990 and CECT11011 could be an ale strain originally heterozygous for wine and non-wine alleles. Another brewing hybrid, strain CECT11002, appeared as related to the brewing hybrids and the clinical isolate, but it did not contain non-wine alleles for the genes under analysis; all these hybrids may also have been originated from a similar ale parental strain. Erny *et al.,* (2012) included in their microsatellite analysis a Chimay strain which clusters with the *S. cerevisiae* brewing strains. We do not know whether their Chimay strain and our CECT11002 (also from Chimay) is the same or not, but at least they should be related, which could corroborate the 'ale' origin of their *S. cerevisiae* parent. Genome sequencing of one of these strains will elucidate this hypothesis.

By using the population genetic information from Arias (2008), we also tried to determine the exact geographic origin of the parental *S. cerevisiae* strains. We looked for particular *S. cerevisiae* strains from different wine regions possessing the combination of alleles present in the hypothetical parental S. *cerevisiae* strains. With the exception of one *CAT8* allele, genotype CG1 was present in strains from Chile, South Africa, Switzerland and Spain; and genotypes CG2 and CG3, with the exception of *BRE5*, were found in strains from Argentina, Chile, Italy, Japan, South Africa, Austria, France and Spain. Other genotypes, with slight differences were found in Argentina, Chile, South Africa, Austria, Slovenia, Switzerland, Italy, Japan, France and Spain. As strains from the new winemaking regions (South America and South Africa in this case) were introduced from Europe with vines and winemaking tools, the most probable geographic origin for

hybridization, according to the *S. cerevisiae* hypothetical parental genotype, is Europe.

The European origin of hybrids is also supported by the phylogenetic analysis of *S. kudriavzevii* alleles. Alleles present in hybrids were detected among European *S. kudriavzevii* pure strains. Three of seven alleles of *S. kudriavzevii* KG1, were found in 3 *S. kudriavzevii* strains from Ciudad Real (Spain), Castellon (Spain) (Lopes *et al.,*, 2010) and Portugal (Sampaio & Gonçalves, 2008). However, other genotypes have not been found among the few *S. kudriavzevii* pure strains available. Future surveys on the genetic variability of European populations of *S. kudriavzevii* may be of interest to decipher the geographic origin of hybridization, because this wild species has not been subjected to human traffic and it may preserve its original population structure in the same way than *S. paradoxus* (Johnson *et al.,*, 2004; Koufopanou *et al.,*, 2006). A recent study (Erny *et al.,*, 2012), complementary to the present one, on the possible origin of a different set of European *S. cerevisiae* x *S. kudriavzevii* hybrids from winemaking (only four Swiss hybrids and VIN7 are in common), carried out by means of microsatellite information, also confirmed the European origin of the putative parental strains of hybrids.

By combining the phylogenetic analysis of gene sequences with all the available information on genetic and genomic characterization of *S. cerevisiae* x *S. kudriavzevii* hybrids (Gonzalez *et al.,*, 2008; Belloch *et al.,*, 2009; Erny *et al.,*, 2012; Peris *et al.,*, 2012a, b), a total of six potential hybridization events were determined. The first hybridization event involved a haploid *S. kudriavevii* parental KG2 with mtDNA K2 and a diploid *S. cerevisiae* parental CG2. This event originated all Swiss hybrids and the related Trappist brewing strains CECT11003 and 11004. This clearly independent origin for Swiss wine hybrids is in accordance with the microsatellite phylogenetic analysis of hybrids performed by Erny *et al.,* (2012).

A second hybridization event involving a haploid *S. kudriavzevii* KG1 with mtDNA type K4 (found in all hybrids from this group) and a diploid *S. cerevisiae* CG1 originated a lineage of hybrids widely distributed in different wine regions such as Austrian hybrids, the Croatian strain SOY3, and the South African hybrid VIN7 of putative European origin according to Erny *et al.,* (2012). These authors observed in their study that VIN7 is included in the same group as other Alsatian and German wine hybrids and bears a close relationship to Hungarian wine hybrids, confirming an European origin for VIN7. Therefore, this is a lineage of wine hybrids widely distributed from the Rhine valley (Alsace and Germany) to the Danube valley (Pannonian region: Austria, Croatia and Hungary).

A third hybridization event was involved in the origin of a lineage of brewing strains also widely distributed in ale breweries from England, Germany, Belgium (Chimay Trappist Abbey), New Zealand and the clinical isolate MR25. This hybridization event involved a haploid *S. kudriavzevii* parental close to K2, KG3 strain, and probably a heterozygous diploid *S. cerevisiae* parental. An “ale” *S. cerevisiae* strain heterozygous for wine and non-wine alleles could be involved in the origin of this group of hybrids.

Hybrid PB7 was probably originated from two diploid cells derived from mosaic strains *S. cerevisiae* CG6 and *S. kudriavzevii* KG5. Its tetraploidy (Peris *et al.,*, 2012b) and the presence in this hybrid of a recombinant mtDNA (Peris *et al.,*, 2012a) supports an independent hybridization event.

Independent origins are postulated for hybrids IF6 and AMH. In the case of AMH, its complex tetraploid genome (Peris *et al.,*, 2012b), in which most of the *S. kudriavzevii* subgenome is lost (Peris *et al.,*, 2012a), led us to suspect a possible scenario in which a diploid *S. cerevisiae* crossed with a haploid *S. kudriavzevii* strain and, after sporulation or a drastic *S. kudriavzevii* genome reduction, a diploid spore or an evolved derivative backcrossed with a diploid *S. cerevisiae.* IF6 was originated from a cross between a diploid *S. cerevisiae*

CG5 mosaic genotype and a haploid *S. kudriavzevii* KG1, identical to the one involved in the origin of Austrian hybrids. Therefore, the possibility of a common origin with Austrian HA, VIN7, and SOY3 hybrids cannot completely ruled out if a heterozygous *S. cerevisiae* ancestor were involved in the hybridization event. However, this hypothesis not only requires the differential loss or segregation of alleles in the IF6 and Austrian lineages, but also the independent acquisition of the mitochondrial genome from the hybrid zygote, *S. cerevisiae* type C2 in IF6 and *S. kudriavzevii* type K4 in the Austrian lineage. This is possible in hybrid zygotes where three types of mitochondrial genomes may be present: two from each parental and a recombinant, generated after mitochondria fusion (Berger and Yaffe, 2000), but mitochondrial sorting occurs from the first budding formation (Shibata and Ling, 2007), generating independent lineages that are difficult to distinguish from independent hybridizations in which parental relatives were involved.

Finally, triple hybrids *S. cerevisiae* x *S. kudriavzevii* x *S. uvarum* are not shown in Figure 4, also due to to their complex origins, in which a secondary hybridization was involved. However, the supernetwork analysis and gene trees information indicates that CBS2834 and CID1 were probably derived from the same (or similar) *S. kudriavzevii* parent (KG6) but different *S. cerevisiae* parental strains, the same than the Austrian strains (CG1) for CBS2834 and similar to PB7 (CG6) for CID1.

Finally, the origin of the triple hybrids CID1 and CBS2834 is not clear due to the additional occurrence of a secondary hybridization event either between a *S. cerevisiae* x *S. kudriavzevii* hybrid or derivative with a *S. uvarum* strain or between a *S. cerevisiae* x *S. uvarum* hybrid or derivative with a *S. kudriavzevii* strain. However, CID1 and CBS2834 were probably originated from independent hybridization events.

Most hybrids seem to have been generated by rare-mating events involving a diploid *S. cerevisiae* strain and a haploid strain of *S. kudriavzevii*

generating different chimerical genomes with ploidy values close to 3n. This is most clear for brewing strains (CECT1388, CECT1990, CECT11011 and MR25) where heterozygous genes could be observed. In PB7, which exhibited a ploidy value of 3.96, two diploid parents could be involved. Rare-mating has already been proposed as a mechanism for natural hybrid generation (Barros Lopes *et al.,*, 2002). Additionally, artificial hybrids generated by rare mating are easily obtained in laboratory conditions (Pérez-Través *et al.,*, 2012).

Hybrid distribution and their physiological properties, together with the conclusions of recent studies on the population-genetic structure of *S. cerevisiae* (Fay & Benavides, 2005; Liti *et al.,*, 2009) as well as the phylogenetic analyses performed in the present study, can be used to speculate a possible scenario for the hybridization process. Grapevine (Arroyo-García *et al.,*, 2006) and barley (Badr *et al.,*, 2000) domestication mainly occurred in the Middle East, where the earliest archaeological evidence of winemaking (McGovern *et al.,*, 1997) and brewing (Michel *et al.,*, 1992) have been discovered. From these areas of domestication, there was a gradual radiation to adjacent areas of the Mediterranean regions of Europe and Africa, following the spread of Phoenician, Greek and Carthaginian civilizations. Finally, the expansion of vine growing and winemaking to temperate regions of Oceanic and Continental climates of Europe, following the main trade fluvial routes, was performed under the influence of the Romans, who would take vine-growing to the limits of their empire, the Rhine and Danube Rivers. By the end of the Roman Empire, grape growing was common in most European locations. In the Middle Ages, viticulture and enology were improved and expanded by Christian monks.

Recent studies on the genetic diversity of *S. cerevisiae* populations (Fay & Benavides, 2005; Aa *et al.,*, 2006; Arias, 2008; Liti *et al.,*, 2009) show that wine strains constitute a genetically differentiated population that could have appeared during the process of adaptation to winemaking conditions, a

process of fortuitous domestication of a *S. cerevisiae* wine strain. The microsatellite population analysis of *Saccharomyces* strains (Legras *et al.,*, 2007) also suggests that this population likely originated in the Near East and spread during the expansion of grapevine and winemaking.

About 2,000 years ago, wine *S. cerevisiae* yeasts were likely taken by the Romans, together with the vines and winemaking tools, to the Northern limit of grapevine distribution. There, *S. cerevisiae* wine strains, even nowadays, have problems when performing wine fermentations at the lower temperatures to which other *Saccharomyces* species are better adapted (Salvadó *et al.,*, 2011). In these regions, cryotolerant species, such as *S. bayanus* var. *uvarum*, may outcompete *S. cerevisiae* (Naumov *et al.,*, 2000; Naumov *et al.,*, 2001; Naumov *et al.,*, 2002; Demuyter *et al.,*, 2004). Under such circumstances, however, hybrids may have advantages over the parental species (Serra *et al.,*, 2005; Arroyo-Lopez *et al.,*, 2009; Belloch *et al.,*, 2008). This is due to the acquisition of physiological properties from both parents, which provide a mechanism for selection of hybrids (Zambonelli *et al.,*, 1997; Masneuf *et al.,*, 1998; Greig *et al.,*, 2002; Gonzalez *et al.,*, 2007). In the case of *S. cerevisiae* x *S. kudriavzevii* hybrids, they acquired good alcohol and glucose tolerances and fast fermentation performances from *S. cerevisiae* (Belloch *et al.,*, 2008; Arroyo-López *et al.,*, 2010) and a better adaptation to low and intermediate temperatures from *S. kudriavzevii* (Gonzalez *et al.,*, 2007; Belloch *et al.,*, 2008; Arroyo-Lopez *et al.,*, 2009).

These *S. cerevisiae* x *S. kudriavzevii* hybrids likely appeared several times, according to this study, and became frequent in some areas of the Northern limit of vine growing, but they could probably spread in Central Europe with the expansion of vine growing and winemaking practices that occurred during the Middle Ages (McGovern, 2010). Winemaking was preserved and improved during the Middle Ages by Christian monks. Benedictine abbeys were the main wine producers and traders, but the

Cistercian reformation made possible the main revolution in winemaking improvements and vine growing extension (Burton & Kerr, 2011).

From their original abbeys in Burgundy, Cistercians spread across Europe during the 11th and 12th centuries to establish more than 300 abbeys. During this expansion, the white monks spread the viticulture and enology practices to the Rhine and Danube valleys and the Pannonian basin of Central Europe (Burton & Kerr, 2011). They extended the Burgundian family of grape varieties, mainly Chardonnay and Pinots, as well as German varieties, and with them likely the hybrid yeasts responsible for wine fermentation.

In the regions where the main lineages of *S. cerevisiae* x *S. kudriavzevii* wine hybrids have been found, winemaking was introduced or improved on by Cistercian monks. In fact, the Cistercian order is given credit for planting in the French regions of Burgundy, Chablis, Loire, Rhone, Champagne (where the *S. cerevisiae* x *S. kudriavzevii* hybrid EPII, also called Epernay 2, was isolated (Dunn *et al.,*, 2012)), Alsace (where many hybrids are also present and predominant (Erny *et al.,* 2012)) and in several other wine regions in Central Europe. Some of these regions are: Rheingau Wine Region in Germany, where hybrid AMH (Assmannshausen) and those from Geisenheim (Erny *et al.,* 2012) were isolated; Thermenregion, Austria, where HA hybrids, characterized in this study, were found as predominant in vineyards (Lopandić *et al.,* 2007); Slavonian Croatia, where SOY3 was isolated; and Hungary, where these hybrids have also been found (Erny *et al.,*, 2012).

Competing interests

None

Author's contribution

EB conceived and supervised this study. DP, CL and AA performed the experimental work. DP and CL performed the data analyses. DP and CL wrote

the first version of the manuscript. DP, CL and EB participated in the final manuscript revision.

Acknowledgments and funding

This work was supported by Spanish Government project AGL2009-12673-CO2-02 to EB, and Generalitat Valenciana grant PROMETEO/2009/019. DP, CL and AA acknowledge a Spanish Government FPI fellowship, a Spanish Government postdoctoral contract, and a Mexican Government PROMEP fellowship, respectively. The authors thank Prof. J. Gafner, Prof. K. Lopandić, Dr. H. Gangl, Dr. Álvarez-Pérez, Dr. S. Orlić, Dr. R. de Llanos, S. Llopis, Lallemand Bio and Anchor Wine Yeasts for providing yeast strains. We are also grateful to Dr. J.L. Legras for his interesting suggestions and comments.

OBJECTIVE 4. Reconstruction of ancestral hybridization events between *Saccharomyces* species

Chapter 1

COX2 sequences from *Saccharomyces* species contain a recombination hotspot

David Peris[1], Armando Arias[1, 2], Eladio Barrio[3,4]

[1]Institut "Cavanilles" de Biodiversitat i Biologia Evolutiva, Universitat de València, Spain

[2]Laboratorio de Biotecnología, Unidad de Botánica y Zoología, CUCBA, Universidad de Guadalajara, Zapopan, Jalisco, Mexico.

[3]Departament de Genètica, Universitat de València, Spain.

[4]Departamento de Biotecnología, Instituto de Agroquímica y Tecnología de Alimentos, CSIC, Valencia, Spain.

***Corresponding author:**

Dr. Eladio Barrio
Institut 'Cavanilles' de Biodiversitat i Biologia Evolutiva
Universitat de València
P.O. Box 22085
E- 46071 València
SPAIN
Tel.: +34 963 543 667
Fax: +34 963 543 670
E-mail: eladio.barrio@uv.es

Abstract

Saccharomyces genus includes seven species based on patterns of breeding. The species boundaries are not clear due to the description of several reticulate events between *Saccharomyces* species, such as Horizontal Gene Transfer (HGT), introgressions and hybridizations. These events are footprints in the genomes that indicate ancestral hybridization events driving to non stable hybrid formations increasing the difficulties of phylogenetic studies due to phylogenetic incongruence. Application of next generation sequencing facilitates the identification of *Saccharomyces* species due to complete nuclear genome information. Some genome sequencing projects using nuclear genome have described the population structure of *S. cerevisiae* and *S. paradoxus* depending on source/country origin. Although mitochondrial genome is not a proper marker for phylogeographical studies in yeast, the aim of this study is support the "pure" nature of the previous strains, using the mitochondrial *COX2* gene sequences. In addition, our results indicate a common recombination hotspot in *COX2* gene between different *Saccharomyces* species due to ancestral hybridization events and we postulate a molecular mechanism involve in the recombination.

In the present study, we analyzed the mitochondrial *COX2* sequences from 532 strains, belonging to the seven pure *Saccharomyces* species and to five types of interspecific hybrids, to determine the extent of recombination in this gene. As a result, patent evidences of extensive interspecific mtDNA recombination are described for the first time in *Saccharomyces* yeasts and a putative mechanism to explain the presence of such recombination hotspot is proposed.

Keywords: *Saccharomyces* genus, phylogeny, hotspot recombination, *COX2*

1. Introduction

The genus *Saccharomyces* encompasses the industrially most exploited species known to man, *S. cerevisiae.* Strains of *S. cerevisiae* are used worldwide in the production of different fermented foods and beverages. In addition to *S. cerevisiae*, at present, several other species are recognized in the genus *Saccharomyces* (Vaughan-Martini and Martini 2011): *S. arboricolus*, *S. bayanus* (with two varieties, *bayanus* and *uvarum*), *S. cariocanus, S. kudriavzevii, S. mikatae, S. paradoxus* and *S. pastorianus.* Strains belonging to the last species, *S. pastorianus* (syn. *S. carlsbergensis*) are employed in lager beer production, and they were described as hybrids between *S. cerevisiae* and *S. bayanus* (Kodama *et al.,* 2005; Rainieri *et al.,* 2006; Dunn and Sherlock, 2008; Nakao *et al.,* 2009; Bond 2009). Some authors proposed the reinstatement of *S. bayanus* var. *uvarum* to the species level, as *S. uvarum*, when the hybrid nature of *S. bayanus* var. *bayanus* (renamed as *S. bayanus*) strains was demonstrated (Nguyen and Gaillardin, 2005; Rainieri *et al.,* 2006). Recently, pure strains of *S. bayanus* var. *bayanus*, renamed as *S. eubayanus*, were found in the Argentinian Patagonia (Libkind *et al.,* 2011). This new species was proposed as the ancestor of *S. bayanus* and *S. pastorianus* hybrids together with *S. uvarum* and *S. cerevisiae*, respectively (Libkind *et al.,* 2011; Nguyen *et al.,* 2011). In the case of *S. paradoxus*, four different populations have been described: American, European, Far Eastern and Hawaiian (Liti *et al.,* 2006; Liti *et al.,* 2009), which seem to be in the process of speciation (Kuehne *et al.,* 2007).

With the exception of the hybrid taxa *S. pastorianus* and *S. bayanus*, *Saccharomyces* species definition is congruent with the biological species concept based on reproductive isolation (Naumov, 1996; Liti *et al.,* 2006; Greig, 2008). Artificial interspecific hybridization experiments have extensively been performed to delimitate the species of the genus

Saccharomyces according to the biological species concept based on postzygotic reproductive isolation (Naumov *et al.,* 2000; Liti *et al.,* 2006; Naumov, 2009). These experiments indicated that *Saccharomyces* interspecific hybrids can easily be formed, and although sterile, they are viable and can be maintained by asexual reproduction.

Although pre-mating reproductive isolation in the form of mate choice can reduce hybridization rates (Maclean and Greig, 2008), hybrids just form when the closest available mate is another species. These *Saccharomyces* hybrids are generally sterile, with less than 1% viable ascospores (Greig, 2008). Activation of the mismatch repair system by sequence divergence between the two parental genomes (Chambers *et al.,* 1996; Hunter *et al.,* 1996; Greig *et al.,* 2003; Liti *et al.,* 2006), and in a lesser extent the presence of reciprocal translocations (Delneri *et al.,* 2003; Liti *et al.,* 2006), were described as the most important postzygotic barriers contributing to the reproductive isolation among *Saccharomyces* species. Finally, recent cases of cytonuclear incompatibility between different nuclear and mitochondrial gene pairs have also been reported as involved in *Saccharomyces* interspecific hybrid sterility (Lee *et al.,* 2008; Chou *et al.,* 2010).

Because reproductive isolation in the genus *Saccharomyces* is mainly postzygotic, interspecific hybridization is possible in natural environments. In addition to *S. pastorianus* (*S. cerevisiae* x *S. eubayanus*), *S. bayanus* (*S. eubayanus* x *S. uvarum*), other new natural *Saccharomyces* hybrids have also been isolated from wine, beer, cider, dietary supplements or clinical samples. They include *S. cerevisiae* x *S. uvarum* (Masneuf *et al.,* 1998; Le Jeune *et al.,* 2007) and *S. cerevisiae* x *S. kudriavzevii d*ouble hybrids (González *et al.,* 2006; Lopandic *et al.,* 2007; González *et al.,* 2008; Erny *et al.,* 2008; Peris *et al.,* 2012a), as well as *S. cerevisiae* x *S. kudriavzevii* x *S.*

uvarum (González *et al.,* 2006) and *S. cerevisiae* x *S. eubayanus* x *S. uvarum* (Rainieri *et al.,* 2006; Nguyen *et al.,* 2011) triple hybrids.

Recent and past events of hybridization leave recombination footprints that can be detected. This way, the genome characterization of *Saccharomyces* hybrids, either by array comparative genome hybridization (aCGH) with DNA chips (Bond *et al.,* 2004; Kodama *et al.,* 2005; Belloch *et al.,* 2009; Peris *et al.,* 2012b) or sequencing (Nakao *et al.,* 2009; Borneman *et al.,* 2012), revealed the presence of chimerical chromosomes generated by recombination between homologous regions of the parent genomes. Moreover, genome sequencing and aCGH analysis of *S. cerevisiae,* and *S. paradoxus* strains showed evidence of introgression in both directions (Liti *et al.,* 2006; Wei *et al.,* 2007; Doniger *et al.,* 2008; Muller and McCusker, 2009; Dunn *et al.,* 2012), likely generated during unstable hybridization between these species followed by backcrosses with one or the other parental species. Recently, introgression from *S. mikatae* to *S. cerevisiae* has also been described (Dunn *et al.,* 2012).

In a recent characterization of new *S. cerevisiae* x *S. kudriavzevii* hybrids (Peris *et al.,* 2012a), the mitochondrial gene *COX2*, encoding subunit II of the cytochrome-c oxidase complex, from these hybrids was sequenced to determine their mitochondrial inheritance. These sequences showed putative evidence of interspecific recombination at the mitochondrial genome from certain hybrids.

Yeast mitochondrial DNA (mtDNA) shows a paucity of genes, which in some species is compensated by a structural complexity in the intergenic regions, resulting from the distribution of optional introns, rearrangements and/or insertions/deletions (Clark-Walker, 1989; Clark-Walker, 1992; Langkjær *et al.,* 2003; Nosek *et al.,* 2006). In *S. cerevisiae*, mitochondria from the two parents positioned towards the central region of the zygote fuse after mating. In these fused mitochondria, parental mtDNAs mix and

recombine with high frequency (Nunnari *et al.,* 1997; Berger and Yaffe, 2000). A large body of data confirming intraspecific mitochondrial recombination in *S. cerevisiae* is available since several decades ago (Dujon *et al.,* 1974; Fonty *et al.,* 1978). In this way, genetic analysis of cells derived from medial buds generated in the central region of the zygote revealed recombinant mtDNA, whereas cells from end buds inherited one of the parental mtDNA types (Dujon *et al.,* 1974; Nunnari *et al.,* 1997; Berger and Yaffe, 2000). This mitochondrial vegetative segregation establishes homoplasmic lineages by a DNA recombination protein-dependent mechanism (Ling and Shibata, 2004; Shibata and Ling, 2007).

In the present study, we analyzed the mitochondrial *COX2* sequences from 532 strains, belonging to the seven pure *Saccharomyces* species and to five types of interspecific hybrids, to determine the extent of recombination in this gene. As a result, patent evidences of extensive interspecific mtDNA recombination are described for the first time in *Saccharomyces* yeasts and a putative mechanism to explain the presence of such recombination hotspot is proposed.

2. Material and Methods

2.1 Saccharomyces *strains and culture media*

Five hundred thirty-two strains from different origins and sources of isolation were used in this study (Table 1). The species assignations for these strains, confirmed by restriction analysis and/or sequencing of the 5.8S-ITS region (Fernández-Espinar *et al.,* 2000), are as follows: 1 strain to *S*. *arboricolus,* 2 *S*. *cariocanus,* 420 *S*. *cerevisiae,* 21 *S*. *paradoxus;* 2 *S*. *mikatae,* 9 *S*. *kudriavzevii,* 28 *S*. *bayanus* var. *uvarum* (or *S. uvarum),* and 15 *S*. *bayanus* var. *bayanus* (or *S*. *bayanus,* i.e. *S. eubayanus* x *S*. *uvarum* hybrids), 5 *S*. *pastorianus* (*S*. *cerevisiae* x *S*. *eubayanus* hybrids), 1 *S*. *cerevisiae* x *S*. *uvarum* hybrid, 25 *S*. *cerevisiae* x *S*. *kudriavzevii* hybrids,

Table 1. Geographic origin of *Saccharomyces* strains. Wine or non-wine sources of *Saccharomyces cerevisiae* strains are also included.

Species	Origin	Strains
S. cerevisiae		**420**
	European	130
	Africa	27
	Far-East	20
	America	227
	Unknown	16
	Wine	131
	Non-wine	289
S. paradoxus		**23**
	European	8
	Far-East	6
	America	9
S. mikatae	Far-East	**2**
S. arboricolus	Far-East	**1**
S. kudriavzevii		**9**
	Far-East	2
	Europe	7
S. uvarum		**28**
	Europe	26
	Unknown	2
S. cerevisiae x *S. kudriavzevii*	Europe	**25**
S. cerevisiae x *S. kudriavzevii* x *S. uvarum*		**2**
	Europe	*2*
S. cerevisiae x *S. eubayanus*		**5**
	Europe	4
	Unknown	1
S. cerevisiae x *S. uvarum*	Europe	**1**
S. eubayanus x *S. uvarum*		**15**
	Europe	7
	Unknown	8

and 2 *S*. *cerevisiae* x *S*. *kudriavzevii* x *S*. *uvarum* hybrids. One strain of *Lachancea castelli* was used as outgroup. Yeast strains were grown at 28°C in GPY medium (2% glucose, 0.5% peptone, 0.5% yeast extract).

2.2 PCR amplification and sequencing of *COX2* genes.

Total yeast DNA was extracted following the procedure described elsewhere (Querol *et al.,* 1992). The mitochondrial gene *COX2* was amplified by PCR, using the primers described in ref. (Belloch *et al.,* 2000). PCR products were cleaned with High Pure PCR Product Purification Kit (Roche Diagnostics, Mannheim, Germany) and both strands of the DNA were directly sequenced using the BigDyeTM Terminator V3.0 Cycle Sequencing Kit (Applied Biosystems, Warrington, UK), following the manufacturer's instructions, in an Applied Biosystems automatic DNA sequencer ABI 37301 (Applied Biosystems).

COX2 gene sequences obtained in this study were deposited into GenBank under accession numbers JN676363-JN676823. Other *COX2* gene sequences, not determined in this study, were retrieved from GenBank, where they are deposited under the accession numbers indicated in Table S1. Other sequences were obtained from the *Saccharomyces* Genome Resequencing Project by Blast searching (http://www.sanger.ac.uk/research/projects/genomeinformatics/sgrp.html).

2.3 Phylogenetic analysis and detection of recombination

COX2 sequences were aligned with MEGA5 (Tamura *et al.,* 2011) and classified into haplotypes by using DNASP v5.10 (Librado and Rozas, 2009). A median joining (MJ) network (Bandelt *et al.,* 1999) was constructed by using Network 4.5 (http://www.fluxus-engineering.com/).

Recombination analyses were performed with one representative sequence from each haplotype. Recombination points were defined by

using the RDPv3.44 package (Martin *et al.,* 2010). This program includes six methods to detect recombination: RDP (Martin and Rybicki, 2000), Bootscanning (Salminen *et al.,* 1995), MaxChi (Smith, 1992), Chimaera (Smith, 1992), GeneConv (Padidam *et al.,* 1999) and Sis-scan (Gibbs *et al.,* 2000). Common settings for all methods were to consider sequences as circular, statistical significances were set at the $P < 0.05$ level, with Bonferroni correction for multiple comparisons. Visual comparison of the polymorphic site distribution was also performed to confirm the results. Different recombination points were detected but two most frequent recombination sites divided the *COX2* gene into two segments, referred as 5'-end (positions 1-496 in the alignment or 124-620 in the reference *COX2* gene sequence from strain S288c, GenelD: 854622) and 3'-end (from position 497 to the end of the alignment, or from 621-708 in the reference S288c *COX2* gene sequence).

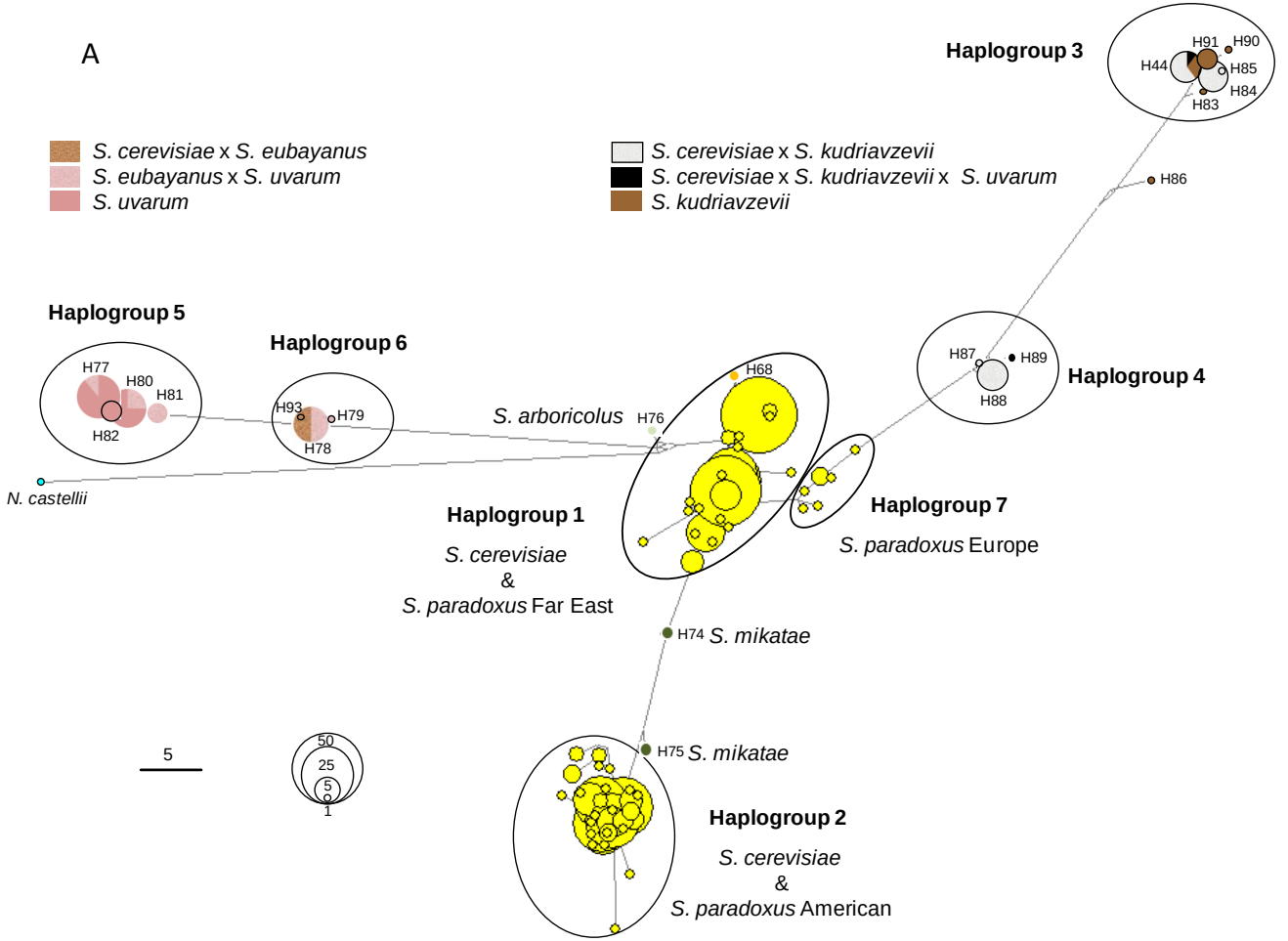
Phylogenetic networks were obtained from the two segments of *COX2* alignment with the Neighbor-Net method, with default settings, included in SPLITSTREE 4 package (Huson and Bryant, 2006). These two *COX2* segments were also used to obtain maximum-likelihood trees with the best suited models of nucleotide substitution defined according to jModeltest (Posada, 2008). Tree Puzzle v5.2 (Schmidt *et al.,* 2002) was used to test the phylogenetic congruence of the two ML trees with respect to a consensus *Saccharomyces* species tree topology obtained by combining phylogenetic trees from previous studies (Rokas *et al.,* 2003; Liti *et al.,* 2006; Wang and Bai, 2008; Scannell *et al.,* 2011). The statistical significance of these comparisons was assessed with the Shimodaira-Hasegawa (Shimodaira and Hasegawa, 1999) and ELW (Strimmer and Rambaut, 2002) tests.

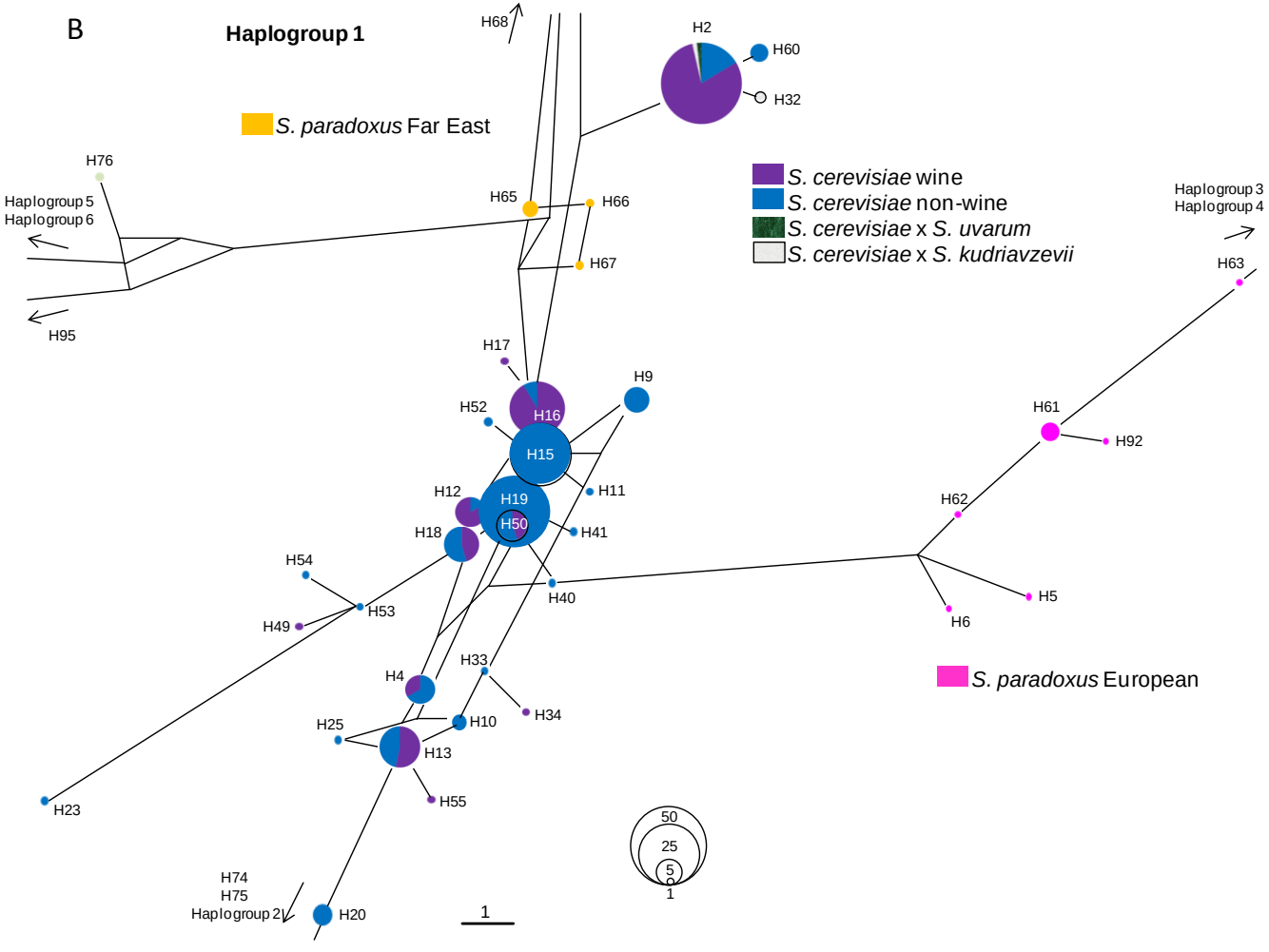
3. Results

3.1 COX2 variability in the Saccharomyces *yeasts.*

The 532 *COX2* coding sequences used in this study correspond to 93 different species-specific haplotypes. Their alignment yields 585 nucleotide positions, of which 98 are variable (16.8%), and of them 77 are phylogenetically informative (13.2%). Half of this nucleotide variability occurs in the 3' end of this gene (from positions 451-585, almost the last fifth of the alignment). Twenty-two variable sites correspond to 0-fold degenerated positions where non-synonymous substitutions occur, but only 7 of them are informative; 31 variable sites correspond to 2-fold degenerated positions, but only 5 of them show non-synonymous substitutions (2 informative and 3 singletons). Finally, 44 variable sites are 4-fold degenerated and 41 of them are informative. Two adjacent codons (positions 535-537 and 538-540) show several non-synonymous substitutions, in the second codon the corresponding amino acid replacements are unique (singletons), but in the first one they are convergent, and a "flip-flop" amino acid replacement pattern is observed. Two other codons (positions 541-543 and 556-558) show several informative and convergent substitutions corresponding to synonymous changes between the two codon families coding for Serine (TCN and AGY), and between the two codon families encoding Threonine (CTN and ACN), respectively (Table 2). The variability in these 3 codons, showing informative and convergent changes, classifies haplotypes in groups, as indicated in Table 2.

The 93 different *COX2* haplotype sequences were used to reconstruct their phylogenetic relationships by means of a MJ network analysis. Phylogenetic networks should be employed when reticulate events such as intra- or interspecific hybridization, horizontal gene transfer, recombination, or gene duplication and loss are believed to be involved (Huson and


A
S. cerevisiae x S. eubayanus
S. eubayanus x S. uvarum
S. uvarum
S. cerevisiae x S. kudriavzevii
S. cerevisiae x S. kudriavzevii x S. uvarum
S. kudriavzevii
Haplogroup 3
H91
H90
H44
H85
H84
H83
H86
Haplogroup 5
Haplogroup 6
H77
H80
H81
H82
H93
H79
H78
S. arboricolus
H76
H68
H87
H89
H88
Haplogroup 4
N. castellii
Haplogroup 1
S. cerevisiae
&
S. paradoxus Far East
Haplogroup 7
S. paradoxus Europe
H74 S. mikatae
5
50
25
5
1
H75 S. mikatae
Haplogroup 2
S. cerevisiae
&
S. paradoxus American


B
Haplogroup 1
S. paradoxus Far East
S. cerevisiae wine
S. cerevisiae non-wine
S. cerevisiae x S. uvarum
S. cerevisiae x S. kudriavzevii
S. paradoxus European
Haplogroup 3
Haplogroup 4
Haplogroup 5
Haplogroup 6
Haplogroup 2
H68
H2
H60
H32
H76
H65
H66
H67
H95
H17
H9
H52
H16
H15
H12
H19
H50
H11
H18
H41
H54
H53
H49
H40
H4
H33
H34
H25
H10
H13
H55
H23
H74
H75
H20
H63
H61
H92
H62
H5
H6
50
25
5
1
1

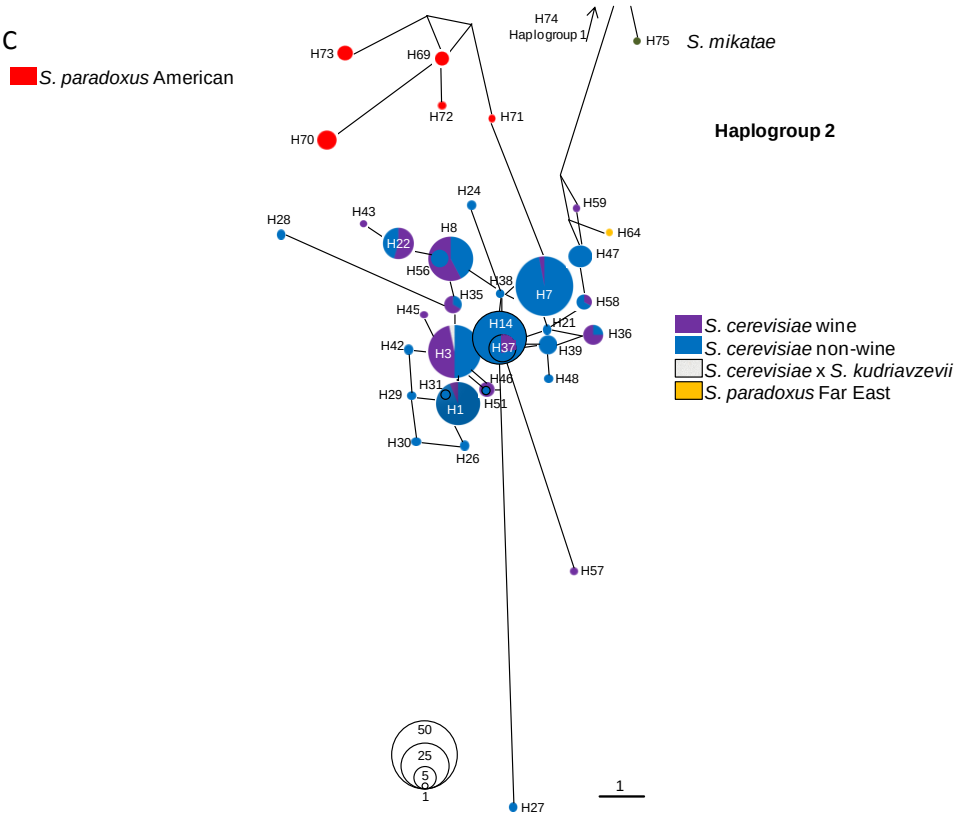

C
S. paradoxus American
H74
Haplogroup 1
H75
S. mikatae
H73
H69
H72
H71
H70
Haplogroup 2
H24
H59
H43
H28
H8
H64
H22
H47
H56
H38
H7
H58
H35
H45
H14
H21
H36
H42
H3
H37
H39
H31
H46
H48
H29
H1
H51
H30
H26
H57
H27
S. cerevisiae wine
S. cerevisiae non-wine
S. cerevisiae x S. kudriavzevii
S. paradoxus Far East
50
25
5
1
1

Figure 1A.- A median joining network of *COX2* mitochondrial gene alignment. Each haplotype is represented by a circle, with the area of the circle proportional to its frequency. Species or hybrid strains were indicated by different colors, with the exception of haplogroups 1 and 2 that were colored in Figures 1B and 1C, respectively. The length of each branch is proportional to the number of mutations on the respective branch.
Figure 1B.- A zoom in of median joining network part of Haplogroup 1.
Figure 1C.- A zoom in of median joining network part of Haplogroup 2.

Bryant, 2006). Seven haplogroups can be described according to the MJ network analysis (Fig. 1A). Relationships among haplotypes within Haplogroups 1 and 2 are given in more detail in Figures 1B and 1C, respectively.

In the MJ phylogenetic network based on *COX2* sequences, haplotype 76 (from *S. arboricolus*) occupies an ancestral position close to the root located by the outgroup *Naumovia castellii COX2* sequence. The first branch to diverge corresponds to haplogroups 5 and 6, which appear as derived and ancestral haplogroups, respectively. These haplogroups comprise haplotypes found in *S. uvarum* and the hybrid species *S. bayanus* (*S. eubayanus* x *S. uvarum* hybrids) and *S. pastorianus* (*S. cerevisiae* x *S. eubayanus* hybrids).

Table 2. Variable *COX2* nucleotide positions among haplotypes (H). Haplotypes are colored according to the color codes depicted in Figures 2. *COX2* regions are also colored according to their similarities. Hg corresponds to the haplogroup according to the phylogenetic network analysis (Fig. 1), and N stands for number of strains exhibiting each haplotype. The 3' end regions contain 3 codons (indicated in red) that are variable for their 3 positions (squared). The first codon is the only showing several informative and convergent non-synonymous substitutions (aminoacids A, N, Q, S or T), the second encodes Serine with a six-codon family, TCN (S1) or AGY (S2), and the third encodes Threonine with an eight-codon family, CTN (T1) or ACN (T2). The variability in these codons classifies haplotypes in groups according to their codon combinations, as indicated.

COX2 variable nucleotide positions (in vertical)

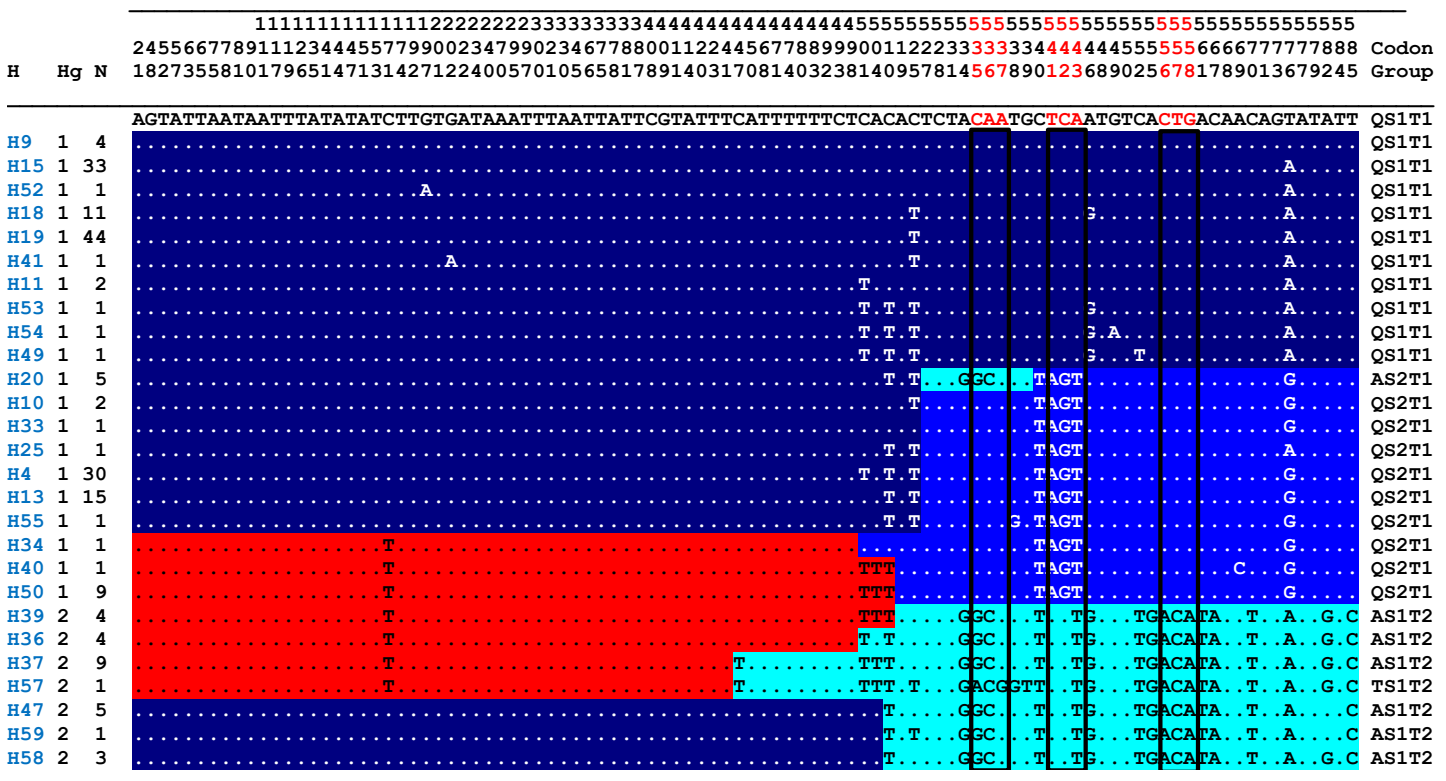


Table 2. Cont.

COX2 variable nucleotide positions (in vertical)

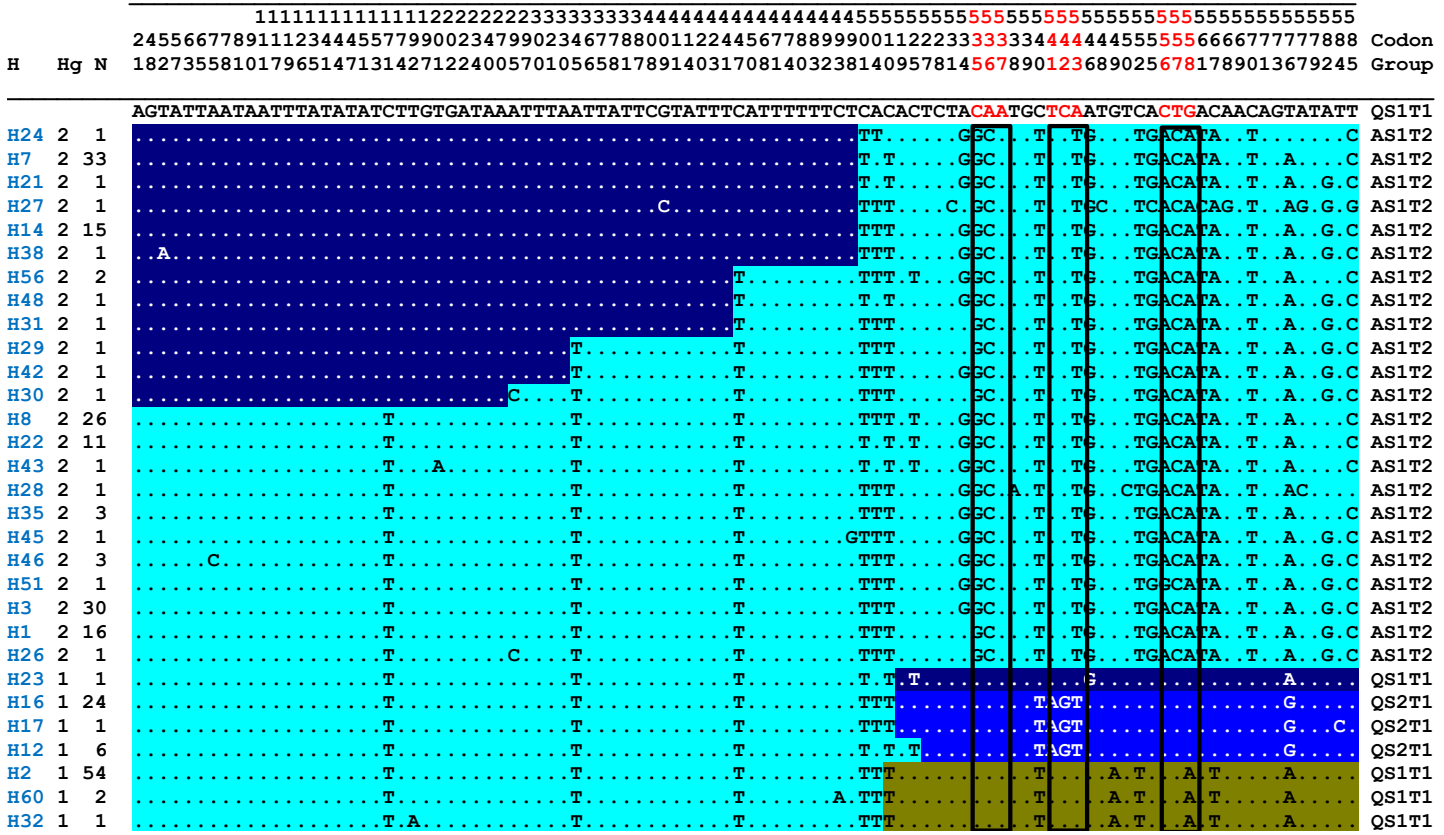


Table 2. Cont.

COX2 variable nucleotide positions (in vertical)

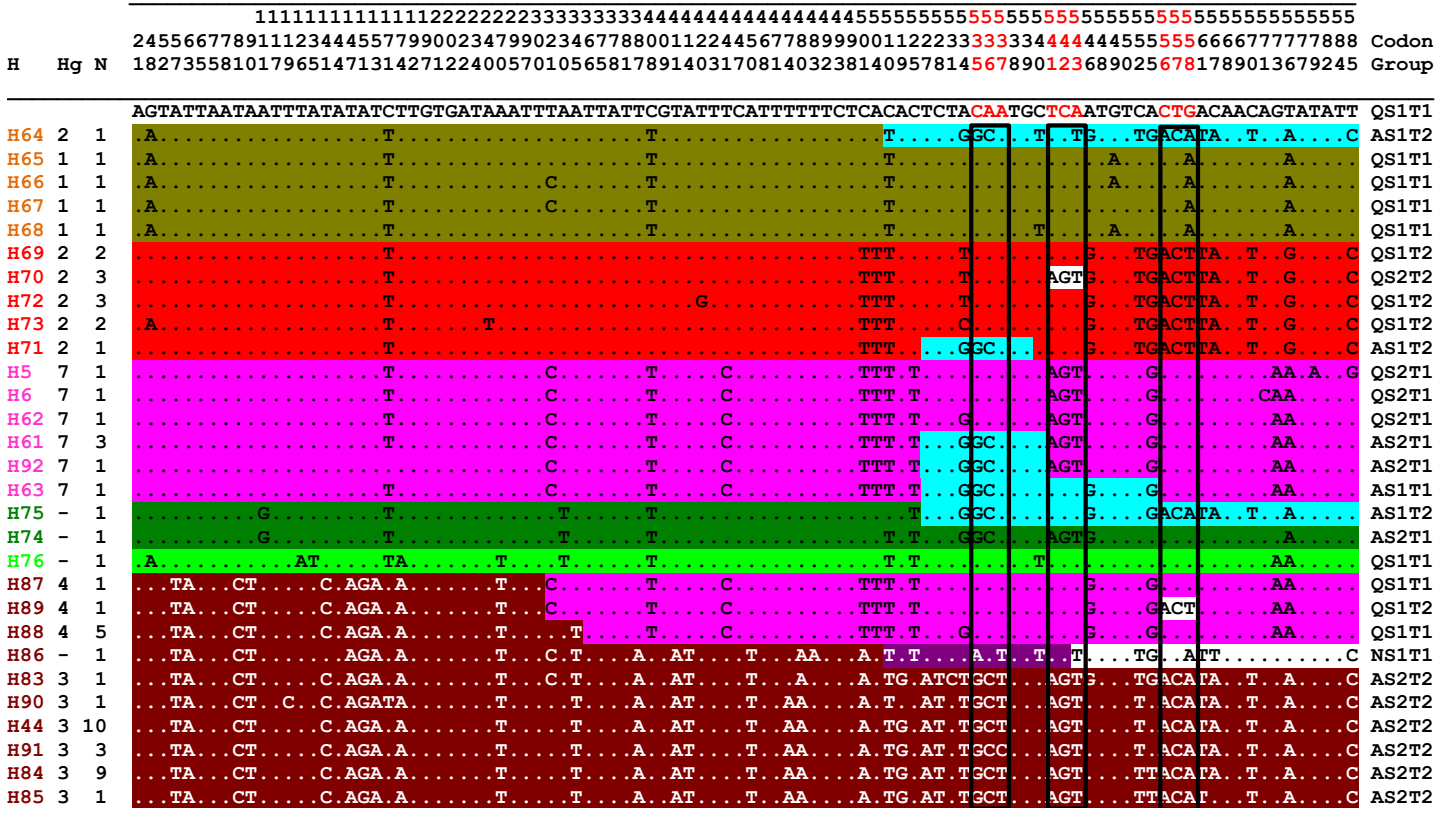


Table 2. Cont.

COX2 variable nucleotide positions (in vertical)

H	Hg	N	1827355810179651471314271224005701056581789140317081403238140957814567890123689025678	17890136792 45	Codon Group
			AGTATTAATAATTTATATATCTTGTGATAAATTTAATTATTCGTATTTCATTTTTTCTCACACTCTACAATGCTCAATGTCACTGACAACAGTATATT		QS1T1
H78	6	10	T....GTT...A..T..AG.TA.......T.AA.T...CA.T.A..A.TTG...AC....T.T...TA.T.....A.T.ACTCT..T..G.....		NS1T2
H79	6	1	T....GTT...A..T..AG.TA.......T.AA.T...CA.T.A..A.TTG...AC....T.T...TA.T.....A.T.ACTCT..T.TG.....		NS1T2
H93	6	1	T....GTT...A..T..AG.TA.......T.AA.T...CA.T.A..A.TTG...AC....T.T...TA.T.....A.T.ACTCT..T..G....C		NS1T2
H81	5	3	T..T.GTT......TCT.G..A....GC.T.AA.TTAATAAT.A..A.TTGC..AC....T.T...TA.T.....A.T.ACTCT..T..G.....		NS1T2
H77	5	17	T..T.GTT......TCT.G..A....GC.T.AA.TTAATAAT.A..A.TTGC..AC....T.T....TCT..T..T....T.ACTCT..T..G.....		SS1T2
H80	5	12	T..T.GTT......TCT.G..A....GC.T.AA.TTAATAAT.A..A.TTGC..AC....T.T....A.T..T.....T.ACTCT..T..G.....		NS1T2
H82	5	4	T..T.GTT......TCT.G..A....GC.T.AA.TTAATAAT.A..A.TTGC..AC....T.T....A.T..T..C...T.ACTCT..T..G.....		NS1T2

Then, Haplogroup 1 appears in a basal position with respect to the remaining haplogroups, located in 2 separated branches. Haplotypes included within haplogroup 1 can be divided in four subgroups (Fig. 1B). The first subgroup comprises haplotypes H2, H60 and H94, corresponding to 55 *S. cerevisiae* strains, mostly from wines, and 2 *S. cerevisiae* x *S. kudriavzevii* (UvCEG and IF6) and 1 *S. cerevisiae* x *S. uvarum* (S6U) hybrids. The second one, located in an intermediate position, includes 4 haplotypes (H65 to H68) exhibited by 5 S. *paradoxus* strains from Far East Siberia. The third subgroup includes the remaining 24 haplogroup 1 haplotypes, corresponding to 197 *S. cerevisiae* strains from different origins and sources of isolation, but mainly from non-wine sources. Finally, the fourth subgroup only includes haplotype H68, found in *S. paradoxus* strain CECT 11424 from Far East Russia, which is located in a separated position.

Closely related to haplogroup 1 appears haplogroup 7, in fact, its haplotypes could be included as a subgroup within haplogroup 1, but as they correspond to European *S. paradoxus* strains and are connecting haplogroup 1 to haplogroups 3 and 4, they were included in a different group. Haplogroups 3 and 4 (ancestral and derived, respectively) include haplotypes present in *S. kudriavzevii* strains and its hybrids *S. cerevisiae* x *S. kudriavzevii* and *S. cerevisiae* x *S. kudriavzevii* x *S. uvarum*.

Haplogroup 2 is the second main group, located in a branch separated from haplogroup 1 by haplotypes H74 and H75, present in the 2 *S. mikatae* strains under analysis. Haplogroup 2 (fig. 1C) comprises two subgroups of haplotypes, the first one includes haplotypes present in *S. cerevisiae* strains from diverse origins, but mainly from non-wine sources, and one S. *paradoxus* strain from Japan (haplotype H64). The second subgroup contains haplotypes found in American S. *paradoxus* (H69-H72) and *S. cariocanus* strains (H73).

As can be seen, there is no correlation between the phylogenetic relationships deduced from mitochondrial *COX2* gene sequences, and the taxonomic assignation of strains based on the restriction analysis and/or sequencing of the 5.8S-ITS region (data not shown), or the expected phylogenetic relationships among *Saccharomyces* species and populations (Liti *et al.,* 2006; Wang and Bai, 2008). This is relevant in the cases of *S. cerevisiae* and *S. paradoxus* strains, which are intermixed within two differentiated groups. In the case of *S. paradoxus* strains, they are grouped according to their geographic origin, Europe, Far East Siberia, Japan and America, but each geographic group is closer to different groups of *S. cerevisiae* strains. In this way, a group of S. *cerevisiae* strains are closely related to *S. paradoxus* from Far Eastern Siberia (haplogroup 1) and from Europe (haplogroup 7), and a second group of *S. cerevisiae* strains are closely related to a Japanese and the American *S. paradoxus* (haplogroup 2). Moreover, although Haplogroup 1 includes a higher number of American S. *cerevisiae* strains, there is no clear correspondence between the *S. cerevisiae* groupings and the strain origin (Figure S1) or source of isolation. Moreover, hybrids exhibiting *S. cerevisiae* COX2 sequences appear in both haplogroups: two S. *cerevisiae* x S. *kudriavzevii* strains IF6, from a dietary complement, and UvCEG, from wine, and one S. *cerevisiae* x S. *uvarum* strain S6U, from wine, are grouped within Haplogroup 1, but the hybrid S. *cerevisiae* x S. *kudriavzevii* AMH, from wine, is Included in Haplogroup 2.

3.2 Recombinant COX2 *sequences.*

Due to the incongruences between the *COX2* phylogenetic relationships and the expected phylogenetic relationships among *Saccharomyces* species and populations based on nuclear genes (Liti *et al.,* 2006; Wang and Bai, 2008), we investigated the possible presence of recombination signals in the *COX2* sequences. For this purpose we used

the different methods to detect recombination implemented in the RDP v3.44 program (see materials and methods), and the results of this analysis were confirmed by the visual inspection of the COX2 nucleotide variability (Table 2). In most cases, the identification of the recombinant sequences and their parental non-recombinant forms is difficult because one of the two recombinant derivatives can be absent in our large sample or because the putative parentals and their recombinants can show slight nucleotide differences either due to divergence since the recombination event or due to the absence of the real parental in the sample but a closely related haplotype to the parental is present.

According to the RPD analysis, many haplotypes are recombinant and in most of them a major recombination site is located between positions 501 and 525 of the *COX2* sequence alignment (Table 2), although in some cases the recombination site is located in other positions, e.g. haplotypes H87, H88 and H89. In most cases, recombinant segments extend beyond the limits of the COX2 gene but in a few cases the recombination segment finishes within the *COX2* gene (e.g. haplotype H63). Finally, in a couple of haplotypes (H20 and H86), two recombination events are postulated. Because, in most cases the recombinant segments extend beyond the limits of the *COX2* gene, the assignation of the recombinant and the parental forms (see Table 2) can be wrong, because the parental may be the recombinant and vice versa.

Recombination in *COX2* has been occurring at both the intraspecific, among *S. cerevisiae COX2* (different haplotypes), and the interspecific levels, between *S. cerevisiae* and *S. paradoxus* or between *S. cerevisiae* and *S. mikatae* (haplotype H75). It is remarkable the case of haplotypes H2, H32 and H60, on one hand, and H64, on the other, which correspond to the two closely related reciprocal recombinant forms resulting from a recombination event between *S. cerevisiae COX2* haplotypes from

haplogroup 2 and Far Eastern *S. paradoxus COX2* haplotypes from haplogroup 1.

It is also worth to note that in several cases the putative recombinant *COX2* haplotypes are present in interspecific hybrids of different nature. This way, the recombinant haplotypes H87 and H88 are present in *S. cerevisiae* x *S. kudriavzevii* hybrids (PB7 and 5 brewery hybrid strains, respectively) and haplotype H89 in the cider strain CID1 (a triple hybrid *S. cerevisiae* x *S. kudriavzevii* x *S. uvarum*). These haplotypes showed a 5' end sequence (alignment positions 1 to 300 or 350, depending on the haplotype) identical to *S. kudriavzevii COX2* sequences, and the 3' end is closely related not to *COX2* from *S. cerevisiae* (as expected), but to *COX2* from European *S. paradoxus* strains (Table 2). Another example comes from recombinant haplotype H81, exhibited by 5 *S. bayanus* strains corresponding to *S. eubayanus x S. uvarum* hybrids. The first 504 nucleotides of the *COX2* gene are identical to the *S. uvarum* haplotypes, but the rest of the gen is identical to haplotype H78 present in other *S. eubayanus* x *S. uvarum* (*S. bayanus*) and *S. eubayanus* x *S. cerevisiae* (*S. pastorianus*) hybrids, which likely corresponds to the *COX2* gene inherited from the *S. eubayanus* parent.

3.3 Phylogenetic networks of *COX2 segments*

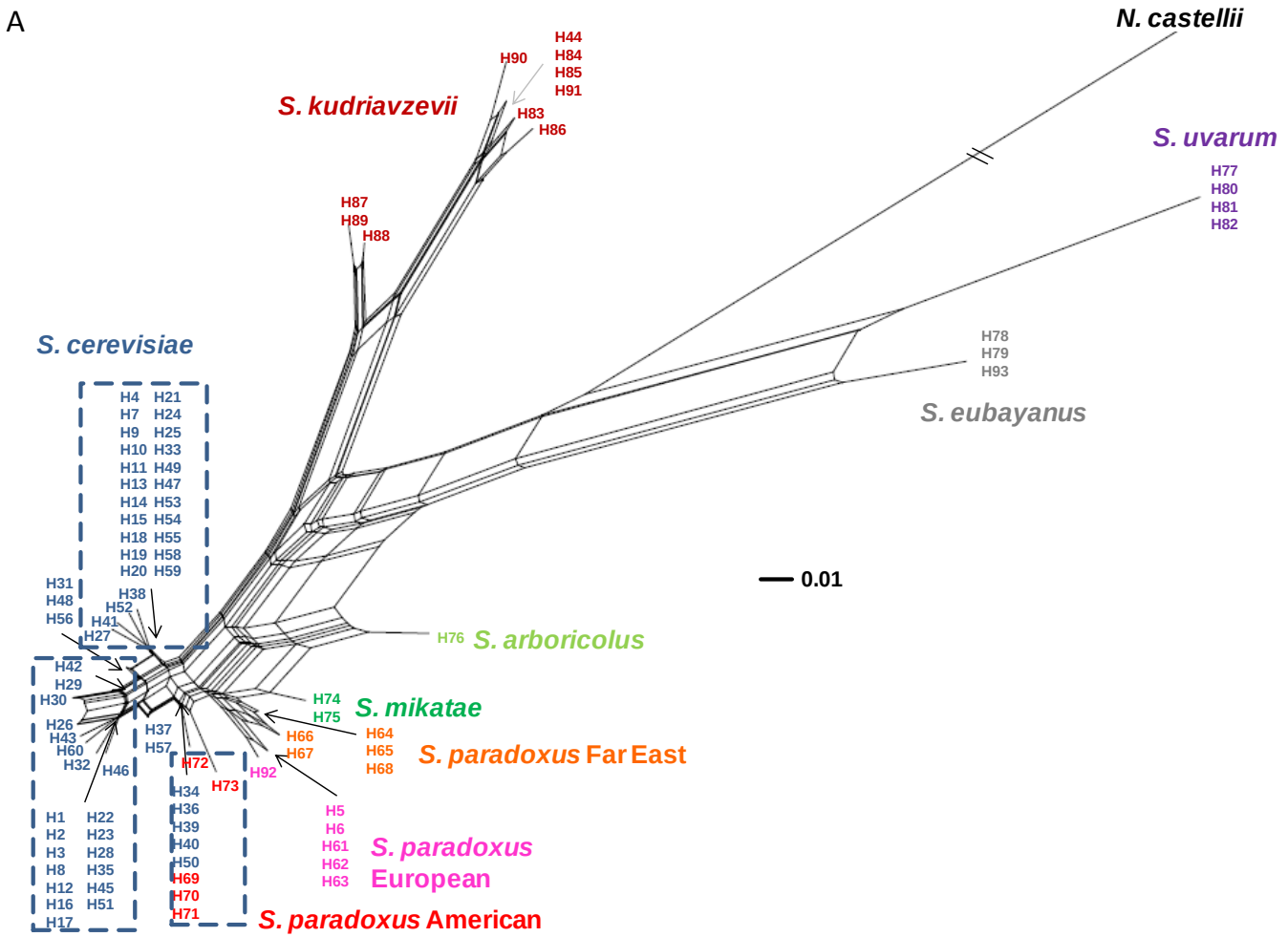
As a result of the recombination analysis, we detected a recombination hotspot present in most recombinant *COX2* haplotypes. Therefore, we constructed maximum-likelihood (ML) phylogenetic trees for each *COX2* segment: 5' end segment, from the beginning to the hotspot, and 3' end from the hotspot to the end. These partial *COX2* phylogenetic trees were compared against the expected topology based on the species tree derived from previous studies (Rokas *et al.,* 2003; Liti *et al.,* 2006;

Wang and Baig, 2008). These comparisons were performed with the Shimodaira-Hasegawa (SH) and Expected Likelihood Weight (ELW) tests.

The topology of the ML phylogenetic tree of the *COX2* 5'-end sequences was not significantly different from the species topology phylogenetic tree. However, the ML tree of the second segment (*COX2* 3'-end sequences) was significantly better than the tree with the topology of the species tree. As phylogenetic networks are better representations of the phylogenetic relationships when reticulated events, such as recombination, are involved in the evolution of the sequences, we obtain phylogenetic networks for each *COX2* segments (Fig. 2A and 2B).

The network of the first COX2 5'-end segment (fig. 2A) showed well defined groups of strains for each species. In the case of *S. eubayanus* haplotypes, they correspond to COX2 sequences from this species present in *S. eubayanus* x *S. uvarum* and *S. eubayanus* x *S. cerevisiae* hybrids. And in the case of *S. uvarum* and *S. kudriavzevii* haplotypes, they correspond to those found in pure strains of these species as well as in their hybrids that inherited *COX2* from these parentals. Haplotypes H87, H88 and H89 appear in a separated group within the *S. kudriavzevii* lineage, but this is due to the fact that their recombination sites are located within the 5'-end segment used for the phylogenetic network reconstruction.

In the case of *S. paradoxus*, strains are grouped according to their geographic origins: Europe, Asian Far East (Japan and Siberian Far East) and America. And in the case of *S. cerevisiae*, strains appear in three main groups (indicated by dashed squares) as well as additional lineages of sequences recombinant for this segment. These groupings are not correlated either with their geographic origins or with their sources of isolation. One of these groups, formed by haplotypes H34, H36, H39, H40 and H50, is the only exception to the species clustering because they are


A
N. castellii
S. kudriavzevii
H44
H84
H85
H91
H90
H83
H86
S. uvarum
H77
H80
H81
H82
H87
H89
H88
S. cerevisiae
H78
H79
H93
S. eubayanus
H4 H21
H7 H24
H9 H25
H10 H33
H11 H49
H13 H47
H14 H53
H15 H54
H18 H55
H19 H58
H20 H59
H31
H48
H56
H38
H52
H41
H27
0.01
H76 S. arboricolus
H42
H29
H30
H74
H75
S. mikatae
H26
H43
H60
H32
H37
H57
H46
H72
H73
H92
H66
H67
H64
H65
H68
S. paradoxus Far East
H34
H36
H39
H40
H50
H69
H70
H71
H5
H6
H61
H62
H63
S. paradoxus
European
H1 H22
H2 H23
H3 H28
H8 H35
H12 H45
H16 H51
H17
S. paradoxus American

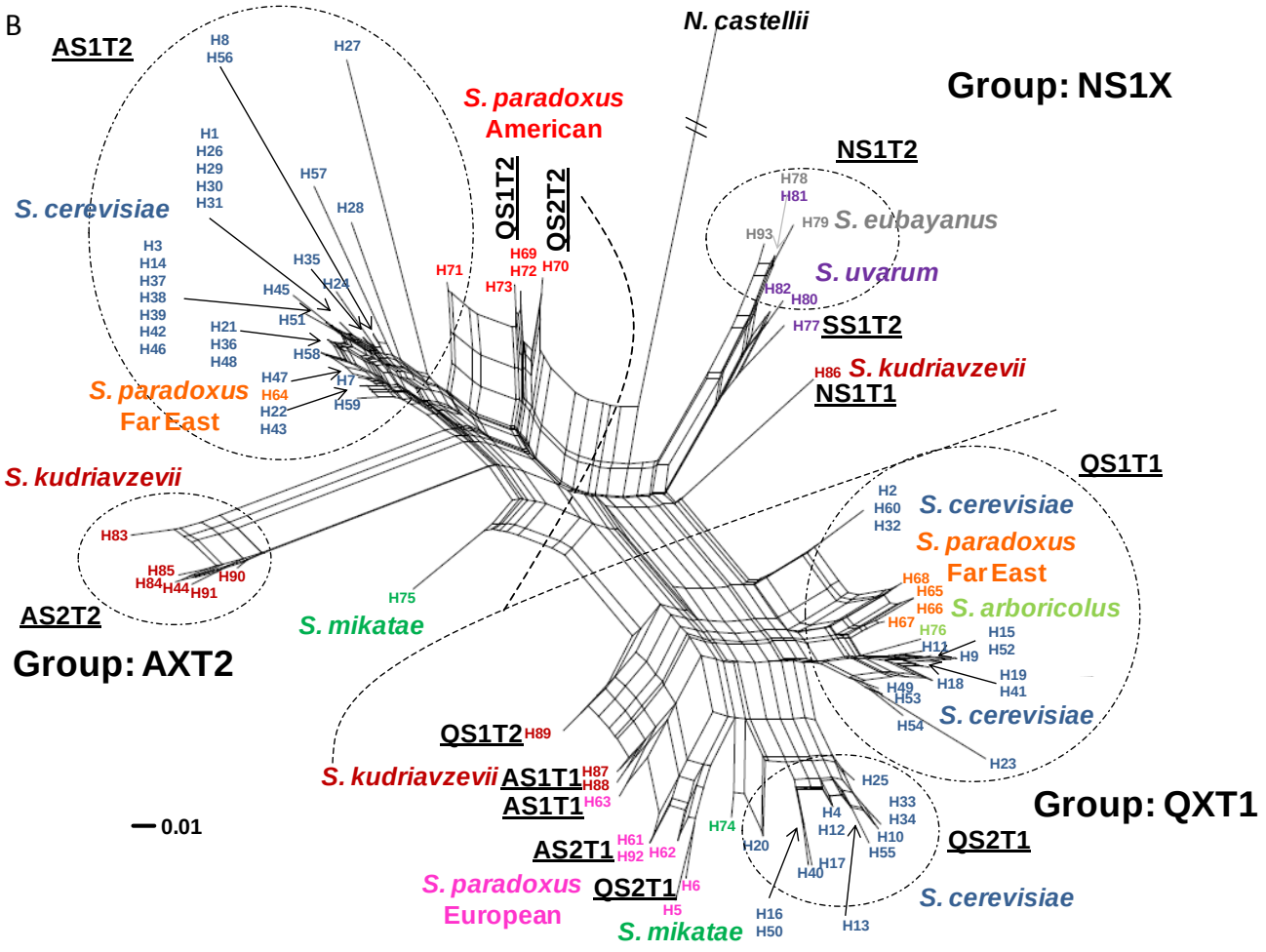

B
AS1T2
H8
H56
H27
N. castellii
Group: NS1X
S. paradoxus
American
QS1T2
QS2T2
NS1T2
H1
H26
H29
H30
H31
H57
H28
S. cerevisiae
H78
H81
H79 S. eubayanus
H93
S. uvarum
H3
H14
H37
H38
H39
H42
H46
H35
H24
H45
H51
H21
H36
H48
H58
H71
H69
H72
H73
H70
H82
H80
H77 SS1T2
H86 S. kudriavzevii
NS1T1
S. paradoxus
Far East
H47
H64
H22
H43
H7
H59
QS1T1
S. kudriavzevii
H2
H60
H32
S. cerevisiae
S. paradoxus
Far East
H83
H85
H84
H44
H91
H90
H68
H65
H66
S. arboricolus
H67
H76
H15
H52
H9
H18
H19
H41
H49
H53
H54
S. cerevisiae
AS2T2
H75
S. mikatae
Group: AXT2
QS1T2 H89
H23
S. kudriavzevii
AS1T1 H87
H88
AS1T1 H63
H25
H33
H34
H4
H12
H10
H55
H74
H20
H17
H40
QS2T1
AS2T1 H61
H92
H62
S. paradoxus
European
QS2T1 H6
H5
S. mikatae
H16
H50
H13
S. cerevisiae
Group: QXT1
0.01

Figure 2A.- Phylogenetic network of the *COX2* first part alignment (1-496). Haplotypes are colored according to the specie included in them. The length of each edge is proportional to the weight of the associated split, this is analogous to the length of a branch in a phylogenetic tree.
Figure 2B.- Phylogenetic network of the *COX2* second part alignment (497-End). Haplotypes are colored according to the specie included in them. The length of each edge is proportional to the weight of the associated split, this is analogous to the length of a branch in a phylogenetic tree.

grouped with the American S. *paradoxus* haplotypes. This tree is congruent with the putative recombinant nature of these haplotypes (see Table 2).

The phylogenetic network of the second segment (Fig. 2B), corresponding to the *COX2* 3'-end sequences, shows radically different relationships due to the recombination events. This tree is congruent with the sequence polymorphisms of this region shown in Table 2.

Haplotypes are clustered in three main groups (indicated in Fig 2B) that are more or less correlated to the combinations of those codons that showed a 'flip-flop' variability (see Table 2). Exceptions to these groups seem to be due to putative recombination events leading to intermediate positions of these recombinant haplotypes.

The first group (indicated as AXT2) comprises a subgroup of *S. cerevisiae* strains (codon combination type AS1T2), in which the recombinant haplotype H64 from the Japanese *S. paradoxus* strain is included, as well as the American *S. paradoxus,* most *S. kudriavzevii* haplotypes (type AS2T2) and the *S. mikatae* recombinant haplotype H75.

The second group (QXT1) includes three different S. *cerevisiae* strain clusters. One is closely related to a subgroup formed by European *S. paradoxus* haplotypes and related recombinant haplotypes from *S. kudriavzevii* hybrids. And the second and third clusters are closely related to Far East *S. paradoxus* haplotypes as well as to the non-recombinant S*. mikatae* haplotype H76.

Finally, the third group (NS1X) includes three lineages, that of the non-recombinant haplotypes of *S. eubayanus* together with the

recombinant haplotype H81, the non-recombinant haplotypes of *S. uvarum*, and the lineage of the recombinant haplotype H86 from the Japanese *S. kudriavzevii* strain IFO1803. The 5' end region of this haplotype (from the beginning to the hotspot) is similar to *S. kudriavzevii COX2* sequences, but part of the 3'-end segment (positions 510 to 542) is identical to *S. uvarum* haplotypes 80 and 82, and the rest (543 to the end) different to any other haplotype (unknown origin).

4. Discussion

Seven species (*S. cerevisiae, S. paradoxus, S. cariocanus, S. mikatae, S. arboricolus, S. kudriavzevii* and *S. bayanus*) have been described in the genus *Saccharomyces*, according to the biological species concept (Naumov *et al.,* 1995a,b, 2000; Wang and Bai 2008). The boundaries between the species of genus *Saccharomyces* are unclear. Previous results have shown that hybrids, *S. cerevisiae* x *S. eubayanus* (some of them known as *S. pastorianus*) and *S. cerevisiae* x *S. kudriavzevii*, are frequent in nature (de Barros Lopes *et al.,* 2002, González *et al.,* 2006, 2008). Natural hybrids between *S. cerevisiae* and *S. paradoxus* have not been discovered, although introgressions have been (Liti *et al.,* 2006, Muller and McCusker, 2009). No natural hybrids between *S. cerevisiae* and *S. mikatae* have been found. However an horizontal gene transfer and introgressions between them has been described (Liti *et al.,* 2005, Dunn *et al.,* 2012).

In the present study we have analyzed the phylogenetic relationship among the different species, strains and their hybrids from genus *Saccharomyces*, using the mitochondrial gene *COX2.* The study of mitochondrial gene is of interest because mitochondrial genes are haploid and they show high variability. *COX2* gene sequence has been used

previously to carry out phylogenetic analysis in ascomicetous yeasts (Belloch *et al.,* 2000, Kurtzman and Robnett 2003).

COX2 gene has a high variability in the *Saccharomyces* genus, we detected up to 94 different haplotypes in the seven species. This variability is not translated to amino acids changes which were lower (27 different *COX2* aa sequences). Different authors have proposed a population structure for *S. cerevisiae* (Liti *et al.,* 2009, Schacherer *et al.,* 2009). We did not find a clear distribution of strains using *COX2* sequences. According to continental isolation our strains were equally distributed around the haplogroup 1 and haplogroup 2. Taking into account the isolation source, only a clear distribution was identified in the laboratory and bakery strains which were enclosed in haplogroup 2. Liti and collaborators (2009) defined five pure subpopulations of *S. cerevisiae,* according to the continental/isolation source. Our results supported the wine/European, West African and Malaysian pure groups which strains where enclosed in one of the haplogroups (Haplogroup 2), the only exception was L1528 (H2) which is in ambiguous position between *S. paradoxus* from Far-East (H65-H68). North American and Sake groups are not well supported. In the case of North American we found that YPS606 (H23) is enclosed in haplogroup 1 and YPS128 (H29) is in haplogroup 2. K11 strain (H1) in Sake group was found in a different haplogroup than Y9 (H33) and Y12 (H34). These results are indicative that North American is not a pure group and Sake could be a pure group, not considering K11 in it. The Liti *et al.,* 2009 study was done using a monosporic culture and heterozigosity information is lost, for this reason strains that apparently are pure, probably are not .

In Schacherer *et al.,* (2009) study the *S. cerevisiae* strains were grouped according to the isolation source. They found three different groups: wine, sake and laboratory strains. Our results supported the Sake group with Y9 and Y12 in Haplogroup 1. Our laboratory strains were found

in Haplogroup 2, and wine/nature/clinical group, used in Schacherer *et al.,* (2009) study, were all of them found in Haplogroup 2 as well.

In the case of *S. paradoxus* strains Liti *et al.,* (2009) described three different populations. In our study we found similar results: America (enclosed in Haplogroup 2), Far-East (enclosed in Haplogroup 1) and European (Haplogroup 7). The only exception was the 11152JP (H26) that appears closest to *S. cerevisiae* strains, in Haplogroup 2.

Mitochondrial inheritance is reported to be uniparental (Basse, 2010). After the mating between two different cells, the daughter cell receives one of the parental mitochondrial. Most of the hybrids have a *COX2* sequence similar to one of the parental strains involved in the hybridization. In the double *S. cerevisiae* x *S. kudriavzevii* hybrids, most of them have the *S. kudriavzevii COX2* gene sequence, with the exception of AMH which inherited the *S. cerevisiae* mitochondrial genome (H3), where *S. cerevisiae* wine strains were also found (Peris *et al.,* 2012a), and IF6 and UvCEG, found between Far East *S. paradoxus* haplotypes. In the case of *S. eubayanus* x *S. uvarum* hybrids and *S. cerevisiae* x *S. eubayanus* they inherited one of the two *COX2* sequences enclosed to Haplogroup 5 or Haplogroup 6. A debate is around the differentiation of *S. bayanus* varieties. Some authors have proposed to consider *S. bayanus* var. *bayanus* and *S. bayanus* var. *uvarum* as different species (Rainieri *et al.,* 2006, Perez-Traves *et al., in preparation*) where other consider them as varieties (Nguyen and Gaillardin, 2005). Our results showed two well defined *S. bayanus COX2* sequences, Haplogroup 5 and 6. Number of nucleotide differences among Haplogroup 5 and 6 were similar to nucleotide differences between *S. arboricolus* and Haplogroup 1 (S. cerevisiae). This is also supported by the recent description of a pure strain, called *S. eubayanus*, postulated as the parental strain of non-*cerevisiae* subgenome in the hybrid *S. pastorianus* (Libkind *et al.,* 2011).

During mating, after the fusion of two different yeast cells, the fusion of mitochondria organelles is followed to form a continuous reticulum denominated heteroplasmic state (Berger and Yaffe, 2000). When the daughter cell is in the middle of the zygote (medial buds), the new mitochondria could inherit mtDNA from both parental cells, in the form of recombinant products, previously described in *S. cerevisiae* strains (Nunnari *et al.,* 1997; Berger and Yaffe, 2000). Our results showed different groups, based in the last polymorphic aminoacid positions, of *S. cerevisiae COX2* sequences clustering with different populations of *S. paradoxus* and other species are in ambiguous position between two different species. The existence of introgressions, HGT and hybrids between different species (Liti *et al.,* 2005 and 2006, Rainieri *et al.,* 2006, Gonzalez *et al.,* 2006 and 2008, Peris *et al.,* 2012a) are indicative that the species of genus *Saccharomyces* are able to produce hybrids with high frequency. In some cases the hybrid strains are stable, maintaining chromosomes copies from the two parental, as *S. cerevisiae* x *S. bayanus* and *S. cerevisiae* x *S. kudriavzevii* hybrids (Dunn and Sherlock, 2008, Belloch *et al.,* 2009, Peris *et al.,* 2012b), but in other are not found hybrids, probably must to genetic incompatibilities, as *S. cerevisiae* x *S. paradoxus* and *S. cerevisiae* x *S. mikatae*. Hybridization between different species is not rare because they are found in the same ecological niche (Sampaio and Gonçalves, 2008). However, the reasons for having a stable or not stable allopolyploid genome must to be studied. In cases where non stable hybridizations have occurred, parts of the missing genome could be maintained as a footprint of this ancestral hybridization event.

Some aminoacid positions appear to show a flip-flop pattern increasing the complexity of the analysis. This pattern are obtained when an aminoacid position revert to the ancestral state due to positive selection.

This kind of pattern has been described in virus (Botosso *et al.,* 2009) and it may be confirm in *COX2* gene.

The existence of a common recombination point is indicative that a molecular process could be involved in the recombination between the two different mtDNA genomes more than a random process. In the same *COX2* transcription unit and next to *COX2* gene, taking some nucleotides of the *COX2* 3' end, is encoded a gene known as RF1 or ORF1 (a maturase-related gene) (Bordonné *et al.,* 1988). Maturases are encoded in homing endonucleases genes (HEGs) which are selfish genetic elements that spread by first cleaving chromosomes that do not contain them and then getting copied across to the broken chromosome as a byproduct of the repair process (Burt and Koufopanou, 2004). We speculate that ORF1 could be a homing endonuclease active in some strains or species. ORF1 could be involved in the recombination of the two different mtDNA genomes, after the fusion of the parental mitochondrial organelles. The sequencing of ORF1 and other genes from mtDNA could shed light if our hypothesis could be truth and how far is the recombination sequences extended.

In conclusion, the boundaries between different species of *Saccharomyces* genus are unclear. Although some species like *S. paradoxus* showed a population structure in *COX2* sequences, it is not clear in *S. cerevisiae* strains. Moreover hybridization between the different species of *Saccharomyces* genus gave us results that increase the complexity of the studies, such as introgressions, HGT and recombinant sequences. These results show that hybridization events between species of *Saccharomyces* genus are really frequent and could have an evolutionary advantage. We demonstrated that *COX2* gene is not a proper gene to show the evolution of different species of *Saccharomyces* genus but it has information about ancestral hybridizations.

Author's contributions

EB conceived and supervised this study. DP and AA contributed equally to the experimental analysis. DP wrote the first version of the manuscript. DP and EB participated in the final manuscript revision.

Acknowledgements

This work was supported by Spanish Government grant AGL2009-12673-C02-02 and Generalitat Valenciana grants PROMETEUS and ACOMP/2012. DP acknowledges to the Spanish Government (MICINN) his FPI fellowship. AA received PROMEP Fellowship from SEP, Mexican government.

Chapter 2

Introgression in the mitochondrial *ORF1* gene as a footprint of ancestral hybridization events in the genus *Saccharomyces*

David Peris[1], Sandi Orlić[1,2], Carmela Belloch[3] and Eladio Barrio[3,4*]

[1]Cavanilles' Institute of Biodiversity and Evolutionary Biology, University of Valencia, Parc Científic, P.O. Box 22085, E-46071 Valencia, Spain

[2]Institute Ruđer Bošković, Center for Marine Research, G.Paliage 5, Rovinj, Croatia.

[3]Department of Biotechnology, Institute of Agrochemistry and Food Technology (CSIC), Valencia, Spain

[4]Departament de Genètica, Universitat de València, Spain.

*Corresponding author:

Dr. Eladio Barrio, Instituto de Agroquímica y Tecnología de Alimentos, CSIC. Lab 309. Avda. Agustín Escardino, 7. E-46980 Paterna, Valencia, Spain. Phone: +34 963 900 022 ext. 2309, Fax: +34 963 636 301.

E-mails: DP: david.peris-navarro@uv.es, SO: Sandi.Orlic@uv.es, CB: belloch@iata.csic.es, EB: Eladio.Barrio@uv.es

Abstract

The *Saccharomyces* genus is comprised of a complex group of species. Natural hybrids have been found involving *S. cerevisiae* and other species, including *S. cerevisiae* x *S. bayanus*, *S. cerevisiae* x *S. kudriavzevii*, and triple hybrids of *S. cerevisiae* x *S. bayanus* x *S. kudriavzevii*. Natural hybrids between *S. cerevisiae* x *S. paradoxus* or *S. cerevisiae* x *S. mikatae* have not been described, although footprints from ancestral unstable hybridizations have been found, as cases of introgression or horizontal gene transfer. Most of these events involved nuclear genomes. Recombination among mitochondria has previously only been described at the intraspecific level between *S. cerevisiae* strains. In this study, we show that the previously described *COX2* recombination hot spot could be due to the activity of a homing endonuclease gene (HEG), *ORF1.* HEGs are selfish elements which are spread quickly in the population, and when spread could involve different species being marked as a footprint of ancestral hybridizations. We describe transfers and recombination events involving *ORF1* between different species of *Saccharomyces*, and infer ancestral unstable hybridization. These findings suggest that species from *Saccharomyces* genus are frequently hybridizing, in most cases, such as wild environments, they are unstable and in biotechnological environments they could be stable. Biotechnological environments and just when conditions are not proper for parental strains, hybrids could be maintained in nature due to their better adaptation. We postulate that *ORF1* gene could be a functional homing endonuclease and its transfer to one species to another an indication of ancestral hybridization.

Keywords: *Saccharomyces* genus, LAGLIDADG homing endonuclease, recombination, mitochondrial genome, yeast evolution, hybridization.

1. Introduction

Several species are enclosed into the *Saccharomyces* genus: *S. cerevisiae, S. paradoxus, S. cariocanus, S. mikatae, S. arboricolus, S. kudriavzevii* and *S. bayanus*, the latter includes the varieties *uvarum* and *bayanus* (Kurtzman, 2003; Naumov *et al.,* 2000, 2010; Wang and Bai, 2008). Although a recent study provides strong evidence that *S. bayanus* var. *bayanus* and *S. bayanus* var. *uvarum* (*S. uvarum)* are genetically and ecologically isolated sister species from two distinct lineages, being the former called *S. eubayanus* (Libkind *et al.,* 2011), we continue using the varieties names here as in the last version of "The Yeast: A Taxonomic Study" (Kurtzman *et al.,* 2011), and following the Biological Species Concept (BSC) (Mayr, 1942), where species are considered to be units reproductively isolated from other such units, but within which interbreeding and genetic recombination reduce divergence.

Species from *Saccharomyces* genus are able to form hybrids. Natural double (*S. cerevisiae* x *S. bayanus*, *S. cerevisiae* x *S. kudriavzevii*) and triple (*S. cerevisiae* x *S. bayanus* var. *uvarum* x *S. kudriavzevii*) hybrids have been identified in alcoholic beverages, dietary supplements and clinical patients (Masneuf *et al.,* 1998; Naumova *et al.,* 2005; González *et al.,* 2006, 2008; Le Jeune *et al.,* 2007; Peris *et al.,* 2012a). Hybrids show sterility, with less than 5% viable ascospores (Naumov *et al.,* 1997). At least three postzygotic barriers contribute to reproductive isolation between *Saccharomyces* species. First, activation of the mismatch repair system by sequence divergence between two parental genomes prevents the crossovers necessary for proper chromosome segregation (Chambers *et al.,* 1996). Multiple reciprocal translocations also lead to aneuploidy meiotic progeny (Delneri *et al.,* 2003). Recently, multiple reciprocal cases of cytonuclear incompatibility were reported between *S. cerevisiae* and its relatives, *S. bayanus* and *S. paradoxus* (Lee *et al.,* 2008; Chou *et al.,* 2010). Thus, the existence of these hybrids

suggests that evolution of the *Saccharomyces* genus is more complex than expected.

Despite their close relationships, natural hybrids between *S. cerevisiae* x *S. paradoxus* and *S. cerevisiae* x *S. mikatae* have been not described. Nonetheless, some *S. cerevisiae* strains contain introgressed *S. paradoxus* sequences as footprints of ancestral hybridizations (Liti *et al.,* 2006; Muller and McCusker, 2009), while other strains contain a selfish genetic element horizontally transferred from *S. mikatae* (Liti *et al.,* 2005). Since, horizontal gene transfer is rare in yeast (Dujon *et al.,* 2010), those footprints could be indicative of ancestral unstable hybridizations. These reticulate events have all involved the transfer of nuclear genes into *S. cerevisiae* from other species.

Mitochondrial recombination occurs readily in yeast (Dujon *et al.,* 1974; Birky *et al.,* 1982; Taylor 1986; MacAlpine *et al.,* 1998), but it has only been described between *S. cerevisiae* strains. The GC cluster and A+T tandemly repeated sequences appear to be involved in the initiation of recombination and rearrangements of the mitochondrial genome (Dieckmann and Gandy, 1987; Skelly and Clark-Walker, 1991; Bouchier, 2009). Other genetic elements are also involved in mitochondrial recombination, such as homing endonuclease genes (HEGs) (Nakagawa *et al.,* 1992), a type of selfish genetic element (Burt and Koufopanou, 2004). The first HEG discovered in *S. cerevisiae* mitochondrial was ω and called I-SceI (Dujon *et al.,* 1974). HEGs can transfer its genetic sequence in a HEG^- genome (Colleaux *et al.,* 1986). The protein encoded in the HEG mediates a double-strand break (DSB) in the genome with a HEG^- allele. The HEG^+ allele is used, by the recombinational repair system, to repair the break. This gene conversion mechanism replaces the HEG^- allele with the HEG^+ allele (Burt and Koufopanou 2004). The *Saccharomyces* mitochondrial genome contains several active HEGs (Sargueil *et al.,* 1991; Séraphin *et al.,* 1992). Several additional open reading frames in the mitochondrial genome are hypothesized to encode homing

endonucleases, such as *ORF1*, a free-standing endonuclease (Séraphin *et al.,* 1987). In the *Saccharomyces* genus, *ORF1* is located 19 nucleotides upstream from 3' end of *COX2.* In *S. cerevisiae ORF1* gene is interrupted by GC clusters and probably inactivated as occurred in *ORF3* (*ENS2* or Endo.SceI). However, GC clusters were not detected in the *S. bayanus* var. *uvarum ENS2* gene, suggesting *ENS2* is active (Séraphin *et al.,* 1987, Nakagawa *et al.,* 1991) and after on demonstrated (Nakagawa *et al.,* 1992).

We previously showed that the *COX2* gene contains a recombination hot spot and hypothesized that the *ORF1* homing endonuclease could be involved (Peris *et al., in preparation*). Here, we report the sequences of the *ORF1* gene and *COX3* gene of different species and strains of *Saccharomyces* genus and compare their sequences with *ORF1* homologs in *Kazachstania* (*Saccharomyces*) *servazii* and *Williopsis saturnus* var. *suaveolens*. These data support the plausible involvement of *ORF1* in the *COX2* recombination hot spot and some evolutionary scenarios for *ORF1* are hypothesized.

2. Material and methods

2.1 Yeast strains

Seventy-two *Saccharomyces* strains from a previous work (Peris *et al., in preparation*) were selected as representative strains of the different groups described (table S1). Forty-seven *Saccharomyces cerevisiae*, eight *Saccharomyces paradoxus*, 1 *Saccharomyces mikatae*, 1 *Saccharomyces arboricolus,* three *Saccharomyces kudriavzevii,* 1 *Saccharomyces bayanus* var. *uvarum* (*S. uvarum*), four *S. bayanus* var. *bayanus (S. eubayanus*) x *S. bayanus* var. *uvarum* (*S. uvarum*) hybrids and seven *S. cerevisiae* x *S. kudriavzevii* hybrids. Yeast strains were grown at 28ºC in GPY medium (2% glucose, 0.5% peptone, 0.5% yeast extract).

2.2 PCR amplification and sequencing of ORF1 and COX3 genes

DNA was extracted using the methodology developed by Querol *et al.,* (1992). The mitochondrial gene *ORF1* was amplified by PCR, using a primer walking approach. Primers were designed using IDT Scitools (http://eu.idtdna.com/SciTools/SciTools.aspx?cat=DesignAnalyze). Primer pairs used in each strain to amplify the different portions of *ORF1* and *COX3* are listed in (table S2). Primer sequences and conditions are listed in table S3. PCR products were cleaned with High Pure Product Purification Kit (Roche diagnostics, Manheim, Germany) and both strands of the PCR product were directly sequenced using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, California, U.S.A.) in an Applied Biosystems (Model 310) automatic DNA sequencer. Sequences were edited and assembled with Staden Package v1.5 (Staden *et al.,* 2000). The new sequences were deposited under the GenBank accession numbers (JN709044-JN709115).

2.3 ORF1 and COX3 alignments

ORF1 and *COX3* sequences from other *S. cerevisiae* strains, not sequenced in this study, were obtained applying a blastn search in the Saccharomyces Genome Re-sequencing Project (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/s_cerevisiae_sgrp). A PSI-Blast search was run to obtain *ORF1* sequences from non-*Saccharomyces* species. Accession numbers of these sequences are listed in (table S2). *ORF1* sequences were aligned using MUSCLE (Edgar 2004) and were further refined by visual inspection in Jalview 4.0.b2 (Waterhouse *et al.,* 2009). *COX3* was directly aligned using ClustalW (Thompson *et al.,* 1994), implemented in MEGA 4.0 (Tamura *et al.,* 2007). *COX3* from *Kluyveromyces lactis* was used as an outgroup.

Tandem repeat sequences in *ORF1* and *COX2* genes were searched using Tandem Repeat Finder software (Benson 1999). *ORF1* domains were

annotated in Jalview according to previous description by Dalgaard *et al.,* (1997) and using Conserved domain tool in NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Marchler-Bauer *et al.,* 2009). WebLogo profiles of LAGLIDADG and NUMOD1 domains were done in WebLogo 2.8.2 tool (http://weblogo.berkeley.edu/.), using a representative sequence from each *ORF1* haplotype (including non-*Saccharomyces* strains). GC insertions found in *ORF1* sequences were classified according to de Zamaroczy and Bernardi *et al.,* (1986). *COX2* accesion numbers are listed in table S2.

2.4 Haplotype classification.

The sequences of this study were selected as representative sequences based on previous work done with *COX2* sequences by Peris *et al.,* (*in preparation*). The new sequences from *ORF1* and *COX3* were classified according to their haplotypes. Haplotype classification was done in DnaSP v5 (Librado and Rozas 2009).

2.5 Phylogenetic analysis and detection of recombination points

Recombination points were defined using RDPv3.44 (Martin *et al.,* 2010). Six methods were used to detect the recombination points: RDP (Martin and Rybicki, 2000), Bootscanning (Salminen, 1995), MaxChi (Smith 1992), Chimaera (Smith 1992), GeneConv (Padidam *et al.,* 1999) and Sis-scan (Gibbs *et al.,* 2000). For all methods, we considered the sequences as circular and set statistical significance at the $P<0.05$ level with Bonferroni correction for multiple comparisons. Similar results were also achieved using GARD method (Pond *et al.,* 2006), implemented in Datamonkey (Delport *et al.,* 2010). Visual comparison of the polymorphic sites at amino acid level was also done to confirm the results.

ORF1 alignment length of 1253 nucleotides, where GC insertions and indels were removed, was used in the next analyses. A phylogenetic network of *ORF1* alignment was constructed using the Neighbor-Net method with default settings, as included in SPLITSTREE 4 package (Huson and Bryant, 2006). The *COX3* sequence evolution model that fits our sequence data best was optimized using the corrected Akaike Information Criterion (AICc) with a BioNJ tree as initial tree, implemented in jModelTest program (Posada 2005). The best fitting model of evolution for *COX3* sequences was TIM2+G (Posada 2009) with a gamma distribution (G) of substitution rates with a shape parameter α= 0.673. The parameters of the model, estimated in the previous analysis, were used to obtain the best trees under optimality criterion of Maximum-likelihood (ML) (Posada, 2003). Tree reliability was assessed using non-parametric bootstrap resampling of 100 replicates. Phylogenetic analyses were performed using PhyML program (Guindon *et al.,* 2010).

The most frequent recombination points were used to define four alignment segments for phylogenetic analyses. The first one is a concatenated sequence taking from 621 nucleotide position in *COX2* gene until the 246 nucleotide position in *ORF1* alignment sequence (corresponding to nucleotide 292 in *ORF1* gene of S288c, AJ011856). The last nineteen nucleotides of *COX2* gene are the first nucleotides of *ORF1* gene (being a segment of 224 nucleotides length). The second takes from 247 to 644 nucleotides, in *ORF1* alignment (from 293 to 704 in S288c *ORF1*). The third was built from 645 to 920 (706-980 in S288c), and the forth from 921 to the end of the alignment (981-1435 in S288c annotation).

2.6 Detection of selection

The single likelihood ancestor counting (SLAC), fixed effects likelihood (FEL), and random effects likelihood (REL) methods (Pond and Frost, 2005) available at the Datamonkey website were used to detect the signatures of

selection operating on *ORF1* protein gene. Different alignments were used to describe selection signatures. In the first approach, a complete *ORF1* alignment (without GC Insertions and indels), previously analyzed with GARD to describe the four partitions, was used. In a second survey, the three different domains were analyzed independently. Codons, under positive or negative selection, analyzed by three methods (SLAC, FEL and REL) and significantly described by two of the three methods were considered as positives. Phylogenetic relationships between *ORF1* gene sequences were inferred with the REV substitution model and phylogenetic trees were reconstructed by the NJ method. Codon-specific selection pressure along the sequences (i.e. site specific dN-dS) was measured and *p*-values were estimated at each site.

3. Results

3.1 ORF1 *gene structure*

To determine the extent of *ORF1* diversity within the *Saccharomyces*, we determined the sequence of an additional 36 *ORF1* genes, including previously unsampled species, resulting in a collection of 72 *ORF1* sequences. All strains have an *ORF1* gene, suggesting it is shared by all *Saccharomyces*. Among these strains, 51 different haplotypes were found (Table S2). *ORF1* start codon is nineteen nucleotides inside the 3' end of *COX2* gene. The average GC-composition of the *ORF1* is 18%. Eleven strains have GTG as *ORF1* start codon (uncommon start codon in mitochondria): haplotypes M2-M7 and M9-M12, while the translation of *ORF1* gene into protein predicts that fourteen strains have premature stop codons: haplotypes M2, M3, M7-M11, M17-M20, M23 and M46. *ORF1 s*equence length range from 1363 nucleotides (ZA17 strain) or 454 amino acids to 1516 nucleotides (VRB strain) or 505 amino acids. Note that we sequenced a partial *ORF1*

gene, 45 nucleotides left comparing to the complete *ORF1* gene of reference strain S288c.

Differences in size between *ORF1* genes were due to the presence of GC insertions and AT repeats. Seven different GC clusters insertion points were found along the *ORF1* alignment (fig. 1). The first GC cluster is found in VRB *ORF1* sequence. Three different types were found in the GC cluster 2, which are in CBS435, CECT 11757 and 120M. CBS 10644 have the third GC cluster. VRB displayed another GC cluster in the fourth GC cluster insertion point. CBS 435 showed a cluster in the fifth insertion point which was similar in structure to the GC cluster of CBS 10644 in the third insertion point (Figure S1). The most number of GC cluster were found in the sixth (40 *S. cerevisiae* strains and one *S. cerevisiae* x *S. kudriavzevii* hybrid), and in the seventh insertion points (36 *S. cerevisiae*, 2 *S. cerevisiae* x *S. kudriavzevii,* 2 *S. paradoxus* Far East and 2 from America) (Figure S2). The CBS 435 GC cluster 6th was oriented in an opposite direction to the other strains (Figure S1). We found three different subtypes for GC cluster 6 and six for GC cluster 7. GC cluster in 120M and CBS 5313 has the same nucleotide sequence than CBS 435, YPS606 and Y9.

Following a previous structure description and classification done by de Zamaroczy and Bernardi (1986), we were able to classify the new GC clusters found in the *ORF1* sequence, with the exception of GC cluster 2. GC cluster 1 and 4 are similar to a1 family, and GC cluster 3 and 5 were similar to a4 family. In the case of GC cluster 1, 2, 4, and 5 are on the opposite strand. As Séraphin *et al.,* (1987) and Weiller *et al.,* (1989) described, the GC clusters in the *ORF1* gene were flanked by TAG and AGGAG, or CTA and CTCCT if cluster was in the other strand (Figure S1). These conserved nucleotides were flanked by A+T rich sequences. Flanking sequences TAG and AG (CTA and CT) are conserved in most of the sequences with and without GC clusters. All GC clusters in *ORF1* belong to group M1 (Weiller, 1989).

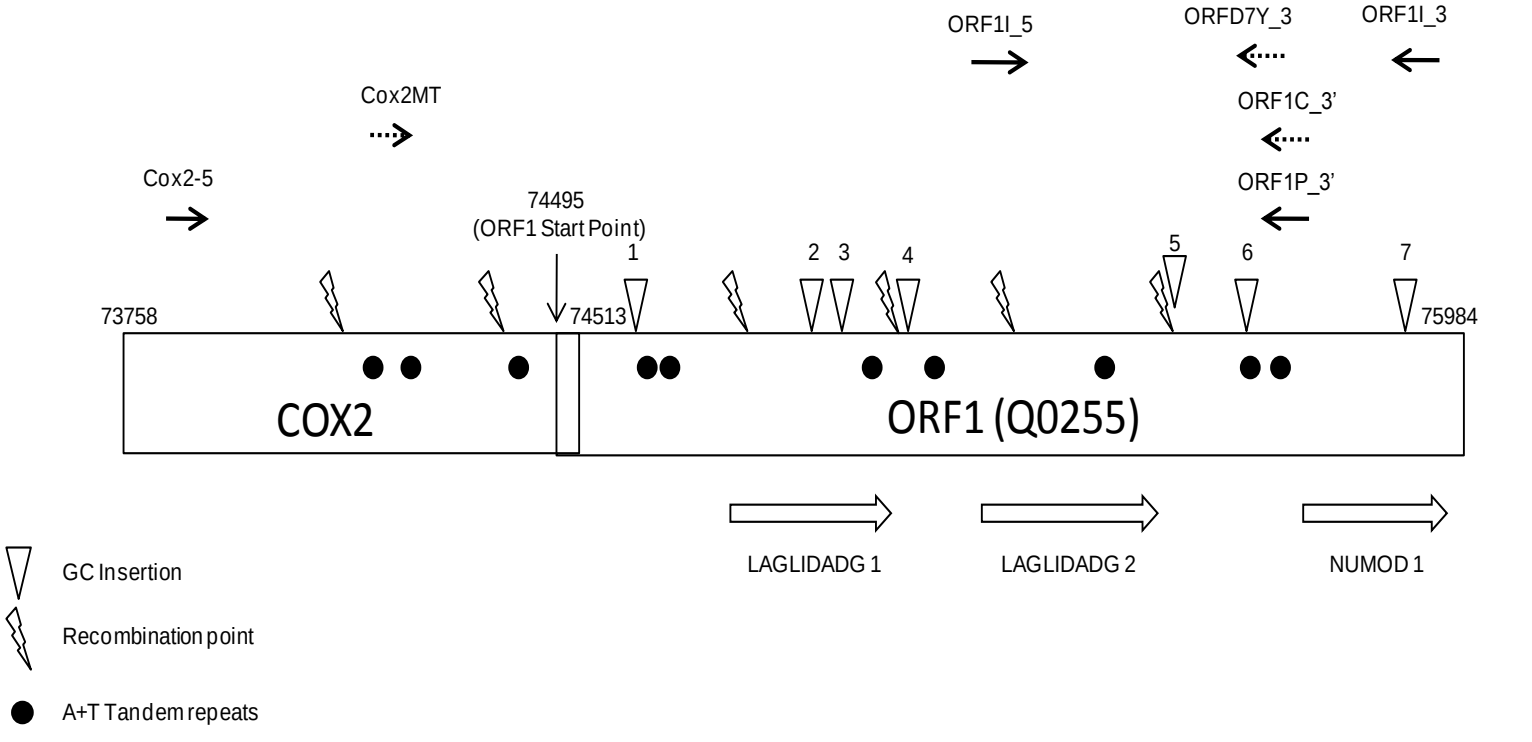


Figure 1 *COX2–ORF1* alignment. A schematic representation of *COX2-ORF1* alignment is shown. Black arrows and dotted arrows represent the primer pairs (also see Table S3). Empty arrows represent the three domains found in *ORF1*. LAGLIDADG 1 and 2 corresponds to previously described P1 and P2. GC clusters, A+T tandem repeats and recombination points detected in the previous work and in this study are drawn in the figure. Nucleotide coordinates correspond to the S288c *S. cerevisiae* reference strain.

On the other hand, differences in size were also due to the presence of AT repeats. Twenty-one different A+T rich sequences that repeated at least twice were found in the *ORF1* alignment (Table S4). The length of A+T rich tandem repeats ranged from three nucleotides to twenty five nucleotides. The most repeated sequence (AAT) was repeated ten times in haplotypes M1, M12, M22, M39, M40, and M50. The A+T rich tandem repeats were located near to GC clusters (fig. 1). In *COX2* gene, we found three A+T rich sequences repeated twice (Table S4).

Three different domains were annotated in all of the sequences, two LADGLIDADG (P1 and P2) and one NUMOD1 (fig 1). The alignment comparison of our sequences and the two homing endonucleases from *Williopsis saturnus* var. *suaveolens* (*ORF1* and *ORF3*), and one from *Kazachstania servazii* (SasefMp08) is showed in the three domains (Figure S3A). *Saccharomyces* strains showed low structural conservation along the three different *ORF1* domains, only in NUMOD1 was near to 50%. In LAGLIDADG 2 we could visually describe two different structures, that we called *ORF1* type I and type II (Figure S3B-D).

To analyze the role of selection, we subjected 417 codons (*ORF1* alignment without GC Clusters and indels) of the total 458 codons to several tests using Datamonkey. These analyses did not reject the neutral evolution model for most codons. A few codons (61 or 15%) were found to be under purifying selection, 43 of which were inside the LAGLIDADGs and NUMOD1 domains (Figure S3A).

3.2 ORF1 *and* COX3 *phylogenetic networks*

To determine how *ORF1* sequences were related, we constructed an *ORF1* neighbor-joining tree using *K. servazii SasefMp08* and *W. saturnus* var. *suaveolens ORF1* and *ORF3* as outgroups. The tree did not match the species phylogeny (Kurtzman and Robnett 2003; Rokas *et al.,* 2003) (Figure

S4). The phylogenetic network of *ORF1* showed two clearly separated groups (fig 2). Type I comprised most haplotypes of different species (*S.* cerevisiae, American *S.* paradoxus, *S.* arboricolus, *S.* mikatae, three *S. cerevisiae* x *S. kudriavzevii* hybrids and *S. bayanus*). Type II included most of the *S. cerevisiae* haplotypes and one *S. paradoxus* from Far East (M51, CECT 11152). The placement of the remaining strains of *S. cerevisiae* was ambiguous*,* as was the placement of *S. paradoxus* from Europe, some *S. paradoxus* from Far East (M19, M20) and the *S. kudriavzevii* haplotypes, including in the latter some *S. cerevisiae* x *S. kudriavzevii* hybrids (tables S1-S2). In contrast, the maximum likelihood tree of *COX3* gene recapitulated the species phylogeny (Kurtzman and Robnett 2003; Rokas *et al.,* 2003), with the exception of American *S. paradoxus* haplotypes (fig. 3). *S. paradoxus* 120M was identical to *S. cerevisiae* haplotype C17.

3.3 Recombination points in ORF1

The low bootstrap values of some branches in *ORF1* phylogenetic tree and the conflict with the species phylogeny could indicate the presence of recombinant sequences or gene transfer from one species to another. To investigate whether some sequences might be recombinant, we used RDP3 and GARD software to partition the *ORF1* alignment. Indels and GC clusters were excluded for analysis. And visual inspection was done to confirm the segments. RDP3 detected up to four different recombination points; all were near A+T rich sequences or GC clusters, with the exception of the third recombination point, which was located in the beginning of LAGLIDADG 2 domain (fig. 1). At least one recombination event involved the haplotypes located in ambiguous position in the previous phylogenetic network (M8-M12, M19, M20, M24-M29, M31-M33, M36 and M42). Four partitions were found using GARD, only the second disagreed with RPD3. The trees inferred from each partition were significantly incongruent each other

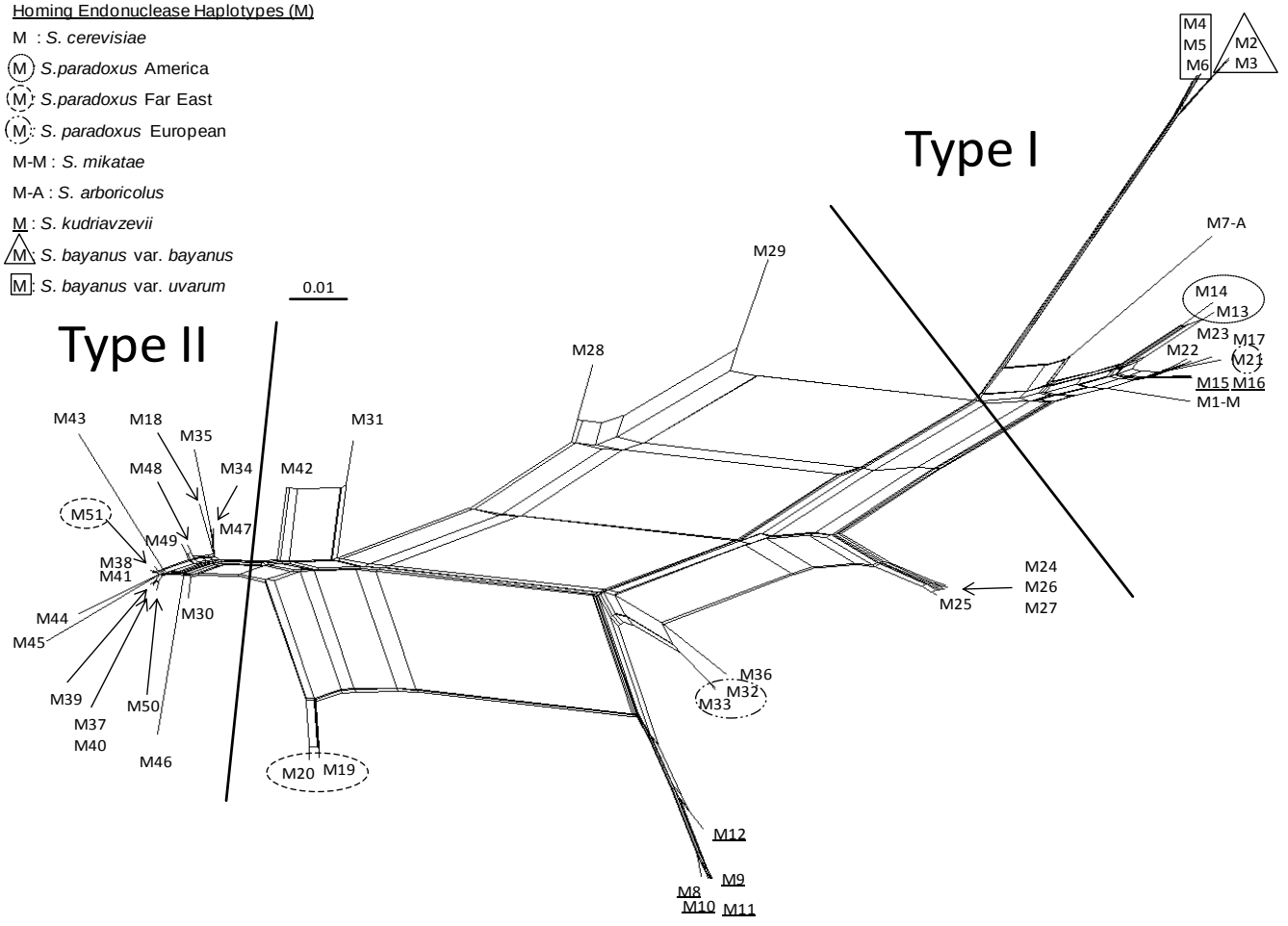

Homing Endonuclease Haplotypes (M)
M : S. cerevisiae
(M) S.paradoxus America
(M) S.paradoxus Far East
(M) S. paradoxus European
M-M : S. mikatae
M-A : S. arboricolus
M : S. kudriavzevii
M S. bayanus var. bayanus
M S. bayanus var. uvarum
0.01
Type I
Type II
M4
M5
M6
M2
M3
M7-A
M29
M14
M13
M23
M17
M22
M21
M15
M16
M1-M
M28
M43
M18
M35
M31
M48
M34
M42
M51
M49
M47
M38
M41
M44
M45
M30
M24
M26
M27
M25
M39
M50
M37
M40
M46
M36
M32
M33
M20
M19
M12
M9
M8
M10
M11

Figure 2 Neighbour-Net phylogenetic network of *ORF1* gene. The length of each edge is proportional to the weight of the associated split, this is analogous to the length of a branch in a phylogenetic tree. Haplotypes for *ORF1* gene are represented in this figure, after remove GC cluster and indels. Hypothetical active homing endonucleases are indicated with a symbol (*).

using Kishino-Hasegawa test (KH) (Kishino & Hasegawa, 1989). ΔAIC_c for the best model (3 breakpoints) was 143.724, and for each breakpoint the p-value was lower than 0.01.

To identify the recombinant haplotypes and minor and major parents, we constructed one phylogenetic network for each of the partitions inferred by GARD (Figure S5). The phylogenetic network of the first partition (3'end *COX2-ORF1,* see M&M section) (Figure S5A) showed two different groups, the first of which (Type I) displayed seven subtypes where different *Saccharomyces* species were included. The second group (Type II) included most of the *S. cerevisiae* strains (haplotypes: M37-M46, M48-M50) and one strain of *S. paradoxus* from Far East (M51).

In the phylonetwork inferred using the second partition, the two types were more clearly separated. Interestingly, several haplotypes have different positions in the network, including to some *S. cerevisiae*, *S. paradoxus* from Europe (54 and CECT 1939), *S. paradoxus* from Far East, *S. kudriavzevii* from previous Subtype I-5, and strains in Subtypes I-1 and I-7 (Figure S5B). For example, *S. cerevisiae* haplotypes M18, M30, M35, M43 and M47 changed their affinities from Type I to Type II (Subtype II-1). Curiously, Y12 (M31), 54 and CECT 1939 (M32 and M33) were circumscribed in a new Subtype II-2 with YIIc17 (M42). Subtypes I-5 and I-3 are now circumscribed into the new Subtype II-3. Subtypes I-1, I-6, and I-7 have merged in one. Haplotypes M24-M27 and M36 were found in an ambiguous position.

The third partition phylonetwork again showed haplotypes in different positions (Figure S5C). *S. cerevisiae* haplotypes M24-M27 and M36, previously in an ambiguous position, and now enclosed in Subtype II-2 with *S.*

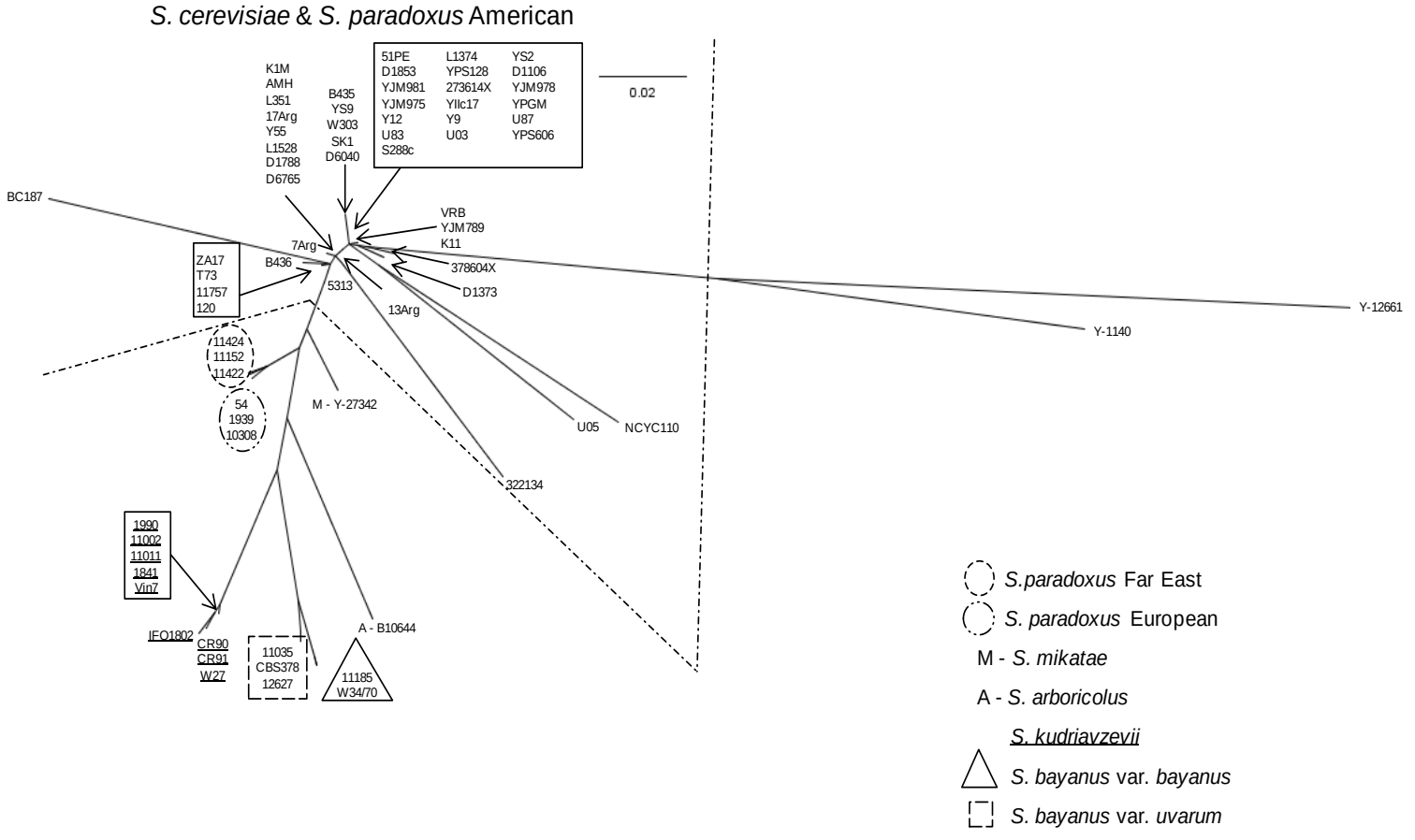


Figure 3 Maximum likelihood tree of partial *COX3* gene sequence. The scale is given in nucleotide substitution per site. Dotted lines separate *S. cerevisiae* and American *S. paradoxus* from the other sequences.

paradoxus from Europe (54, CECT 1939) and two *S. cerevisiae* M31 and M42. *S. cerevisiae* YPS606 (M28) was located in Subtype II-1. *S. paradoxus* from Far East were located in a separate subtype (Subtype II-4) between Subtypes II-1 and II-3.

In the fourth phylonetwork reconstruction, we recovered Type I subtypes with the exception of Subtype I-3 (Figure S5D). Subtype I-6 included the *S. cerevisiae* haplotypes M24-26, which were located in Subtype II-2 in the third partition. In Type II, the Subtypes II-1, II-2 and II-4 were maintained, but Subtype II-2 also contained several *S. cerevisiae* haplotypes (M18, M28, M35, M47 and M49).

In summary, we detected at least four different major recombination points (Table 1). In the case of M24-M27 and M36, we also detected a recombination point in the middle of the second partition of the alignment (Table 1), leading to their ambiguous placement in the second partition phylonetwork (Figure S5B). Thus, there is substantial evidence for each possible type of recombination between and within the two main types (Type I x Type I, Type I x Type II, and Type II x Type II).

4. Discussion

4.1 ORF1 *an active homing endonuclease*

The *Saccharomyces* genus includes seven species: *S. cerevisiae, S. paradoxus, S. cariocanus, S. mikatae, S. arboricolus, S. kudriavzevii* and *S. bayanus*, according to the biological species concept (Naumov *et al.,* 1995a, b, 2000; Wang and Bai, 2008). The boundaries between the *Saccharomyces* species are fuzzy, and many sterile hybrids have been found, such as *S. cerevisiae* x *S. bayanus, S. cerevisiae* x *S. kudriavzevii* (de Barros Lopes *et al.,* 2002; González *et al.,* 2006, 2008; Peris *et al.,* 2012a). Although no natural hybrids have been described between *S. cerevisiae* x *S. paradoxus*

Table 1. A representative sequence of each haplotype of the *ORF1* Homing Endonuclease alignment sequence is shown. Only polymorphic sites are displayed and colored according to similarity.

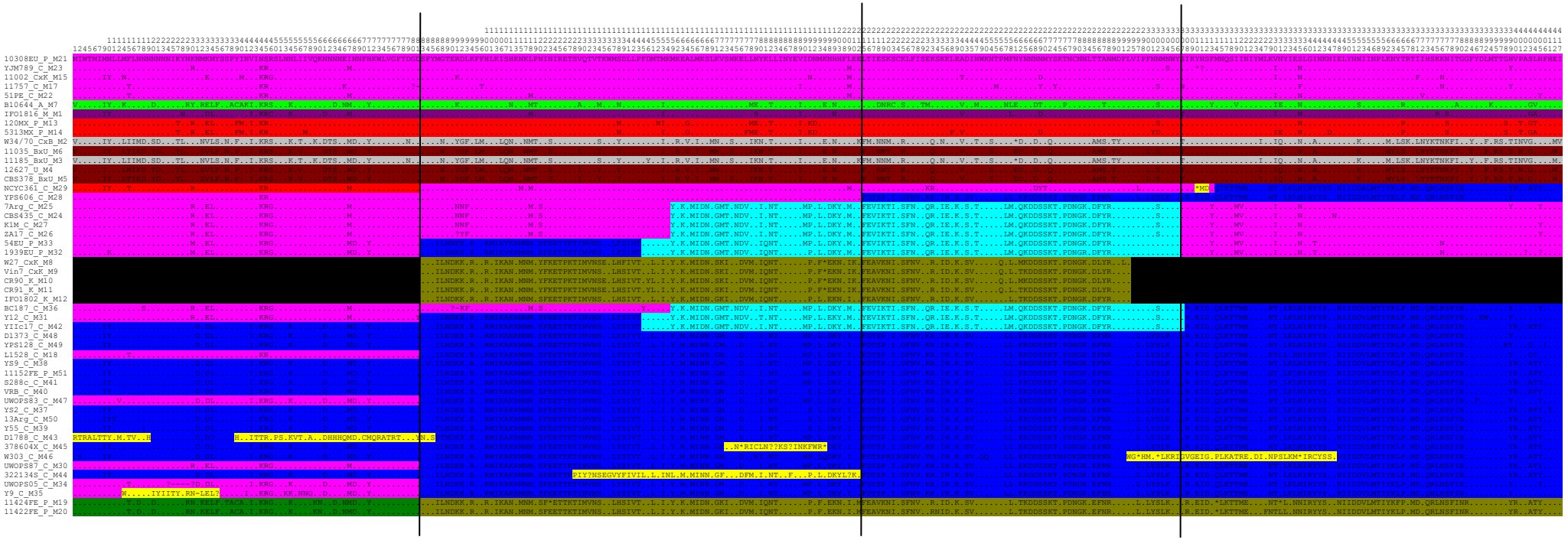
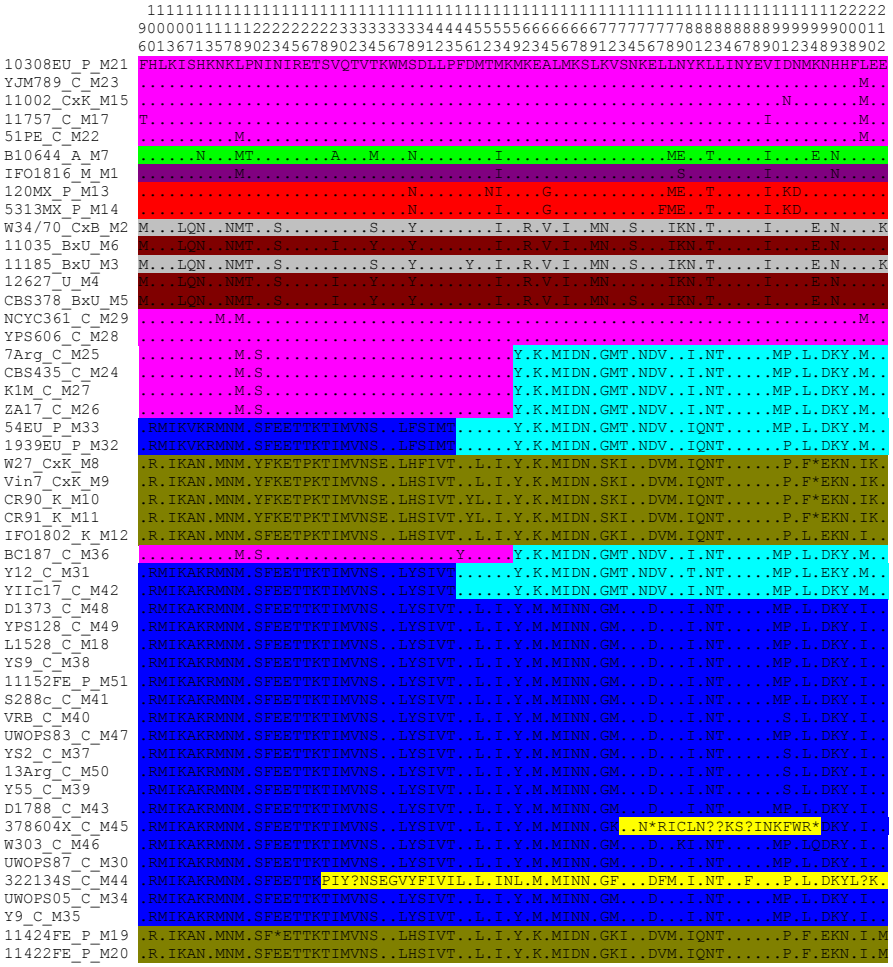
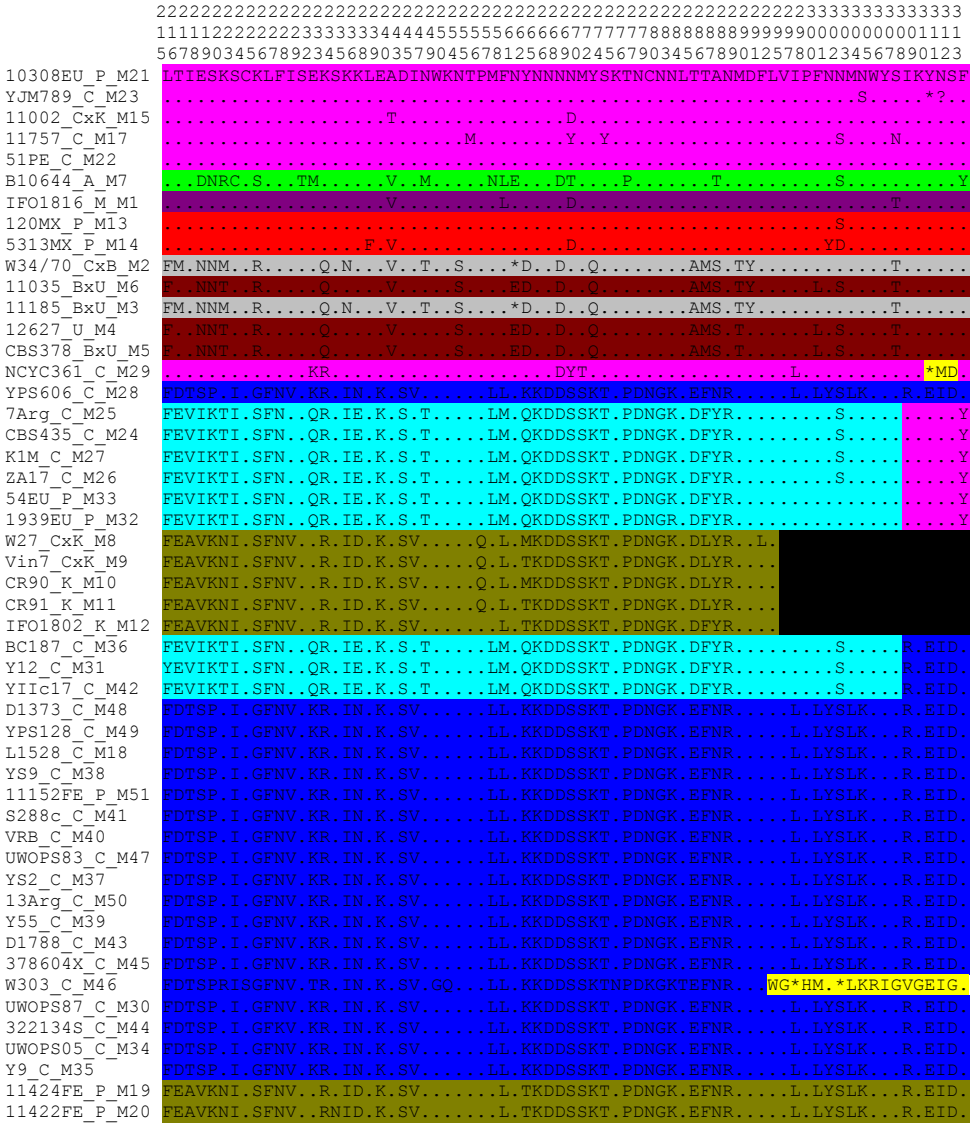
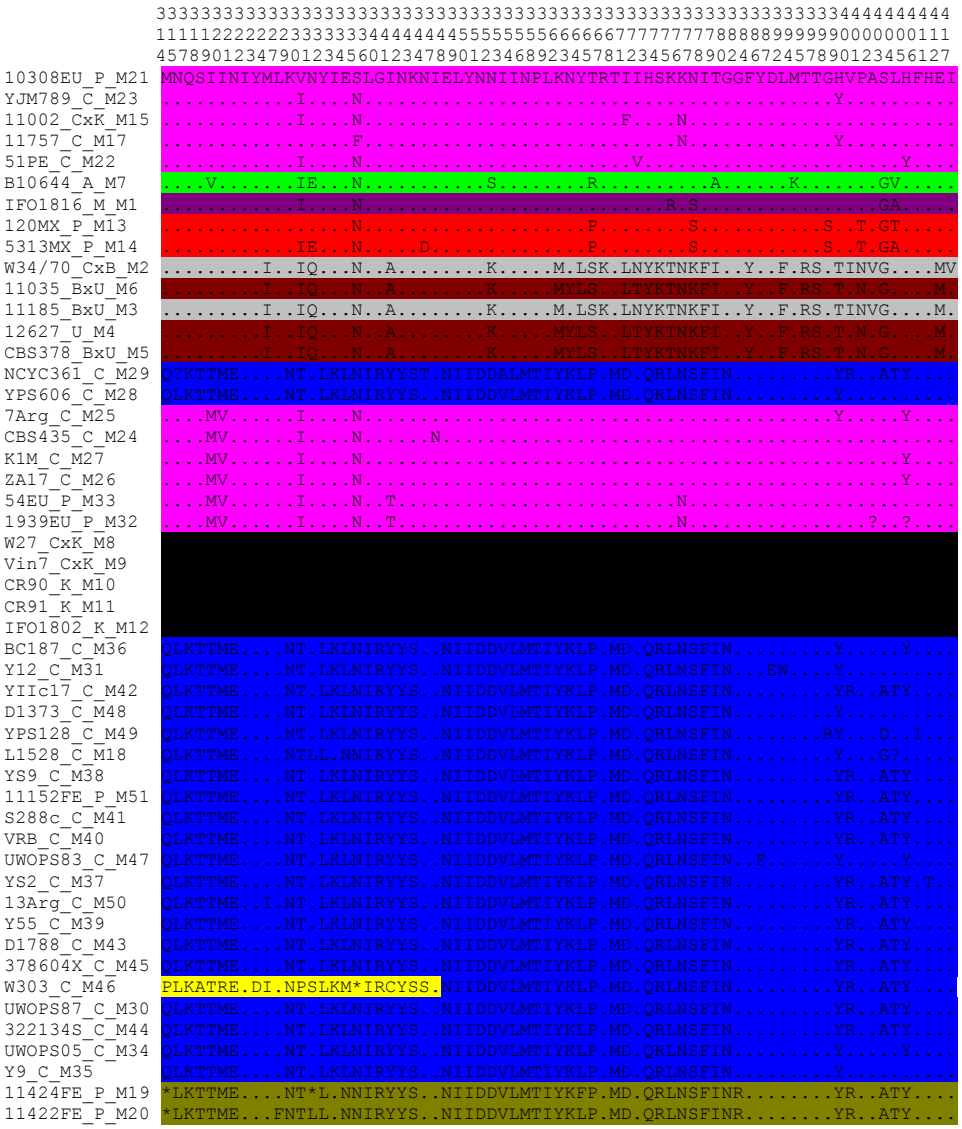


Table 1 by segments.


10308EU_P_M21
YJM789_C_M23
11002_CxK_M15
11757_C_M17
51PE_C_M22
B10644_A_M7
IFO1816_M_M1
120MX_P_M13
5313MX_P_M14
W34/70_CxB_M2
11035_BxU_M6
11185_BxU_M3
12627_U_M4
CBS378_BxU_M5
NCYC361_C_M29
YPS606_C_M28
7Arg_C_M25
CBS435_C_M24
K1M_C_M27
ZA17_C_M26
54EU_P_M33
1939EU_P_M32
W27_CxK_M8
Vin7_CxK_M9
CR90_K_M10
CR91_K_M11
IFO1802_K_M12
BC187_C_M36
Y12_C_M31
YIIc17_C_M42
D1373_C_M48
YPS128_C_M49
L1528_C_M18
YS9_C_M38
11152FE_P_M51
S288c_C_M41
VRB_C_M40
UWOPS83_C_M47
YS2_C_M37
13Arg_C_M50
Y55_C_M39
D1788_C_M43
378604X_C_M45
W303_C_M46
UWOPS87_C_M30
322134S_C_M44
UWOPS05_C_M34
Y9_C_M35
11424FE_P_M19
11422FE_P_M20


10308EU_P_M21
YJM789_C_M23
11002_CxK_M15
11757_C_M17
51PE_C_M22
B10644_A_M7
IFO1816_M_M1
120MX_P_M13
5313MX_P_M14
W34/70_CxB_M2
11035_BxU_M6
11185_BxU_M3
12627_U_M4
CBS378_BxU_M5
NCYC361_C_M29
YPS606_C_M28
7Arg_C_M25
CBS435_C_M24
K1M_C_M27
ZA17_C_M26
54EU_P_M33
1939EU_P_M32
W27_CxK_M8
Vin7_CxK_M9
CR90_K_M10
CR91_K_M11
IFO1802_K_M12
BC187_C_M36
Y12_C_M31
YIIc17_C_M42
D1373_C_M48
YPS128_C_M49
L1528_C_M18
YS9_C_M38
11152FE_P_M51
S288c_C_M41
VRB_C_M40
UWOPS83_C_M47
YS2_C_M37
13Arg_C_M50
Y55_C_M39
D1788_C_M43
378604X_C_M45
W303_C_M46
UWOPS87_C_M30
322134S_C_M44
UWOPS05_C_M34
Y9_C_M35
11424FE_P_M19
11422FE_P_M20


10308EU_P_M21 MNQSIINIYMLKVNYIESLGINKNIELYNNIINPLKNYTRTIIHSKKNITGGFYDLMTTGHVPASLHFHEI
YJM789_C_M23
11002_CxK_M15
11757_C_M17
51PE_C_M22
B10644_A_M7
IFO1816_M_M1
120MX_P_M13
5313MX_P_M14
W34/70_CxB_M2
11035_BxU_M6
11185_BxU_M3
12627_U_M4
CBS378_BxU_M5
NCYC361_C_M29
YPS606_C_M28
7Arg_C_M25
CBS435_C_M24
K1M_C_M27
ZA17_C_M26
54EU_P_M33
1939EU_P_M32
W27_CxK_M8
Vin7_CxK_M9
CR90_K_M10
CR91_K_M11
IFO1802_K_M12
BC187_C_M36
Y12_C_M31
YIIc17_C_M42
D1373_C_M48
YPS128_C_M49
L1528_C_M18
YS9_C_M38
11152FE_P_M51
S288c_C_M41
VRB_C_M40
UWOPS83_C_M47
YS2_C_M37
13Arg_C_M50
Y55_C_M39
D1788_C_M43
378604X_C_M45
W303_C_M46
UWOPS87_C_M30
322134S_C_M44
UWOPS05_C_M34
Y9_C_M35
11424FE_P_M19
11422FE_P_M20

and *S. cerevisiae* x *S. mikatae,* introgression of nuclear genes (Liti *et al.,* 2006; Muller and McCusker, 2009) and HGT of selfish elements (Liti *et al.,* 2005) have both been detected. Mitochondrial recombination occurs readily in yeast (Dujon *et al.,* 1974; Birky *et al.,* 1982; Taylor 1986; MacAlpine *et al.,* 1998), but until now, mtDNA recombination has only been described between different strains of *S. cerevisiae.* In our study we show that mtDNA recombination occurred between different species of *Saccharomyces* suggesting several hybridization events.

In a previous study we found a recombination hot spot in *COX2* gene, involving different species of *Saccharomyces* genus (Peris *et al., in preparation*). We hypothesized that recombination could be mediated by the *ORF1* (*RF1*) free-standing HEG. As previously described with ω (I-*Sce*I), the proteins codified in HEGs are selfish genetic elements that can infect other genomes without the HEG allele (HEG^-) (Colleaux *et al.,* 1986; Burt and Koufopanou 2004). A HEG protein mediates a DSB in the HEG^- allele, and the HEG^+ allele is used, by the recombinational repair system to repair the break, spreading the HEG into other genomes (Burt and Koufopanou 2004).

Surprisingly, we found two major groups of sequences for *ORF1* gene, each of which contained strains from different species (fig. 2). For example, *S. cerevisiae* and *S. paradoxus* strains could be found in both groups, Type I and Type II. Both groups also contained *ORF1*-encoded homing endonucleases that are predicted to be active (e.g. M51 in Type I and M1, M4-M6, M15, M16, M21, and M22 in Type II). We also previously described recombination points between *Saccharomyces* spp. in the *COX2* gene, which overlaps with *ORF1* (Peris *et al., in preparation*). Thus, as the *ORF1* homing endonuclease spread rapidly between *Saccharomyces* genomes by creating and repairing DSBs, it appears to have created multiple recombinant alleles of *COX2* and itself, including variants that cross species boundaries.

4.2 Recombination hot spots

We described several recombination points along the *ORF1* gene (Table 1). Most GC clusters and A+T tandem repeats were inserted near recombination points (fig. 1), consistent with previous observations that GC clusters are favored sites for mitochondrial recombination (Dieckmann and Gandy, 1987; Bouchier, 2009). Tandem repeat polymorphisms are commonly associated with GC insertions, indicating than GC clusters and their associated A+T repeats may have transposed as a unit (Skelly and Clark Walker, 1991). Based on our analyses, it seems likely that the internal recombination points found in the *ORF1* gene (fig. 1) are mediated by the presence of GC cluster and A+T rich sequences. For example, some haplotypes of *S. kudriavzevii* (M10-M12) and *S. cerevisiae* x *S. kudriavzevii* hybrids (M8 and M9) received segments from Far East *S. paradoxus* (CECT 11424 and CECT 11422) (Table 1, Figure S5B-C). These results indicate that interspecific recombination does occur in nature and suggest that the common flanking sequences may also mediate internal gene recombinations.

To determine the extent of the recombination hot spot, we sequenced the *COX3* gene. In *S. cerevisiae, S. bayanus*, and the ancestral *Saccharomyces* genome, *COX3* was near *COX2,* although they were separated by *ORF1* and the tRNA cluster (Groth *et al.,*, 2000). Although *S. paradoxus COX3* has been described in a different location (Groth *et al.,*, 2000), we still compared *COX3* gene sequences to its haplotypes correlated with the *COX2-ORF1* region. *S. mikatae, S. arboricolus*, and *S. kudriavzevii* mitochondrial genomes have been not described, so we assumed that *COX3* is near *COX2-ORF1-tRNAs*, as in the ancestral *Saccharomyces* mitochondrial genome. Another free standing homing endonuclease *ORF2* (*RF2*) that also contains GC clusters is near the *COX3* gene (Michel, 1984). Our results showed species-specific groups, with the exception of 120M and CBS 5313, indicating that recombination hot spots are probably located in the *COX2-ORF1* region and not *COX3*. If our model

about the *ORF1* HEG-mediated recombination is correct, it may suggest *ORF2* is inactive in most strains (fig. 3) or absent, as has been recently published in *S. paradoxus* CBS432 (Procházka *et al.,* 2012).

4.3 ORF1 *gene inactivation*

HEGs follow a life cycle (Goddard and Burt 1999). An empty site is infected by a HEG^+ allele, the functional HEG degenerates and loses functionality, and finally suffers a precise lost, starting again the cycle. In the case of *ORF1* gene, we found two different types, both of which included active variants. Moreover, all the strains sequenced in this study or in other projects (as SGRP) have an *ORF1* gene sequence. These findings indicate that the *ORF1* gene is widespread and often functional in the *Saccharomyces* genus.

Two different I-SceI types have been described: an inactive one containing a GC cluster that breaks the reading frame in *S. cerevisiae* (de Zamaroczy and Bernardi, 1986) and an active one without the GC cluster in *S. bayanus* var. *uvarum* (Séraphin, 1987). For *ORF1*, we described up to 7 GC clusters inserting points, most of which were found in Type II. Importantly, our haplotype classification procedure did not take GC clusters and indels into account. Some strains grouped into the same haplotypes showed differences in GC clusters, which suggest that GC cluster insertion occurred after *ORF1* infection. In the case of *S. paradoxus* 120M and CBS 5313, which have maintained a mtDNA from *S. cerevisiae* (see below) but an *ORF1* type different to those observed in most *S. cerevisiae* strains, they showed a GC cluster similar to *S. cerevisiae* strains. This suggests that GC cluster jumped from a different mtDNA region (*S. cerevisiae* mtDNA region) and inserted in the *ORF1* gene (Figure S2).

Most strains from the M40 haplotype have two GC clusters (6 and 7), although VRB strain had two additional GC clusters (1 and 4), and L351 and

AMH had only GC cluster 6. Similarly, SK1 of the M41 haplotype had only GC cluster 6. The introduction of GC clusters is apparently independent but region specific, because some haplotypes have GC cluster 6 but not 7, while some have GC cluster 7 but not 6.

Sequences with and without GC clusters maintain the flanking regions necessary for GC insertion (Figure S1), as previously described for GC cluster 6 and 7 (Weiller *et al.,*, 1989; Séraphin *et al.,* 1987). After infection by an *ORF1* homing endonuclease, GC clusters could be introduced into the *ORF1* gene as a defense mechanism. According to our results, strains with similar mtDNA (*COX2* and *COX3* sequences) but with different *ORF1* genes have similar GC clusters. This suggests that GC cluster origin is from mtDNA. GC cluster insertion is not a unique mechanism to inactivate *ORF1*. Accumulation of mutations can produce premature stop codons as occurred in *S. cerevisiae* haplotypes M26 and M23; *S. kudriavzevii* haplotypes M8, M10 and M11; and *S. bayanus* var. *bayanus x S. bayanus* var. *uvarum* M2 and M3.

Analysis of dN/dS showed that most of the amino acid residues are evolving neutrally, suggesting that *ORF1* is not performing an important function in these strains, consistent with their role as selfish genetic elements (Doolittle and Sapienza, 1980). The conservation of some amino acids in LAGLIDADGs and NUMOD domains are probably because rapid re-infection continually restores the original amino acid sequences of *ORF1* domains.

4.4 Evolution of ORF1 *homing endonuclease*

The evolution of the *ORF1* gene is quite complex. The most plausible scenario is that type I *ORF1* homing endonuclease is the most ancestral sequence inherited by the *Saccharomyces* common ancestor. Most of the *Saccharomyces* strains are circumscribed in the Type I group, including *S. cerevisiae* strains (M17, M22 and M23), *S. paradoxus* from America (M13 and M14) and Europe (M21), *S. arboricolus*, *S. mikatae,* S. *cerevisiae* x *S.*

kudriavzevii (CECT 1990, CECT 1388 and CECT 11002), *S. bayanus* var. *bayanus* and *S. bayanus* var. *uvarum*. Type I is probably more infective than Type II and may serve as a reservoir of functional alleles, since we found that most putatively active *ORF1* genes are Type I. Some *S. cerevisiae* and *S. paradoxus* from FE (CECT 11152) have lost the Type I *ORF1* gene, but they were re-infected with a new version of *ORF1* that we called Type II. The origin of Type II is unknown, but probably it was received from a member of the *Saccharomycetaceae* family. However, the high similarity of *ORF1* Type II with *ORF1* Type I and their phylogenetic placement (Figure S4) suggest that the donor was not a distant relative. Alternatively, *ORF1* Type II could have suffered rapid evolution and degeneration after inactivation of a Type I-like ancestor.

Homing endonucleases usually are found inside introns or inteins. They facilitate the splicing of intron and inteins, receiving the name of maturases (Belfort 2003). In the case of *ORF1,* it is not found inside an intron or intein, being a free-standing homing endonuclease. It has been described in *Schizosaccharomyces pombe* the presence of intron and intron-less strains in *COX2,* containing a homing endonuclease encode inside the intron (Schäfer *et al.,* 1998). Probably, in the evolution of *Saccharomyces* genus the intron of *COX2* was lost but the infective character of homing endonucleases has allowed it to be maintained in the nearest of *COX2* gene. However, it surprising that the homing endonuclease of *S. pombe* in *COX2,* an H-N-H family (Schäfer *et al.,* 1998), is from a different family than *S. cerevisiae COX2*, LAGLIDAD family. It suggests that the evolution of these homing endonucleases is independent and conclusions are more complex to obtain. In our opinion, the self-splicing nature of group I and group II mitochondrial introns (Belfort 2003) indicate that maturases are not needed. However, the efficiency of splicing is increased, and in other cases could be necessary

(Belfort 2003). This could explain the presence of different homing endonucleases in the *COX2* of *S. pombe* and *Saccharomyces* genus.

4.5 Footprint of ancestral hybridization

Though intraspecific recombination is well-documented in *S. cerevisiae*, our results suggest that the mitochondrial genome can retain the footprint of prior interspecific infections by homing endonucleases. When mating occurs between two different strains or even species, fusion of mitochondrial organelles is followed by the formation of a continuous reticulum-denominated heteroplasmic state (Berger and Yaffe, 2000). During this process, the Orf1 homing endonuclease spreads from $ORF1^+$ mtDNA to $ORF1^-$ mtDNA by targeted gene conversion. GC insertions and A+T repeated sequences also seem to facilitate recombination between different $ORF1^+$ alleles. Inheritance of mtDNA is uniparental and a homoplasmic state is quickly reached (Basse, 2010). Thus, after recombination, daughter cells could receive a unique recombinant type of mtDNA. As we showed here (Table 1 and Figure S5) and previously (Peris *et al., in preparation*), interchanges of nucleotides sequences have occurred several times between different *Saccharomyces* species (Figure S5B).

We also observed a complex mtDNA in 120M and CBS 5313, two strains with 5 nuclear gene sequences from an American *S. paradoxus* (Arias 2008). 120M and CBS 5313 displayed *COX2* sequences similar to *S. cerevisiae* strains (Peris *et al., in preparation*), and here we show that *COX3* gene also matches *S. cerevisiae.* Thus, it seems that an ancestral unstable hybridization occurred between an American strain of *S. paradoxus* and a strain of *S. cerevisiae* with an *ORF1* type I homing endonuclease closer to *S. paradoxus* from Europe, suggesting that the parental donor could be a strain similar to 51PE. In the case of *S. cerevisiae* x *S. kudriavzevii* hybrids containing a mtDNA type K5, K6 and K10 in Peris *et al.* (2012a) or haplotypes

H87-H89 in Peris *et al.* (*in preparation*) two scenarios are possible. In one of them, a European *S. paradoxus* could transfer its *ORF1* to the mtDNA of a European *S. kudriavzevii* strain and inherited by these hybrids after mating with a brewery or ale, depending of the parental strain *S. cerevisiae*, as we shown in the evolutionary reconstruction of these hybrids (Peris *et al.,* accepted). In the contrary, *ORF1* from a European *S. paradoxus* could be transferred to a wine *S. cerevisiae* as we have shown the high similarity between CECT 11757 and European *S. paradoxus ORF1* sequences, and a derivative strain could hybridize with a European *S. kudriavzevii* (Peris *et al.,* accepted) and a recombinant mtDNA *S. kud-S.cer* could be maintained in these hybrids.

In conclusion, as more mitochondrial genomes will be sequenced, we could more thoroughly trace the complex evolution of *Saccharomyces* genus. Apparently, the evolution of this genus is not strictly linear. The recently demonstration that some species can live in sympatric association (Sampaio and Gonçalvez, 2008), species of this genus could hybridize with a high frequency yielding unstable hybrids, maintaining the portion of one parental genome, or new cells with a genome from predominantly one parent in most of the cases (Dunn and Sherlock 2008; Peris *et al.,* 2012b). Nonetheless, in the former case some footprints are occasionally left in the nuclear genome (through introgression and HGT) or the mitochondrial genome, as we have shown in this work, occurring with the *ORF1* allele. This indicates that hybridization is a random and frequent mechanism, which in some conditions, as biotechnological environment, the hybrid genome could be maintain or stabilized.

Aknowledgements

We thank to C.P. Kurtzman to provide us the *S. mikatae* NRRL Y-27342. We thank Chris Todd Hittinger for paper revision and comments which

have improved the draft. DP and EB conceived the study. DP performed the amplification and sequence analysis. DP drafted the manuscript. SO and CB participated in the manuscript revision. This work was supported by Spanish Government project AGL2009-12673-CO2-02 to EB. DP acknowledges to the Spanish Government (MICINN) his FPI fellowship.

SUMMARY OF RESULTS & DISCUSSION

In this doctoral thesis we have elucidated important questions regarding *S. cerevisiae* x *S. kudriavzevii* hybrid yeasts. We have identified and characterized new hybrids from different environments. We have researched the genome structure of the hybrids unveiling the origin of the parent strains and their role in the hybrid genomes. We have estimated the minimum number of hybridization events, the geographical origin of the hybrids and we detected ancestral hybridization between different *Saccharomyces* species supporting a high frequency of this phenomenon. Finally, we have explored the biotechnological advantages of the hybridization process underwent by *S. cerevisiae* and other *Saccharomyces* species as well as the evolutionary consequences of species hybridization for the genus *Saccharomyces*.

1. Identification of natural hybrids

PCR-restriction fragment length polymorphisms of 35 nuclear genes distributed around the 16 chromosomes of *Saccharomyces* were used for identification of new natural hybrids *S. cerevisiae* x *S. kudriavzevii*. Four new natural hybrids, PB7 and SOY3 isolated from wines in the Southern of Europe and, MR25 and IF6 isolated from clinical and food environment, respectively, have been described. The double parental origin of these natural hybrids *S. cerevisiae* x *S. kudriavzevii* was confirmed by gene sequencing of seven nuclear genes (*BRE*5, *CAT8*, *EGT2*, *GAL4*, *MET6*, *CYR1* and *CYC3*).

RFLPs data have revealed high diversity in the genomes of the double hybrids *S. cerevisiae* x *S. kudriavzevii* and triple hybrids *S. bayanus* var. *uvarum* x *S. cerevisiae* x *S. kudriavzevii*. Different degrees of *S. kudriavzevii* gene loss could be deduced from RFLP data, ranging from hybrids with a copy of each parent RFLP allele to hybrids with a complete set of *S. cerevisiae* RFLP alleles but lower content of *S. kudriavzevii* RLFP

alleles, such as AMH and IF6. A similar trend of *S. kudriavzevii* gene loss was found in brewing and a group of Swiss wine *S. cerevisiae* x *S. kudriavzevii* hybrids, using the same methodology (González *et al.,* 2008). The high genome diversity found among the *S. cerevisiae* x *S. kudriavzevii* hybrids could be indicative of several hybridization events, as it is explained in the next sections.

The discovery of the new natural *S. cerevisiae* x *S. kudriavzevii* hybrids has expanded the present limits of their geographical distribution, from the Continental Europe where previous *S. cerevisiae* x *S. kudriavzevii* hybrids had been isolated, to the southern Mediterranean countries of the European wine regions (Schütz & Gafner 1994; González *et al.,* 2006; Lopandic *et al.,* 2007). These new hybrids, from Southern Europe wine areas, have appeared at low frequencies, coexisting with the dominant *S. cerevisiae* strains during the first stages of the wine fermentation. Perhaps the milder temperatures of these regions, at which spontaneous fermentations occur, still allow the coexistence of *S. cerevisiae* and hybrids.

2. Genome structure of S. cerevisiae x S. kudriavzevii *hybrids*

The genome structure of 14 *S. cerevisiae* x *S. kudriavzevii* hybrids has been elucidated using a combination of RFLPs, array Comparative Genome Hybridization (aCGH) and flow cytometry. Regarding DNA content, the interpretation of caryoscope diagrams obtained from aCGH analyses combined with RFLPs and flow cytometry indicated that most hybrids are triploids, except of AMH and PB7 that are tetraploids. Hybrids displayed a complex genome structure containing aneuploidies and several rearrangements due to recombinations between homeologous chromosomes generating chimerical chromosomes. Different types of chromosome rearrangements were detected: i) the complete loss of a *S. kudriavzevii* parent chromosome compensated by an extra copy of the *S.*

cerevisiae chromosome; ii) aneuploidies, involving more or less chromosome copies; iii) and the presence of chimerical chromosomes. Hybrids displayed frequent chromosome rearrangements, as genome sequencing projects of wine hybrid VIN7 and lager brewing hybrid W34/70 (*S. cerevisiae* x *S. eubayanus*) have confirmed (Nakao *et al.,* 2009; Borneman *et al.* 2011). In yeast, any kind of aneuploidies decreases cellular fitness (Torres *et al.,* 2008), although aneuploidies generate significant phenotypic variation conferring fitness advantages in not common conditions, such as stress conditions (Pavelka *et al.,* 2010). The frequency of aneuploidies or chromosome rearrangements in hybrid strains could be due to their adaptation to winemaking and beer production, two processes characterized by a succession of physicochemical stress conditions (Ivorra *et al.,* 1999; Carrasco *et al.,* 2001; Briggs *et al.,* 2004). Future studies about engineered hybrids with different composition of aneuploid chromosomes could reveal which of them confers fitness gains in different stress conditions.

Previous studies had shown similar complex genome structure of hybrid W27 *S. cerevisiae* x *S. kudriavzevii* isolated from wine in Switzerland (Belloch *et al.,* 2009). Using propidium iodide, to quantify the DNA content, these authors described W27 and other Swiss *S. cerevisiae* x *S. kudriavzevii* hybrids as diploids (Belloch *et al.,* 2009). In this thesis, quantification of DNA content using SYTOX Green revealed that W27 and other Swiss hybrids are triploids. Recent studies, using a similar approach, determined hybrids ploidy between 2n and 4n, being most cases 3n (Erny *et al.,* 2012). Genome sequencing of VIN7 (Borneman *et al.,* 2011) and W27 (data not published), is in accordance to 3n DNA content.

S. cerevisiae x *S. kudriavzevii* hybrids show a complete copy of *S. cerevisiae* chromosomes and a complete or partial copy of *S. kudriavzevii* chromosomes. The reduction of the non *S. cerevisiae* genome in

Saccharomyces hybrids was already reported in artificial *S. cerevisiae* x *S. uvarum* hybrids (Antunovics *et al.,* 2005). In contrast, hybrids *S. cerevisiae* x *S. eubayanus* group 1 maintain the genome of *S. eubayanus* and can lose *S. cerevisiae* chromosomes (Dunn and Sherlock 2008). These differences could be due to the different mtDNA inheritance. In the case of artificial *S. cerevisiae* x *S. uvarum* hybrids, all hybrids inherited a *S. cerevisiae* mtDNA genome (Antunovics *et al.,* 2005); in contrast, *S. cerevisiae* x *S. eubayanus* group 1 hybrids inherited the ancestor *S. bayanus* mtDNA (Dunn and Sherlock 2008). Is important to note the different behavior in *S. cerevisiae* x *S. kudriavzevii* hybrids where most of them inherited a *S. kudriavzevii* mtDNA and they can lose the *S. kudriavzevii* nuclear genes.

Despite the loss of *S. kudriavzevii* genes all hybrids shared a common pool of *S. kudriavzevii* genes. GO analysis showed a significant overrepresentation of *S. kudriavzevii* genes associated with the physiological adaptation of hybrids to growth at low temperatures, such as fatty acid transport and N-glycosilation of proteins, important to maintain the integrity of membranes. Furthermore, GO terms related with stress tolerance such as ergosterol, phospholipids and aminoacid metabolism, as well as temperature inducible protein (*TIP1*) and seripauperins (*PAU*) were conserved in most of hybrids. Previous studies showed a better adaptation of *S. kudriavzevii* at lower temperatures (Belloch *et al.,* 2008) attributed to having a different lipid composition of membrane compared to *S. cerevisiae* (Tronchoni *et al.,* 2012). However, *S. kudriavzevii* shows low tolerance to ethanol compared with *S. cerevisiae* (Belloch *et al.,* 2008; Arroyo-López *et al.,* 2009).

The gene reduction observed in the *S. kudriavzevii* moiety and maintenance of genes from the *S. cerevisiae* moiety might be a direct effect of selective pressure under fermentative or propagation conditions where

the *S. cerevisiae* parent has advantages. For this reason, the most recent hybrids could be those showing a complete copy of each parent strain, such as PB7, whereas the most ancient hybrids would be those with genomes showing less percentage of *S. kudriavzevii* genome and more rearrangements. This is supported by the comparison of the genome structure of VIN7 analyzed in this thesis with the genome sequence of VIN7 from Borneman *et al.,* (2011) study. The latter is the original strain and the former is a derived commercial strain which presents the absence of *S. kudriavzevii* chromosome III compared with the original strain probably due to the process of propagation under stress conditions. Nevertheless, the *S. kudriavzevii* genes would be maintained as a counterpart of the *S. cerevisiae* genome distinguishing these hybrids with higher fermentative power at lower temperature of fermentation.

3. mtDNA of S. cerevisiae *x* S. kudriavzevii *hybrids*

COX2 gene sequencing was used to infer the mtDNA inheritance in *S. cerevisiae* x *S. kudriavzevii* hybrids. Most of the natural hybrids inherited the *S. kudriavzevii* mitochondrial DNA, except hybrids AMH and IF6 which have inherited the mtDNA from the *S. cerevisiae* parent. Mitochondrial inheritance is uniparental, therefore hybrids can only inherit the mtDNA from *S. cerevisiae*, *S. kudriavzevii* or occasionally a recombinant mtDNA (Basse 2010).

Harboring any of the different mtDNAs might impose limitations in conserving specific genes in the nuclear genome. The presence of the *S. kudriavzevii* mtDNA in the hybrid forces the retention of most *S. kudriavzevii* nuclear genes involved in mitochondrial functions. GO analysis of nuclear *S. kudriavzevii* genes shows a significant enrichment in genes related to mitochondrial functions. Strains retaining the *S. cerevisiae* mtDNA, such as AMH and IF6, have suffered the highest *S. kudriavzevii*

gene reduction. Recent studies have shown the presence of nucleo-mitochondrial incompatibilities between *S. bayanus* nuclear gene *AEP2* and *S. cerevisiae* mitochondrial gene *OLI1* (Lee *et al.,* 2008). Moreover, genes *MRS1* and *AIM22* are associated with cytonuclear incompatibilities among *S. cerevisiae, S. paradoxus* and *S. bayanus* hybrids (Chou *et al.,* 2010). In *S. cerevisiae* x *S. kudriavzevii* hybrids, the retention of a high number of nuclear genes of *S. kudriavzevii* carrying the *S. kudriavzevii* mtDNA compared with hybrids containing the *S. cerevisiae* mtDNA could be attributed to cytonuclear incompatibilities.

It's noteworthy that mtDNA inheritance has an important role in nuclear gene expression (Parikh *et al.,* 1987), influencing in the metabolic fluxes that could increase or decrease the respiratory or fermentative pathways. Artificial hybrids *S. cerevisiae* x *S. uvarum* showed different carbohydrate metabolism and hexose transport depending on the parental mtDNA (Solieri *et al.,* 2008). Moreover, the lipid composition of the membrane is essential to yeast adaptation at different environmental temperatures (Henschke and Rose, 1991; Beltran *et al.,* 2008; Redón *et al.,* 2011; Torija *et al.,* 2003). Yeast adaptation to the beginning of fermentation (in wine or beer) requires an initial step of respiration for lipid synthesis (Hammond, 2000; Briggs *et al.,* 2004). Recently, it has been shown the different lipid composition of *S. cerevisiae*, *S. kudriavzevii* and their hybrids (Tronchoni *et al.,* 2012). Here, we postulate that these differences could be done by new interactions between mtDNA and nuclear genome in the hybrids, changing the flux of respiration and fermentation, obtaining wines and beers with new organoleptic properties, where glycerol, aromatic compounds and ethanol concentrations are modified compared to the *S. cerevisiae* strains. This is supported by the different gene expression between hybrids and their parents (Tronchoni *et al., in preparation*) conducting to a different aroma

profiles (Gangl *et al.,* 2009; Gamero *et al.,* 2011) and glycerol and ethanol production (Arroyo-Lopez *et al.,*, 2010).

A group of hybrids (most brewing hybrids, PB7, CID1 and MR25) enclosed in type (haplotype) K5 (H89), K6 (H88) and K10 (H87), have inherited a recombinant version of mtDNA. However, recombination only involves the 3'-end of *COX2* gene and *ORF1*, as we will describe in more detail in the *Mitochondrial recombination* section. The hybrids above, harbor a mtDNA recombinant sequence which one of the segments of the gene is similar in sequence to a European *S. paradoxus* strain, although hybrids between *S. kudriavzevii* x *S. paradoxus* have not been found. However, this finding opens the possibility that a *S. kudriavzevii* strain, not found yet, could have hybridized with a European *S. paradoxus*. The *S. kudriavzevii* strain could have inherited a recombinant version of mtDNA *S. kudriavzevii-S. paradoxus* and later on, it could have been involved in the origin of the *S. cerevisiae* x *S. kudriavzevii* hybrids. Nevertheless, some *S. cerevisiae* strains have a recombinant mtDNA *S. cerevisiae-S. paradoxus*, in the same manner as the hybrids, although in those cases the *COX2* gene sequences correspond to *S. cerevisiae* and the 3'-end and *ORF1* to a European *S. paradoxus.* In this case, the parent *S. cerevisiae* would be the donor of the *S. paradoxus* sequences to the *S. cerevisiae* x S. *kudriavzevii* hybrids.

4. On the origins of the hybrids

The results obtained from RFLPs, aCGH, flow cytometry and nuclear and mitochondrial gene sequencing were used for the construction of phylogenetic trees (Neighbor-Joining and Maximum Parsimony methods) and supernetworks to infer the minimum number of hybridization events necessary for explaining the genome diversity found in the hybrids, as well as the number of parental strains involved in their generation. In addition,

the comparison of nuclear and mitochondrial sequences from *S. cerevisiae* and *S. kudriavzevii* strains from different geographical origins was used to deduce the geographical origin of the hybrids.

Differences between wine and non-wine related *S. cerevisiae* strains have been previously described. Arias (2008) demonstrated the existence of genetic polymorphisms shared by most wine strains forming a “wine allele haplogroup”. Moreover, genome sequencing of monosporic cultures of *S. cerevisiae*, isolated from different origins and sources, has demonstrated different groups of "pure" *S. cerevisiae* strains (Liti *et al.,* 2009). One of them corresponds to the wine/European *S. cerevisiae* strains. As showed in this thesis "wine alleles" from Arias (2008) are clustered with wine/European strains from Liti *et al.,* (2009). Similarly, Carreto *et al.,* (2008), using microarray analysis, established that *S. cerevisiae* strains isolated from wine fermentations shared a common pool of depleted genes. This common pool of depleted genes was also observed by microarray analysis in the hybrids. The *S. cerevisiae* alleles from seven nuclear genes of *S. cerevisiae* x *S. kudriavzevii* hybrids have been shown to cluster in the "wine haplogroup", where wine alleles from Arias (2008) and wine/European from Liti *et al.,* (2009) were also found. These findings support the wine origin of the putative *S. cerevisiae* parent strain of most *S. cerevisiae* x *S. kudriavzevii* natural hybrids. In the case of CECT 11011 (brewing hybrid strain), a non-wine allele was found, indicating that probably an "ale" strain could have originated the brewing group of hybrids that clustered with CECT 11011. In addition, CECT 1388, 1990, 11011 and MR25 have heterozygous alleles which is another feature of brewing *S. cerevisiae* strains (Arias 2008). CHIMAY brewing hybrid, using microsatellite data, was clustered with brewing *S. cerevisiae* strains (Erny *et al.,* 2012), supporting our hypothesis. A genome sequencing project of the brewing hybrids could elucidate the "ale" origin of this group.

A similar approach was used in the case of *S. kudriavzevii* alleles. It has been shown that *S. kudriavzevii* strains from Japan and from Europe are different at metabolic level, such as GAL pathway (Hittinger *et al.,* 2010), as at genetic level (Lopes *et al.,* 2010). The *S. kudriavzevii* alleles from the seven nuclear gene sequences and RFLP allele patterns of *S. cerevisiae* x *S. kudriavzevii* hybrids, described in this thesis, were also shared by some European *S. kudriavzevii* strains.

Our results indicate that at least six different hybridization events would have been necessary to generate the double hybrids *S. cerevisiae* x *S. kudriavzevii* and two additional origins for the triple hybrids *S. cerevisiae* x *S. kudriavzevii* x *S. uvarum,* studied in this thesis. Six wine/European *S. cerevisiae* and six European *S. kudriavzevii* strains could have been involved in the generation of the hybrids. *COX2* gene sequences have displayed two clear haplogroups where wine/European *S. cerevisiae* strains could be found, supporting at least two clear groups of parental *S. cerevisiae* strains, as we found in the supernetworks. Some parental *S. cerevisiae* strains could be mosaics from the two meaning groups. Previous studies had postulated two different origins for the *S. cerevisiae* x *S. kudriavzevii* hybrids (Gonzalez *et al.,* 2008); however the results generated in this thesis have increased this number. Although more hybridization events can be postulated than in Erny *et al.,* (2012) study, where three hybridization events were proposed, our results are in agreement with that study. We propose an independent origin for the Swiss and two brewing (CECT 11003 and CECT 11004) hybrids, what is in accordance with the microsatellite phylogenetic analysis of hybrids, performed by Erny *et al.,* (2012). A second origin, proposed for the Alsatian VIN7, Austrian and Croatian SOY3 hybrids, is also in accordance to Erny *et al.,* (2012) work, as these hybrids are close relatives of the Hungarian wine hybrids. Therefore, this is a lineage of wine hybrids widely distributed along the Rhine valley

(Alsace and Germany) to the Danube valley (Pannonian region: Austria, Croatia and Hungary).

A different origin, for most of the brewing hybrid strains and MR25 (a clinical isolate), is also proposed in this thesis. Microsatellite analysis has demonstrated that CHIMAY hybrid strain is close to *S. cerevisiae* from beer (Erny *et al.,* 2012). In addition, the presence of a non-wine allele in CECT 11011 and CECT 1388, and two alleles for *BRE5* in some brewing and MR25 are indicative of a brewing origin of the parent *S. cerevisiae*, supporting the independence of this group.

In this thesis independent origins for PB7, AMH and IF6 hybrids are also proposed. And finally, the origin of the triple hybrids CID1 and CBS2834 is not clear due to the additional occurrence of a secondary hybridization event either between the *S. cerevisiae* x *S. kudriavzevii* hybrid or derivative and a *S. uvarum* strain or between a *S. cerevisiae* x *S. uvarum* hybrid or derivative and a *S. kudriavzevii* strain. The *S. cerevisiae* CBS 2834 parent strain seem to be that originating the Austrian-Croatian hybrids, but not the *S. kudriavzevii* parent; although similar mtDNA sequence, inherited from *S. kudriavzevii* parent, was observed.

The DNA content (ploidy) observed in the *S. cerevisiae* x *S. kudriavzevii* hybrids indicates different types of crosses or mating types. For most *S. cerevisiae* x *S. kudriavzevii* hybrids, a cross between a diploid *S. cerevisiae* and a haploid *S. kudriavzevii* might have occurred. Moreover, the genome composition of the *S. cerevisiae* x *S. kudriavzevii* natural hybrids indicates a high ploidy of the ancestral hybrid, around triploid or tetraploid, that is compatible with a rare-mating mechanism between a diploid *S. cerevisiae* and a haploid/diploid *S. kudriavzevii* (Pretorius 2000; de Barros Lopes *et al.,* 2002). Hybrid PB7 seems to be originated by a cross between two diploid strains. The genome structure of hybrid AMH indicates that two hybridization events might have occurred. In the first one,

a diploid *S. cerevisiae* would have mated with a haploid *S. kudriavzevii*, and in the second a cross between the hybrid or derivative and a diploid *S. cerevisiae* would have occurred. Hybrid generation by two crosses or mating events have been already described to elucidate natural triple hybrids CID1 and CBS 2834 (Groth 1999; González *et al.,* 2006). Although Erny *et al.,* (2012) proposes spore matings and duplication of the complete set of *S. cerevisiae* chromosomes as the meaning mechanism of hybrid generation, heterozygotes brewing hybrids and the formation of artificial hybrids generated by rare-mating in laboratory, which appears to be very easy (Pérez-Través *et al.* .2012), supports our hypothesis.

The results generated in this thesis do not allow to infer the primal country where the first *S. cerevisiae* x *S. kudriavzevii* hybrids were originated; however, gene similarity between the *S. kudriavzevii* subgenome of the hybrids and the European *S. kudriavzevii*, and between the *S. cerevisiae* subgenome of the hybrids and *S. cerevisiae* strains isolated in Europe indicated that Europe could be considered the primary continental origin for the natural *S. cerevisiae* x *S. kudriavzevii* hybrids. Although, our Swiss group of hybrids were not close relative to any of described European *S. kudriavzevii* populations, the recent efforts focus in sampling *S. kudriavzevii* strains around Europe are discovering new *S. kudriavzevii* strains. We suspect that the new group of *S. kudriavzevii* from Ardèche (France) (Erny *et al.,* 2012), quite different to those from Portugal and Spain probably could be related to the parent *S. kudriavzevii* from where Swiss hybrids were originated.

From a historical perspective, the expansion of the *S. cerevisiae* wine strains can be seen as a consequence of the spread of vineyards around the world (Chambers & Pretorius 2010; Sicard & Legras 2011). These "domesticated" *S. cerevisiae* strains (Fay and Benavides *et al.,* 2005) have problems when performing wine fermentations at the lower temperatures to

which other *Saccharomyces* species are better adapted (Salvadó *et al.,* 2011). Under such circumstances, hybrids have clear advantages over the parent species (Serra *et al.,* 2005; Belloch *et al.,* 2008; Arroyo-López *et al.,* 2009). Acquiring good alcohol and glucose tolerances and fast fermentation performances from *S. cerevisiae* (Belloch *et al.,* 2008; Arroyo-López *et al.,* 2010) and a better adaptation to low and intermediate temperatures from *S. kudriavzevii* (González *et al.,* 2007; Belloch *et al.,* 2008; Arroyo-López *et al.,* 2009). In the regions where principal successful lineages of *S. cerevisiae* x *S. kudriavzevii* wine hybrids have been found (Rhine and Danube valleys: Pannonian basin of Central Europe, Alsace and other oceanic and continental climate regions) (Erny *et al.,* 2012 and this thesis), winemaking had been introduced or improved on by Cistercian monks. In all these regions, winemaking and beer production were located in abbeys established by Cistercians monks or new orders derivative from Cistercians, such as Trappist monks. Cistercians monks established abbeys around Europe between 11th and 13th centuries. Around 300 Cistercian abbeys where established in the Rhine and Danube valleys and Trappist abbeys were located in Belgium (Burton and Kerr, 2011; Courtray, 1986). Hybrids have been isolated from these regions, indicative that hybrids and Cistercians/Trappist monks could be related.

Another interesting point to consider is the environment in which the hybrids might have surfaced. *S. kudriavzevii* has been not found in biotechnological environments; in contrast it is found in oaks, in sympatric association with *S. cerevisiae* and *S. paradoxus* (Sampaio & Gonçalves 2008). Therefore, the origin of the hybrids might be linked with the arboreal environment. Experiments done to favor hybrid generation between *S. cerevisiae* and *S. kudriavzevii* under fermentative conditions demonstrated that the biotechnological environment is not adequate for hybrid generation (Arroyo-López *et al.,* 2011). The fermentations underwent in the past were

not sophisticated as nowadays and many insects, principal vectors of yeasts (Reuter *et al.,* 2007), as wasps from Italian regions, recently demonstrated (Stefanini *et al.,* 2012), could have been in contact with the first steps of the fermentations process. If *S. cerevisiae* x *S. kudriavzevii* hybrids were randomly originated and maintained as a consequence of the better adaptation to lower temperatures, of continental and oceanic climates in the wild, and put in contact with the substrates of wine and beer, they might be selected by an unconscious manner by humans and maintained until today due to their good fermentative properties. After on, their genomes could evolve as a consequence of the new fermentative environment conditions. However, it must be confirmed the existence of hybrids in nature or insects from Central Europe.

5. Nuclear recombination hotspots

Caryoscopes diagrams from aCGH and phylogenetic trees were used to define several recombination hotspots in the genome of the hybrids.

Genome structure of the hybrids revealed the presence of several chimerical chromosomes which structure was shared by hybrids from putatively different origin. A detailed study of the regions where recombination had been occurred revealed the presence of highly recombining sequences, such as ARS, Ty elements, Y' elements, rRNA regions and conserved coding genes. Recombination between homeologous chromosomes could then have been mediated by highly recombining sequences, as described in previous studies (Kim *et al.,* 1998; Pérez-Ortín *et al.,* 2002; Di Rienzi *et al.,* 2009) . The mismatch repair system (MMR) that favors the loss of one partner when the recombination is initiated between homeologous chromosomes would have mediated the complete or partial chromosome losses of the *S. kudriavzevii* subgenome in the hybrids. In addition, this process increases the LOH of many parts of

the hybrid genomes, as we observed in the region ITS5.8-rDNA of IF6 hybrid.

Presence of chimerical chromosomes has also been related with low viability of spores. Most of *S. cerevisiae* x *S. kudriavzevii* hybrids have shown low spores viability (<1%), with the exception of hybrid PB7 which showed close to 100% spores viability. Hybrid PB7 is an exception within the hybrids as it shows two copies of each parent chromosome for most of the chromosomes and very few chromosomal rearrangements. The low spore viability observed in hybrids could be due to as a consequence of post-zygotic barriers, created by nucleo-cytoplasm incompatibilities, chimerical chromosomes or other unknown mechanisms. Consequences of low spore viability are the lack of successful sexual reproduction and prevalence of reproduction by mitosis. Although recombination have been associated with meiosis, mitotic recombination has also been described occurring at low frequency (Andersen & Sekelsky 2010) being the most plausible mechanism occurring in hybrids.

6. Mitochondrial recombination as a footprint of ancestral hybridizations

The Phylogenetic analysis using networks of *COX2* mitochondrial gene sequences from different species of *Saccharomyces* genus has supported the population structure of some "pure" *S. cerevisiae* from Liti *et al.,* (2009) and Schacherer *et al.,* (2009) and hinted the presence of a recombination hotspot in the mitochondrial genome.

The median-joining network based on *COX2* sequences revealed two independent haplogroups for *S. cerevisiae* strains, 1 and 2, which have not shown a relationship depending on country or source of isolation, with the exception of laboratory and bakery strains. This indicates that studies based on *COX2* sequencing and mitochondrial genomes are not proper for

phylogeographical analysis in *S. cerevisiae*. Studies from Liti *et al.,* (2009) and Schacherer *et al.,*(2009) defined different populations for *S. cerevisiae* and *S. paradoxus* coincident with the country or source of isolation. *COX2* sequences can be used to support these analysis based on nuclear genomes. The wine/European, West African and Malaysian pure groups described by Liti *et al.,* (2009) and the wine, sake and laboratory groups from Schacherer *et al.,* (2009), have been found in specific haplogroups. However the strains enclosed in the groups North American and Sake from Liti *et al.,* (2009) have been located in both haplogroups, indicating that these groups are not pure. In the case of heterozygous genes, the sequencing of a monosporic culture drives to the loss of one allele, thus losing information. Probably, in the North American and Sake groups these alleles, that could indicate that this group is not pure, have not been sequenced, as we suspect with *COX2* gene sequence results.

In the case of *S. paradoxus,* we found three haplogroups named Far-East (haplogroup 1), American *(*haplogroup 2) and European (haplogroup 7), as have been described by Liti *et al.,* (2009). Special discussion will be done in American *S. paradoxus* below.

One *S. kudriavzevii* from Japan, the European *S. kudriavzevii*, the triple hybrid CBS 2834 and most of the *S. cerevisiae* x *S. kudriavzevii* hybrids were enclosed in the Haplogroup 3. In addition, *S. arboricolus* and *S. mikatae* have been located in independent positions.

S. bayanus var. *uvarum* (*S. uvarum*) and *S. bayanus* var. *bayanus* (or *S. eubayanus*) sequences were divided into haplogroups 5 and 6, respectively. A recent study have found a pure strain of *S. bayanus* var. *bayanus*, called *S. eubayanus* (Libkind *et al.,* 2011). Our results have shown that *S. uvarum* and *S. eubayanus COX2* sequences have nucleotide differences similar to those found between other *Saccharomyces* species. This supports the idea that *S. eubayanus* (*S. bayanus* var. *bayanus*) and *S.*

uvarum could be different species, and not varieties, as some authors claimed (Libkind *et al.,* 2011).

Analysis of *COX2* gene sequences revealed the presence of a frequent recombination point in the 3' end of the gene. Alignment of the 3' end sequences revealed the presence of intra and interspecific recombinations, involving some species within the *Saccharomyces* genus. Regarding the hybrids *S. cerevisiae* x *S. kudriavzevii* included in the mtDNA types K5, K6 and K10, all of them displayed a recombinant *COX2* end between *S. kudriavzevii* and the European *S. paradoxus*, as we commented in the *mtDNA of S. cerevisiae x S. kudriavzevii hybrids* section.

Some representative strains, of the previous phylogenetic analysis with *COX2* sequences, were selected for sequencing the *COX2* adjacent genes, *ORF1* and *COX3* (not in all species *COX3* is near to *COX2-ORF1-tRNA*). *ORF1* gene encodes a free-standing homing endonuclease (HEG) whose start codon is located in the *COX2* 3'-end. *ORF1* sequences provided information about the length of the recombination occurring in the mitochondrial genome. Gene *COX3*, codifying the subunit 3 of the cytochrome oxidase, could provide information about the mechanism involved in the recombination as this gene is located in a transcriptional unit different to *COX2* and *ORF1*. Mitochondrial gene order differences have been described to occur at transcriptional unit level and at gene level (Groth *et al.,* 2000).

The *ORF1* gene sequencing has revealed two different types of sequences for *ORF1* gene in the *Saccharomyces* species. *ORF1* translation frame could be found truncated by the presence of premature stop codons and GC insertions causing a shift in the open reading frame. In addition to the recombination in *COX2,* other recombination points have also been found in the *ORF1* gene, located near to GC insertions and A+T tandem repeats. The evolution of the *ORF1* gene was difficult to explain

due to the lack of information (sequences) deposited in the databases. Future efforts in this field could elucidate the origin of this maturase like-protein, and other HEGs located in the mitochondrial genomes. We speculate about the ancestrality of *ORF1* type I due to it's observed in most *Saccharomyces* species, and *ORF1* type II could be a derivative of *ORF1* type I or transferred from a *Saccharomycetaceae* member.

No recombination could be found in the sequence of the *COX3* gene, and the phylogenetic analysis showed similar species specific clusters to the ones found with the first segment of *COX2* gene, with the exception of two American *S. paradoxus* now enclosed in the *S. cerevisiae* cluster.

Our results showed frequent recombinant mtDNA in different species and hybrids within the genus *Saccharomyces*. This recombinant mtDNA might be the result of an ancestral hybridization what is supported by the presence of mtDNA from *S. cerevisiae* in some *S. paradoxus* from America, as we demonstrated in *COX2* 5'-end and *COX3* mitochondrial gene sequences. Mitochondrial DNA recombination can be explained due to fusion of mitochondria organelles during cell mating driving to a heteroplasmic state (Berger & Yaffe 2000). However, *Saccharomyces* yeasts are homoplasmic and daughter cells in polar positions could inherit one of the two parent mtDNA whereas daughter cells in the middle of the zygote (medial buds) could inherit recombinant mtDNA after mitochondrial fusion as has been observed in *S. cerevisiae* (Nunnari *et al.,* 1997; Berger & Yaffe 2000).

In the case of two American *S. paradoxus* (120M and CBS 5313), gene sequencing of five nuclear genes have not shown the presence of *S. cerevisiae* sequences (Arias 2008). However, we found a high gene sequence similarity of American *S. paradoxus COX2*-5' end and *COX3* partial gene with *S. cerevisiae* sequences, indicating that mtDNA could be inherited from a parent *S. cerevisiae* after a hybridization with an American

S. paradoxus. A nuclear and mitochondrial genome sequencing project of these *S. paradoxus* strains could demonstrate introgressions between *S. cerevisiae* and American *S. paradoxus*, as previously have been described for other *S. paradoxus* strains (Liti *et al.,* 2006; Muller and McCusker, 2009; Dunn *et al.,* 2012) or the hybrid character of these strains. This has been also observed in hybrids *S. cerevisiae* x *S. kudriavzevii,* as we have described in the *mtDNA of* S. cerevisiae *x* S. kudriavzevii *hybrids* section. At *COX2* level we detected more putative recombinant mtDNA not explored at *ORF1* and *COX3* level, such as IFO 1803 (*S. kudriavzevii* from Japan), IFO 1816 (*S. mikatae*) and other strains of *S. paradoxus*. The complete mtDNA gene sequencing of more strains of *Saccharomyces* will elucidate accurately the evolutionary history of these strains.

Our results suggest that *COX2* recombinant hotspot could be mediated by Orf1p due to a common recombination point that is present in most of the recombinant sequences. Mitochondrial *ORF1* gene has been annotated as a maturase like-protein (or free standing HEG-homing endonuclease gene). Homing endonucleases are selfish genetic elements. A HEG protein seems to mediate a double strand break (DSB) in the HEG less genome allele (HEG^-), whereas the HEG^+ allele is used by the recombinational repair system to repair the break (Burt & Koufopanou 2004). After a hybridization event the two types of mitochondrial genomes, one containing a HEG^+ and the other a HEG^- would fuse, and the homing endonuclease protein could mediate the gene conversion of HEG^- to HEG^+, as have been described in other homing endonucleases at mitochondrial and nuclear level (Burt and Koufopanou, 2004). In this way, the daughter cells would inherit the recombinant version of the mtDNA. Moreover, the lack of recombinant sequences in *COX3* supports that mitochondrial recombination is located in the *COX2-ORF1* region and, in addition, the

high frequency of mitochondrial recombination in the *COX2-ORF1* region supports that a molecular mechanism is involved.

The observation of recombination within *ORF1* gene indicates that other mechanisms are taking place after hybridization. In some cases, a recombination between GC clusters or regions enriched with A+T tandem repeats should have happened. This is supported by other studies where recombination is also mediated by GC clusters and A+T tandem repeats between *S. cerevisiae* mitochondrial genomes (Dieckmann & Gandy 1987; Bouchier *et al.,* 2009; Skelly & Clark-Walker 1991). In this case, both mtDNAs could be HEG^{+} and the highly recombinant sequences, such as A+T tandem repeats and GC clusters, could mediate recombination between the mitochondrial genomes.

GENERAL DISCUSSION

1. The biological species concept in yeast

The occurrence of natural hybrids between *S. cerevisiae* and other species in the genus *Saccharomyces* is in contradiction with the biological species concept in yeast. The presence of *S. cerevisiae* x *S. kudriavzevii* (Gonzalez *et al.,* 2008; Peris *et al.,* 2011a), *S. cerevisiae* x *S. eubayanus* (Dunn & Sherlock 2008; Libkind *et al.,* 2011), *S. uvarum* x *S. eubayanus, S. cerevisiae* x *S. uvarum* (Rainieri *et al.,* 2006) hybrids, horizontal gene transfer between *Saccharomyces* species (Liti *et al.,* 2005), introgressions (Liti *et al.,* 2006; Muller & McCusker, 2009; Dunn *et al.,* 2012) and recombinant mitochondrial DNA indicates that the frequency of hybridization is very high in this genus.

The independent evolution of the *Saccharomyces* species genomes has generated a postzygotic barrier between them, supporting in some cases the Dobzhansky-Muller model as reproductive isolation mechanism (Dobzhansky 1937). However, there is a margin that consents gene survival throughout hybridization under unfavorable living conditions for the parent strains. As we have described in this thesis, hybridization is a mechanism occurring by chance, that leads to the production of new individuals which could be better adapted to new growing conditions where parent strains cannot compete.

A reformulation of the species concept in yeast must be done to be according the different processes that are not taking into account in the BSC. Genotypic cluster criterion could be an alternative to BSC.

2. Yeast hybridization as an adaptive mechanism and its biotechnological applications

Hybridization of two diploid strains can be considered as a Whole Genome Duplication (WGD), originating an allotetraploid. Some hybrids maintain the tetraploid state; however in most cases hybrids are as triploids

or diploids that indicate a diploidization process or the mating of diploid and haploid strains.

Whole Genome Duplications have been described in vertebrates, plants and yeast (Ramsey & Schemske 1998; Otto & Whitton 2000; Wolfe & Shields 1997). Some authors have postulated that WGD in yeasts might have occurred due to hybridization between two ancestral diploid strains giving as a consequence an allotetraploid/polyploid strain (Kellis *et al.,* 2004; Andalis *et al.,* 2004). From this allotetraploid strain would have arisen the present post-WGD yeasts which are adapted to fermentation producing high rates of ethanol in aerobic conditions (Crabtree positives) (Merico *et al.,* 2007; Conant & Wolfe 2007). WGD events have also been associated as a process that emerge after a decrease in the number of species in the world and/or the existence of new niches (Edger & Pires 2009). In the case of yeasts, WGD occurred 100 Ma when angiosperm plants and fruits with high sugar content appeared (Wolfe & Shields 1997; Kellis *et al.,* 2004).

In another way, recent hybridizations, as the ones observed between *Saccharomyces* species, are maintained due to human activity which has made new artificial niches, such as fermentative ones. In the case of *S. cerevisiae* x *S. kudriavzevii* hybrids, as we have explained in this thesis, fermentations at low temperature during winemaking or beer production have made possible the replacement of the less competitive parent strains.

For this reason, data is pointing out that WGD by allotetraploidy process seems to be an excellent adaptive mechanism. If it is demonstrated, that WGD is due to hybridization, in yeast, more than a duplication of the genome, we could understand the consequences of WGDs using artificial hybrids as a model.

Natural and artificial hybrids have intermediate physiological properties compared with their parental species (Greig *et al.,* 2002; Belloch *et al.,* 2008). In environmental conditions where the parental strains are not well

adapted, as low temperature for *S. cerevisiae* and higher ethanol concentrations for *S. kudriavzevii,* hybrids between these two species would be maintained in the population and finally replace their parent. Moreover, hybrids might produce wines with different organoleptic properties when compared with the parental strains, as in the case of aromas (Gangl *et al.,* 2009; Gamero *et al.,* 2011). This appears to be one of the reasons driving to an unconscious human selection of these strains/hybrids. This opens the possibility to originate personalized hybrids to improve the actual alcoholic beverages in the market. Generation of new hybrid strains for winemaking, beer and bioethanol production, detoxification or bakery industry is the new goal of the biotechnological companies.

3. Future perspectives

We have discussed the relevance of hybridization for the evolution of yeast and their biotechnological application.

Understanding how the mechanism of hybridization works and how hybrid genomes evolve is key for the application of hybridization to create new biotechnological strains. Questions that have remained unresolved at the end of this thesis are: i) What are the real parents of the different hybrids found in nature? ii) What are the ecological properties where hybrids are obtained and who has spread them around Europe? iii) How are protein interaction networks rewired to maintain the cell active and what consequences it has at the physiological level? iv) Although hybrids are maintained in the population by clonal division (mitosis), evolved hybrids could originate new species able to divide by meiosis (recovering the sexual character)?

Finally, in the case of the HEG, *ORF1,* from mitochondrial DNA some questions remain unclear: i) what is the mechanism of *ORF1* infection and its origin? ii) Could it have a function in the cell?

Future questions regarding the hybrids will be answered by applying next generation sequencing, functional genomics, proteome analysis and experimental evolution.

FINAL CONCLUSIONS

The general conclusions deduced from this doctoral thesis are summarized as follows:

- The geographical distribution of natural *S. cerevisiae* x *S. kudriavzevii* hybrids is limited to European regions of Oceanic and Continental climate characterized by cold winters and warm and dry summers. Expansion of wine *S. cerevisiae* x *S. kudriavzevii* hybrids to Southern European regions has brought them to these regions, corresponding to the northern limit of vine growing in Europe.
- Hybridization might be a very important mechanism of evolution in yeasts as hybrids seem to be better adapted to new environments or fermentative conditions, successfully replacing the parental strains
- In general, genome compositions of *S. cerevisiae* x *S. kudriavzevii* hybrids show a trend to maintain a complete copy of the parental *S. cerevisiae* genome and to lose the *S. kudriavzevii* fraction.
- The genome structure of natural *S. cerevisiae* x *S. kudriavzevii* hybrids is highly diverse. Their DNA content varies from 3.0 to 4.0 times that of the haploid *S. cerevisiae* reference strain. All hybrids showed differences in the number of chromosomal rearrangements, varying from genomes without rearrangements to genomes highly rearranged indicating different evolutionary times for hybrid formation.
- Chimerical chromosomes appear to be the result of chromosomal crossing over between homologous chromosomes mediated by highly recombinogenic sequences, such as ARS, Ty elements, Y' elements, rRNA regions and conserved coding genes, activating the mismatch repair system (MMR) mechanism.
- The role of *S. cerevisiae* parental genes in the hybrid seems to be the maintenance of a high fermentative capability whereas the role of *S.*

kudriavzevii parental genes related with adaptation to stress during fermentation at lower temperatures.

- Most natural *S. cerevisiae* x *S. kudriavzevii* hybrids inherited their mtDNA from the *S. kudriavzevii* parent, which imposes a restriction to the loss of those *S. kudriavzevii* nuclear genes involved in mtDNA functions. The few natural *S. cerevisiae* x *S. kudriavzevii* hybrids bearing mtDNA from the *S. cerevisiae* parent exhibit a higher reduction of the *S. kudriavzevii* subgenome. These results support cytonuclear incompatibilities in the *Saccharomyces* hybrid genomes.
- At least six different hybridization events are necessary to explain the origin of the natural double hybrids *S. cerevisiae* x *S. kudriavzevii.* Brewing strains (except CECT 11003 and 11004) and MR25 likely derived from the same hybridization event. Similarly, Swiss wine hybrids and brewing hybrids CECT 11003 and 11004 share a common hybrid ancestor. Austrian hybrids, Vin7 and SOY3, clustered together indicating also a common origin. The dietary supplement IF6 seems to be similar to hybrids participating in the formation of brewing and wine hybrids. Finally, PB7, AMH and triple hybrids seem to derive from independent hybridization events.
- In most cases, the parental strains involved in hybridization were European wine strains of *S. cerevisiae* and European *S. kudriavzevii* yeasts. The exception maybe the hybrid ancestor of brewing yeasts which could derived from a European heterozygous ale *S. cerevisiae* yeast.
- Most *S. cerevisiae* x *S. kudriavzevii* hybrids originated by rare mating between diploid *S. cerevisiae* strains and haploid *S. kudriavzevii* cells. However, wine hybrid PB7 likely was the result of a hybridization event between diploid *S. cerevisiae* and *S. kudriavzevii* cells; and strain AMH appears to be the result of two successive hybridization events, the first involving a diploid *S. cerevisiae* and a haploid *S. kudriavzevii* and the

second concerning the cross between the result of the first hybridization and a diploid *S. cerevisiae*.

- The recombination hotspot found in *COX2* mtDNA could be the result of the action of the neighboring homing endonuclease encoded by gene ORF1. *Orf1*p activity could have been lost several times in different *S. cerevisiae* and *S. paradoxus* strains but recovered after mitochondrial fusion due to hybridization event between the two species.
- Internal recombinations in *ORF1* and other uncommon recombining regions in *COX2* could be mediated by GC clusters and A+T rich regions.

REFERENCES

Aa E, Townsend JP, Adams RI, Nielsen KM, & Taylor JW (2006) Population structure and gene evolution in *Saccharomyces cerevisiae*. FEMS Yeast Res 6: 702-715.

Acuña G, Würgler FE, & Sengstag C (1994) Reciprocal mitotic recombination is the predominant mechanism for the loss of a heterozygous gene in *Saccharomyces cerevisiae*. Environ Mol Mutagen 24: 307-316.

Aguilera J, Randez-Gil F, & Prieto JA (2007) Cold response in *Saccharomyces cerevisiae*: new functions for old mechanisms. FEMS Microbiol Rev 31: 327-341.

Andalis AA, Storchova Z, Styles C, Galitski T, Pellman D, & Fink GR (2004) Defects Arising From Whole-Genome Duplications in *Saccharomyces cerevisiae*. Genetics 167: 1109-1121.

Andersen MP, Nelson ZW, Hetrick ED, & Gottschling DE (2008) A genetic screen for increased loss of heterozygosity in *Saccharomyces cerevisiae*. Genetics 179: 1179-1195.

Andersen SL & Sekelsky J (2010) Meiotic versus mitotic recombination: Two different routes for double-strand break repair. BioEssays 32: 1058-1066.

Antunovics Z, Nguyen HV, Gaillardin C, & Sipiczki M (2005) Gradual genome stabilisation by progressive reduction of the *Saccharomyces uvarum* genome in an interspecific hybrid with *Saccharomyces cerevisiae*. FEMS Yeast Res 5: 1141-1150.

Arias A (2008) Diversidad genética de las especies del complejo *Saccharomyces sensu stricto* de fermentaciones tradicionales. University of Valencia. Available at: http://www.tesisenred.net/bitstream/handle/10803/9940/arias.pdf?sequence=1. Accessed on 17th of September 2012.

Arroyo-García R, Ruiz-García L, Bolling L, Ocete R, López MA, *et al.,* (2006) Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp. *sativa*) based on chloroplast DNA polymorphisms. Mol Ecol 15: 3707-3714.

Arroyo-Lopez FN, Orlic S, Querol A, & Barrio E (2009) Effects of temperature, pH and sugar concentration on the growth parameters of *Saccharomyces cerevisiae*, *S. kudriavzevii* and their interspecific hybrid. Int J Food Microbiol 131: 120-127.

Arroyo-López FN, Salvadó Z, Tronchoni J, Guillamón JM, Barrio E, & Querol A (2010) Susceptibility and resistance to ethanol in *Saccharomyces* strains isolated from wild and fermentative environments. Yeast 27: 1005-1015.

Arroyo-López FN, Pérez-Través L, Querol A, & Barrio E (2011) Exclusion of *Saccharomyces kudriavzevii* from a wine model system mediated by *Saccharomyces cerevisiae*. Yeast 28: 423-435.

Badr A, Müller K, Schäfer-Pregl R, Rabey HE, Effgen S *et al.,* (2000) On the origin and domestication history of Barley (*Hordeum vulgare*). Mol Biol Evol 17: 499-510.

Bandelt HJ, Forster P, & Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16: 37-48.

Barrio E, González SS, Arias A, Belloch C, & Querol A (2006) Molecular mechanisms involved in the adaptive evolution of industrial yeasts. In *Yeasts in Food and* Beverages, (Querol A & Fleet GH, eds), pp. 153-174. Springer-Verlag, Berlin.

Barros Lopes M, Bellon JR, Shirley NJ, & Ganter PF (2002) Evidence for multiple interspecific hybridization in *Saccharomyces sensu stricto* species. FEMS Yeast Res 1: 323-331.

Basse CW (2010) Mitochondrial inheritance in fungi. Curr Opin Microbiol 13: 712-719.

Belfort M (2003) Two for the price of one: a bifunctional intron-encoded DNA endonuclease-RNA maturase. Gene Dev 17: 2860-2863.

Belloch C, Querol A, Garcia MD, & Barrio E (2000) Phylogeny of the genus *Kluyveromyces* inferred from the mitochondrial cytochrome-c oxidase II gene. Int J Syst Evol Microbiol 50: 405-416.

Belloch C, Orlic S, Barrio E, & Querol A (2008) Fermentative stress adaptation of hybrids within the *Saccharomyces sensu stricto* complex. Int J Food Microbiol 122: 188-195.

Belloch C, Pérez-Torrado R, González SS, Pérez-Ortin JE, García-Martínez J*, et al.,* (2009) The chimerical genomes of natural hybrids between *Saccharomyces cerevisiae* and *Saccharomyces kudriavzevii*. Appl Environ Microb 75: 2534-2544.

Beltran G, Novo M, Guillamón JM, Mas A, & Rozès N (2008) Effect of fermentation temperature and culture media on the yeast lipid composition and wine volatile compounds. Int J Food Microbiol 121: 169-177.

Beltran G, Novo M, Leberre V, Sokol S, Labourdette D, *et al.,* (2006) Integration of transcriptomic and metabolic analyses for understanding the global responses of low-temperature winemaking fermentations. FEMS Yeast Res 6: 1167-1183.

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acid Res 27: 573-580.

Berger KH & Yaffe MP (2000) Mitochondrial DNA inheritance in *Saccharomyces cerevisiae*. Trends Microbiol 8: 508-513.

Birky CW, Arlene RA, Rosemary D, & Maureen C (1982) Mitochondrial transmission genetics: replication, recombination, and segregation of mitochondrial DNA and its inheritance in crosses. In *Mitochondrial Genes.* Cold Spring Harbor Monograph Archive.

Bond U (2009) The Genomes of Lager Yeasts. In *Advances in Applied Microbiology* (Allen & Laskin, eds.) Volume 69 pp. 159-182. Academic Press.

Bordonné R, Dirheimer G, & Martin RP (1988) Expression of the oxi1 and maturase-related RF1 genes in yeast mitochondria. Curr Gen 13: 227-233.

Borneman AR, Desany BA, Riches D, Affourtit JP, Forgan AH, *et al.,* (2011) Whole-Genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. PLoS Genet 7: e1001287-

Borneman AR, Desany BA, Riches D, Affourtit JP, Forgan AH, *et al.,* (2012) The genome sequence of the wine yeast VIN7 reveals an allotriploid hybrid genome with *Saccharomyces cerevisiae* and *Saccharomyces kudriavzevii* origins. FEMS Yeast Res 12: 88-96.

Botosso VF, Zanotto PM, Ueda M, Arruda E, Gilio AE, *et al.,* (2009) Positive selection results in frequent reversible amino acid replacements in the G protein gene of human respiratory syncytial virus. PLoS Pathog 5: e1000254.

Bouchier C, Ma L, Créno S, Dujon B, & Fairhead C (2009) Complete mitochondrial genome sequences of three *Nakaseomyces* species reveal invasion by palindromic GC clusters and considerable size expansion. FEMS Yeast Res 9: 1283-1292.

Bradbury J, Richards K, Niederer H, Lee S, Rod Dunbar P, & Gardner R (2006) A homozygous diploid subset of commercial wine yeast strains. A van Leeuw 89: 27-37.

REFERENCES

Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, *et al.,* (2003) ArrayExpress--a public repository for microarray gene expression data at the EBI. Nucleic Acid Res 31: 68-71.

Briggs DE, Boulton CA, Brookes PA, & Stevens R (2004) *Brewing: science and practice*. Woodhead, Cambridge, UK.

Bull JJ, Huelsenbeck JP, Cunningham CW, Swofford DL, & Waddell PJ (1993) Partitioning and combining data in phylogenetic analysis. Syst Biol 42: 384-397.

Burt A & Koufopanou V (2004) Homing endonuclease genes: the rise and fall and rise again of a selfish element. Curr Opin Genet Dev 14: 609-615.

Burton J & Kerr J (2011) *The Cistercians in the Middle Ages*. Woodbridge, UK: The Boydell Press.

Butler G, Rasmussen MD, Lin MF, Santos MAS, Sakthikumar S, *et al.,* (2009) Evolution of pathogenicity and sexual reproduction in eight Candida genomes. Nature 459: 657-662.

Byrne KP & Wolfe KH (2005) The Yeast Gene Order Browser: Combining curated homology and syntenic context reveals gene fate in polyploid species. Genome Res 15: 1456-1461.

Carrasco P, Querol A, & del Olmo Ml (2001) Analysis of the stress resistance of commercial wine yeast strains. Arch Microbiol 175: 450-457.

Carreto L, Eiriz M, Gomes A, Pereira P, Schuller D, & Santos M (2008) Comparative genomics of wild type yeast strains unveils important genome diversity. BMC Genomics 9: 524.

Casaregola S, Nguyen HV, Lapathitis G, Kotyk A, & Gaillardin C (2001) Analysis of the constitution of the beer yeast genome by PCR, sequencing and subtelomeric sequence hybridization. Int J Syst Evol Micr 51: 1607-1618.

Casaregola S, Weiss S, & Morel G (2011) New perspectives in hemiascomycetous yeast taxonomy. C R Biol 334: 590-598.

Chambers PJ & Pretorius IS (2010) Fermenting knowledge: the history of winemaking, science and yeast research. EMBO Rep advance online publication:

Chambers SR, Hunter N, Louis EJ, & Borts RH (1996) The mismatch repair system reduces meiotic homeologous recombination and stimulates recombination-dependent chromosome loss. Mol Cell Biol 16: 6110-6120.

Chou JY, Hung YS, Lin KH, Lee HY, & Leu JY (2010) Multiple molecular mechanisms cause reproductive isolation between three yeast species. PLoS Biol 8: e1000432.

Ciani M, Mannazzu I, Marinangeli P, Clementi F, & Martini A (2004) Contribution of winery-resident *Saccharomyces cerevisiae* strains to spontaneous grape must fermentation. A van Leeuw 85: 159-164.

Clark-Walker GD (1989) In vivo rearrangement of mitochondrial DNA in *Saccharomyces cerevisiae*. PNAS 86: 8847-8851.

Clark-Walker GD (1992) Evolution of Mitochondrial Genomes in Fungi. In *International Review of Cytology* (David, Wolstenholme & Kwang, eds.) Volume 141 pp. 89-127. Academic Press.

Cliften P, Sudarsanam P, Desikan A, Fulton L, Fulton B, *et al.,* (2003) Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. Science 301: 71-76.

Colleaux L, d'Auriol L, Betermier M, Cottarel G, Jacquier A, *et al.,* (1986) Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into E. coli as a specific double strand endonuclease. Cell 44: 521-533.

Corran HS (1975) *A history of brewing*. Newton Abbot, David & Charles, London.

Coton E, Coton M, Levert D, Casaregola S, & Sohier D (2006) Yeast ecology in French cider and black olive natural fermentations. Int J Food Microbiol 108: 130-135.

Courtray A (1986) *Histoire de La Valsainte*. Edition Chartreuse de La Valsainte, Fribourg, Switzerland.

Coyne J & Orr H (1998) The evolutionary genetics of speciation. Proc R Soc Lond353: 287-305.

Cracraft J (1989) Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. In *Speciation and its Consequences,* (Otte D & Endler JA, eds) pp. 28-59. Sinauer Associates, Sunderland, Massachusetts.

Dalgaard JZ, Klar AJ, Moser MJ, Holley WR, Chatterjee A, & Mian IS (1997) Statistical modeling and analysis of the LAGLIDADG family of site- specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. Nucleic Acid Res 25: 4626-4638.

de Llanos R, Querol A, Planes AM, & Fernández-Espinar MT (2004) Molecular characterization of clinical *Saccharomyces cerevisiae* isolates and their association with non-clinical strains. Syst Appl Microbiol 27: 427-435.

De Vero L, Pulvirenti A, Gullo M, Bonatti P, & Giudici P (2003) Sorting of mitochondrial DNA and proteins in the progeny of *Saccharomyces* interspecific hybrids. Annals of Microbiology 53: 219-231.

de Zamaroczy M & Bernardi G (1986) The GC clusters of the mitochondrial genome of yeast and their evolutionary origin. Gene 41: 1-22.

Delneri D, Colson I, Grammenoudi S, Roberts IN, Louis EJ, & Oliver SG (2003) Engineering evolution to study speciation in yeasts. Nature 422: 68-72.

Delport W, Poon AFY, Frost SDW, & Kosakovsky Pond SL (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26: 2455-2457.

Demuyter C, Lollier M, Legras JL, & Le Jeune C (2004) Predominance of *Saccharomyces uvarum* during spontaneous alcoholic fermentation, for three consecutive years, in an Alsatian winery. J Appl Microbiol 97: 1140-1148.

Despons L, Baret P, Frangeul L, Louis V, Durrens P, & Souciet JL (2010) Genome-wide computational prediction of tandem gene arrays: application in yeasts. BMC Genomics 11: 56.

Di Rienzi SC, Collingwood D, Raghuraman MK, & Brewer BJ (2009) Fragile genomic sites are associated with origins of replication. Genome Biol Evol 1: 350-363.

Dieckmann C & Gandy B (1987) Preferential recombination between GC clusters in yeast mitochondrial DNA. EMBO J 6: 4197-4203.

Dietrich FS, Voegeli S, Brachat S, Lerch A, Gates K, *et al.,* (2004) The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* Genome. Science 304: 304-307.

Dobzhansky T (1937) Genetic nature of species differences. American Naturalist 71: 404-420.

Doniger S, Salomonis N, Dahlquist K, Vranizan K, Lawlor S, & Conklin B (2003) MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. Genome Biol 4: R7.

Doniger SW, Kim HS, Swain D, Corcuera D, Williams M, *et al.,* (2008) A catalog of neutral and deleterious polymorphism in yeast. PLoS Genet 4: e1000183.

Doolittle WF & Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. Nature 284: 601-603.

Drinnenberg IA, Weinberg DE, Xie KT, Mower JP, Wolfe KH, *et al.,* (2009) RNAi in budding yeast. Science 326: 544-550.

Dujon B, Slonimski PP, & Weill L (1974) Mitochondrial genetics IX: A model for recombination and segregation of mitochondrial genomes in *Saccharomyces cerevisiae*. Genetics 78: 415-437.

Dujon B (2006) Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. Trends Gen : TIG 22: 375-387.

Dujon B (2010) Yeast evolutionary genomics. Nat Rev Genet 11: 512-524.

Dunn B & Sherlock G (2008) Reconstruction of the genome origins and evolution of the hybrid lager yeast *Saccharomyces pastorianus*. Genome Res 18: 1610-1623.

Dunn B, Levine RP, & Sherlock G (2005) Microarray karyotyping of commercial wine yeast strains reveals shared, as well as unique, genomic signatures. BMC Genomics 6: 53.

Dunn B, Richter C, Kvitek DJ, Pugh T, & Sherlock G (2012) Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. Genome Res online accession.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acid Res 32: 1792-1797.

Edger P & Pires J (2009) Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome Res17: 699-717.

Erasmus DJ, van der Merwe GK, & van Vuuren HJJ (2003) Genome-wide expression analyses: Metabolic adaptation of *Saccharomyces cerevisiae* to high sugar stress. FEMS Yeast Res 3: 375-399.

Erny C, Raoult P, Alais A, Butterlin G, Delobel P, *et al.,* (2012) Ecological success of a group of *Saccharomyces cerevisiae*/*Saccharomyces kudriavzevii* hybrids in the Northern European wine making environment. Appl Environ Microb online accession.

Fabrizio P & Longo VD (2003) The chronological life span of *Saccharomyces cerevisiae*. Aging Cell 2: 73-81.

Fay JC & Benavides JA (2005) Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. PLoS Genet 1: e5.

Feder JL (1998) The apple maggot fly, *Rhagoletis pomonella*: Flies in the face of conventional wisdom. In *Endless Forms. Species and Speciation* (Howard DJ & Berlocher SH, eds) pp. 130-144. Oxford University Press, New York.

Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.69. University of Washington, Seattle.

Fernández-Espinar M, Esteve-Zarzoso B, Querol A, & Barrio E (2000) RFLP analysis of the ribosomal internal transcription spacers and the 5.8S rRNA gene region of the genus *Saccharomyces*: a fast method for species identification and differentiation of flor yeasts. A van Leeuw 78: 87-97.

Fernández-Espinar M, Barrio E, & Querol A (2003) Analysis of the genetic variability in the species of the *Saccharomyces sensu stricto* complex. Yeast 20: 1213-1226.

Fischer G, James SA, Roberts IN, Oliver SG, & Louis EJ (2000) Chromosomal evolution in *Saccharomyces*. Nature 405: 451-454.

Fitzpatrick DA (2011) Horizontal gene transfer in fungi. FEMS Microbiol Lett 329: 1-8.

Fogel S & Welch JW (1982) Tandem gene amplification mediates copper resistance in yeast. PNAS 79: 5342-5346.

Fonty G, Goursot R, Wilkie D, & Bernardi G (1978) The mitochondrial genome of wild-type yeast cells: VII. Recombination in crosses. J Mol Biol 119: 213-235.

Gamero A, Manzanares P, Querol A, & Belloch C (2011) Monoterpene alcohols release and bioconversion by *Saccharomyces* species and hybrids. Int J Food Microbiol 145: 92-97.

Gibbs MJ, Armstrong JS, & Gibbs AJ (2000) Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16: 573-582.

Goddard MR & Burt A (1999) Recurrent invasion and extinction of a selfish gene. PNAS 96: 13880-13885.

Gogarten P & Townsend J (2005) Horizontal gene transfer, genome innovation and evolution. Nat Rev Microbiol 3: 679-687.

Gojkovic Z, Knecht W, Zameitat E, Warneboldt J, Coutelis JB, *et al.,* (2004) Horizontal gene transfer promoted evolution of the ability to propagate under anaerobic conditions in yeasts. Mol Genet Genomics 271: 387-393.

González SS, Barrio E, Gafner J, & Querol A (2006) Natural hybrids from *Saccharomyces cerevisiae*, *Saccharomyces bayanus* and *Saccharomyces kudriavzevii* in wine fermentations. FEMS Yeast Res 6: 1221-1234.

González SS, Barrio E, & Querol A (2008) Molecular characterization of new natural hybrids between *S. cerevisiae* and *S. kudriavzevii* from brewing. Appl Environ Microb 74: 2314-2320.

Greig D, Louis EJ, Borts RH, & Travisano M (2002) Hybrid speciation in experimental populations of yeast. Science 298: 1773-1775.

Greig D & Travisano M (2003) Haploid superiority. Science 299: 524-525.

Greig D (2008) Reproductive isolation in *Saccharomyces*. Heredity 102: 39-44.

Gromicho M, Coutanceau JP, Ozouf-Costaz C, & Collares-Pereira M (2006) Contrast between extensive variation of 28S rDNA and stability of 5S rDNA and telomeric repeats in the diploid-polyploid Squalius alburnoides complex and in its maternal ancestor *Squalius pyrenaicus* (Teleostei, Cyprinidae). Chromosome Res14: 297-306.

Groth C, Hansen J, & Piskur J (1999) A natural chimeric yeast containing genetic material from three species. Int J Syst Bacteriol 49: 1933-1938.

Groth C, Petersen RF, & Piskur J (2000) Diversity in organization and the origin of gene orders in the mitochondrial DNA molecules of the genus *Saccharomyces*. Mol Biol Evol 17: 1833-1841.

Guindon S & Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by Maximum Likelihood. Syst Biol 52: 696-704.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, & Gascuel O (2010) New algorithms and methods to estimate Maximum-Likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59: 307-321.

Gunge N & Nakatomi Y (1972) Genetic mechanisms of rare matings of the yeast *Saccharomyces cerevisiae* heterozygous for mating type. Genetics 70: 41-58.

Haase SB & Reed SI (2002) Improved flow cytometric analysis of the budding yeast cell cycle. Cell Cycle 1: 117-121.

Hammond J (1993) Brewers yeast. In *The Yeasts* (Rose AH & Harrison JS, eds) pp. 7-67. Academic Press, London, UK.

Hansen J & Kielland-Brandt MC (1994) *Saccharomyces carlsbergensis* contains two functional MET2 alleles similar to homologues from *S. cerevisiae* and *S. monacensis*. Gene 140: 33-40.

Henschke PA & Rose AH (1991) Plasma membrane. In *The Yeasts: Yeast Organelles IV* (Rose AH & Harrison JS, eds) pp. 297-345. Academic Press Limited, London, UK.

Higgins VJ, Beckhouse AG, Oliver AD, Rogers PJ, & Dawes IW (2003) Yeast genome-wide expression analysis identifies a strong ergosterol and oxidative stress response during the initial stages of an industrial lager fermentation. Appl Environ Microb 69: 4777-4787.

Hittinger CT, Goncalves P, Sampaio JP, Dover J, Johnston M, & Rokas A (2010) Remarkably ancient balanced polymorphisms in a multi-locus gene network. Nature 464: 54-58.

Hunter N, Chambers SR, Louis EJ, & Borts RH (1996) The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. EMBO J 15: 1726-1733.

Huson DH, Tobias D, Tobias K, & Steel MA (2004) Phylogenetic Super-Networks from partial trees. IEEE/ACM Trans Comput Biol Bioinformatics 1: 151-158.

Huson DH & Bryant D (2006) Application of Phylogenetic Networks in evolutionary studies. Mol Biol Evol 23: 254-267.

Ivorra C, Pérez-Ortín JE, & del Olmo Ml (1999) An inverse correlation between stress resistance and stuck fermentations in wine yeasts. A molecular study. Biotechnol Bioeng 64: 698-708.

Johnson LJ, Koufopanou V, Goddard MR, Hetherington R, Schafer SM, & Burt A (2004) Population genetics of the wild yeast *Saccharomyces paradoxus*. Genetics 166: 43-52.

Joly S, Rauscher JT, Broyles S, Brown AHD, & Doyle JJ (2004) Evolutionary dynamics and preferential expression of homeologous 18S-5.8S-26S nuclear ribosomal genes in natural and artificial Glycine allopolyploids. Mol Biol Evol 21: 1409-1421.

Jones G, White M, Cooper O, & Storchmann K (2005) Climate change and global wine quality. Climatic Change 73: 319-343.

Kellis M, Birren BW, & Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature 428: 617-624.

Kellis M, Patterson N, Endrizzi M, Birren B, & Lander ES (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423: 241-254.

Kim JM, Vanguri S, Boeke JD, Gabriel A, & Voytas DF (1998) Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. Genome Res 8: 464-478.

Kishino H & Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. J Mol Evol 29: 170-179.

Kodama Y, Kielland-Brandt MC, & Hansen J (2005) Lager brewing yeast. In *Comparative Genomics using fungi as* models (Sunnerhagen p & Piskur J, eds.) pp. 145-164. Springer-Verlag, Berlin, Germany.

Koufopanou V, Hughes J, Bell G, & Burt A (2006) The spatial scale of genetic differentiation in a model organism: the wild yeast *Saccharomyces paradoxus*. Philosophical Transactions B 361: 1941-1946.

Kuehne HA, Murphy HA, Francis C, & Sniegowski PD (2007) Allopatric divergence, secondary contact, and genetic isolation in wild yeast populations. Curr Biol 17: 407-411.

Kurtzman CP & Fell JW (1998) *The Yeasts, A taxonomic study*. 4th ed. Elservier Science, Amsterdam.

Kurtzman CP & Robnett CJ (2003) Phylogenetic relationships among yeasts of the '*Saccharomyces* complex' determined from multigene sequence analyses. FEMS Yeast Res 3: 417-432.

Kurtzman CP, Fell JW, & Boekhout T (2011) *The Yeasts: A Taxonomic Study*. 5th ed. Elservier, Amsterdam.

Langkjaer RB, Casaregola S, Ussery DW, Gaillardin C, & Piskur J (2003) Sequence analysis of three mitochondrial DNA molecules reveals interesting differences among *Saccharomyces* yeasts. Nucleic Acid Res 31: 3081-3091.

Le Jeune C, Lollier M, Demuyter C, Erny C, Legras JL, *et al.,* (2007) Characterization of natural hybrids of *Saccharomyces cerevisiae* and *Saccharomyces bayanus* var. *uvarum*. FEMS Yeast Res 7: 540-549.

Lee HY, Chou JY, Cheong L, Chang NH, Yang SY, & Leu JY (2008) Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. Cell Mol Life Sci 135: 1065-1073.

Legras JL, Merdinoglu D, Cornuet JM, & Karst F (2007) Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. Mol Ecol 16: 2091-2102.

Libkind D, Hittinger CT, Valério E, Gonçalves C, Dover J, *et al.,* (2011) Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. PNAS 108: 14539-14544.

Librado P & Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451-1452.

Ling F & Shibata T (2004) Mhr1p-dependent concatemeric mitochondrial DNA formation for generating yeast mitochondrial homoplasmic cells. Mol Biol Cell 15: 310-322.

Liti G, Peruffo A, James SA, Roberts IN, & Louis EJ (2005) Inferences of evolutionary relationships from a population survey of LTR-retrotransposons and telomeric-associated sequences in the *Saccharomyces sensu stricto* complex. Yeast 22: 177-192.

Liti G, Barton DB, & Louis EJ (2006) Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. Genetics 174: 839-850.

Liti G, Carter DM, Moses AM, Warringer J, Parts L, *et al.,* (2009) Population genomics of domestic and wild yeasts. Nature 458: 337-341.

Lopandic K, Wallner E, Tscheik G, Leitner G, Querol A, *et al.,* (2007) Genetically different wine yeasts isolated from Austrian vine-growing regions influence wine aroma differently and contain putative hybrids between *Saccharomyces cerevisiae* and *Saccharomyces kudriavzevii*. FEMS Yeast Res 7: 953-965.

Lopes CA, Barrio E, & Querol A (2010) Natural hybrids of S. *cerevisiae* x S. *kudriavzevii* share alleles with European wild populations of *Saccharomyces kudriavzevii*. FEMS Yeast Res 10: 412-421.

MacAlpine DM, Perlman PS, & Butow RA (1998) The high mobility group protein Abf2p influences the level of yeast mitochondrial DNA recombination intermediates in vivo. PNAS 95: 6739-6743.

Maclean C & Greig D (2008) Prezygotic reproductive isolation between *Saccharomyces cerevisiae* and *Saccharomyces paradoxus*. BMC Evol Biol 8: 1.

Mallet J (1995) A species definition for the modern synthesis. Trends Ecol Evol 10: 294-299.

Martin D & Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. Bioinformatics 16: 562-563.

Martin DP, Lemey P, Lott M, Moulton V, Posada D, & Lefeuvre P (2010) RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26: 2462-2463.

Martini A (1993) Origin and domestication of the wine yeast *Saccharomyces cerevisiae*. J Wine Res 4: 165-176.

Masneuf I, Hansen J, Groth C, Piskur J, & Dubourdieu D (1998) New hybrids between *Saccharomyces Sensu stricto* yeast species found among wine and cider production strains. Appl Environ Microb 64: 3887-3892.

Mayr E (1942) *Systematics and the Origin of Species*. Columbia University Press, New York.

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

McGovern PE (2010) *Uncorking the Past: The quest for wine, beer, and other alcoholic beverages*. University of California Press, California and London.

McGovern PE, Hartung U, Badler VR, Glusker DL, & Exner LJ (1997) The beginning of winemaking and viniculture in the ancient Near East and Egypt. Expedition 39: 3-21.

McGovern PE, Zhang J, Tang J, Zhang Z, Hall GR, *et al.,* (2004) Fermented beverages of pre- and proto-historic China. PNAS 101: 17593-17598.

Meinhardt F, Kempken F, Kämper J, & Esser K (1990) Linear plasmids among eukaryotes: fundamentals and application. Curr Gen 17: 89-95.

Merico A, Sulo P, Piskur J, & Compagno C (2007) Fermentative lifestyle in yeasts belonging to the *Saccharomyces* complex. FEBS Journal 274: 976-989.

Michel F (1984) A maturase-like coding sequence downstream of the *OXI2* gene of yeast mitochondrial DNA is interrupted by two GC clusters and a putative end-of-messenger signal. Curr Gen 8: 307-317.

Michel RH, McGovern PE, & Badler VR (1992) Chemical evidence for ancient beer. Nature 360: 24-24.

Mortimer RK (2000) Evolution and variation of the yeast (*Saccharomyces*) genome. Genome Res 10: 403-409.

Mortimer RK, Romano P, Suzzi G, & Polsinelli M (1994) Genome renewal: A new phenomenon revealed from a genetic study of 43 strains of

Saccharomyces cerevisiae derived from natural fermentation of grape musts. Yeast 10: 1543-1552.

Muller LAH & McCusker JH (2009) A multispecies-based taxonomic microarray reveals interspecies hybridization and introgression in *Saccharomyces cerevisiae*. FEMS Yeast Res 9: 143-152.

Murphy N, Banks JC, Whitfield JB, & Austin AD (2008) Phylogeny of the parasitic microgastroid subfamilies (Hymenoptera: Braconidae) based on sequence data from seven genes, with an improved time estimate of the origin of the lineage. Mol Phylogenet Evol 47: 378-395.

Myers CL, Dunham MJ, Kung SY, & Troyanskaya OG (2004) Accurate detection of aneuploidies in array CGH and gene expression microarray data. Bioinformatics 20: 3533-3543.

Nakagawa K, Morishima N, & Shibata T (1991) A maturase-like subunit of the sequence-specific endonuclease endo.SceI from yeast mitochondria. J Biol Chem 266: 1977-1984.

Nakagawa K, Morishima N, & Shibata T (1992) An endonuclease with multiple cutting sites, Endo.SceI, initiates genetic recombination at its cutting site in yeast mitochondria. EMBO J 11: 2707-2715.

Nakao Y, Kanamori T, Itoh T, Kodama Y, Rainieri S, *et al.,* (2009) Genome sequence of the lager brewing yeast, an interspecies hybrid. DNA Res 16: 115-129.

Naumov GI (1987) Genetic basis for classification and identification of the ascomycetous yeasts. Stud Mycol. 30: 469-475.

Naumov GI, Naumova ES, & Louis EJ (1995) A new genetically isolated population of the *Saccharomyces sensu stricto* complex from Japan. J Gen App Microbiol 41: 499-505.

Naumov GI, Naumova ES, Hagler AN, Mendonça-Hagler LC, & Louis EJ (1995) A new genetically isolated population of the *Saccharomyces sensu stricto* complex from Brazil. A van Leeuw 67: 351-355.

Naumov GI (1996) Genetic identification of biological species in the *Saccharomyces sensu stricto* complex. J Ind Microbiol Biot 17: 295-302.

Naumov GI, Naumova ES, & Sniegowski PD (1997) Differentiation of European and Far East Asian populations of *Saccharomyces paradoxus* by allozyme analysis. Int J Syst Bacteriol 47: 341-344.

Naumov GI (2000) *Saccharomyces bayanus* var. *uvarum* comb. nov. a new variety established by genetic analysis. Mikrobiologiia 69: 410-414.

Naumov GI, James SA, Naumova ES, Louis EJ, & Roberts IN (2000) Three new species in the *Saccharomyces sensu stricto* complex: *Saccharomyces cariocanus*, *Saccharomyces kudriavzevii* and *Saccharomyces mikatae*. Int J Syst Evol Micr 50: 1931-1942.

Naumov GI, Masneuf I, Naumova S., Aigle M, & Dubourdieu D (2000) Association of *Saccharomyces bayanus* var. *uvarum* with some French wines: genetic analysis of yeast populations. Res Microbiol 151: 683-691.

Naumov GI, Nguyen H-V, Naumova ES, Michel A, Aigle M, & Gaillardin C (2001) Genetic identification of *Saccharomyces bayanus* var. *uvarum*, a cider-fermenting yeast. Int J Food Microbiol 65: 163-171.

Naumov GI, Naumova ES, Antunovics Z, & Sipiczki M (2002) *Saccharomyces bayanus* var. *uvarum* in Tokaj wine-making of Slovakia and Hungary. App Microbiol Biot 59: 727-730.

Naumov GI, Naumova ES, & Masneuf-Pomarède I (2010) Genetic identification of new biological species *Saccharomyces arboricolus* Wang et Bai. A van Leeuw 98: 1-7.

Naumova E, Naumov GI, Masneuf-Pomarède I, Aigle M, & Dubourdieu D (2005) Molecular genetic study of introgression between *Saccharomyces bayanus* and *S. cerevisiae*. Yeast 22: 1099-1115.

Nguyen HV & Gaillardin C (2005) Evolutionary relationships between the former species *Saccharomyces uvarum* and the hybrids *Saccharomyces bayanus* and *Saccharomyces pastorianus*; reinstatement of *Saccharomyces uvarum* (Beijerinck) as a distinct species. FEMS Yeast Res 5: 471-483.

Nguyen HV, Lepingle A, & Gaillardin C (2000) Molecular typing demonstrates homogeneity of *Saccharomyces uvarum* strains and reveals the existence of hybrids between S. *uvarum* and S. *cerevisiae*, including the S. *bayanus* type strain CBS 380. Syst Appl Microbiol 23: 71-85.

Nguyen HV, Legras JL, Neuvéglise C, & Gaillardin C (2011) Deciphering the hybridisation history leading to the lager lineage based on the mosaic genomes of *Saccharomyces bayanus* strains NBRC1948 and CBS380. PLoS ONE 6: e25821.

Nosek J, Tomaska L, Bolotin-Fukuhara M, & Miyakawa I (2006) Mitochondrial chromosome structure: an insight from analysis of complete yeast genomes. FEMS Yeast Res 6: 356-370.

Novo M, Bigey F, Beyne E, Galeote V, Gavory F, *et al.,* (2009) Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. PNAS 106: 16333-16338.

Nunnari J, Marshall WF, Straight A, Murray A, Sedat JW, & Walter P (1997) Mitochondrial transmission during mating in *Saccharomyces cerevisiae* is determined by mitochondrial fusion and fission and the intramitochondrial segregation of mitochondrial DNA. Mol Biol Cell 8: 1233-1242.

Ohno S (1970) *Evolution by Gene Duplication*. Springer, New York.

Orlic S, Arroyo-López FN, Huic-Babic K, Lucilla I, Querol A, & Barrio E (2010) A comparative study of the wine fermentation performance of *Saccharomyces paradoxus* under different nitrogen concentrations and glucose/fructose ratios. J Appl Microbiol 108: 73-80.

Otto SP & Whitton J (2000) Polyploid incidence and evolution. Annu Rev Genet 34: 401-437.

Padidam M, Sawyer S, & Fauquet CM (1999) Possible emergence of new Geminiviruses by frequent recombination. Virology 265: 218-225.

Parikh V, Morgan M, Scott R, Clements L, & Butow R (1987) The mitochondrial genotype can influence nuclear gene expression in yeast. Science 235: 576-580.

Pavelka N, Rancati G, Zhu J, Bradford WD, Saraf A, *et al.,* (2010) Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. Nature 468: 321-325.

Payen C, Koszul R, Dujon B, & Fischer G (2008) Segmental duplications arise from Pol32-dependent repair of broken forks through two alternative replication-based mechanisms. PLoS Genet 4: e1000175.

Peña-Castillo L & Hughes TR (2007) Why are there still over 1000 uncharacterized yeast genes? Genetics 176: 7-14.

Perez-Ortin JE, Querol A, Puig S, & Barrio E (2002) Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains. Genome Res 12: 1533-1539.

Pérez-Través L, Lopes C A, Barrio E, & Querol A (2012) Evaluation of different genetic procedures for the generation of artificial hybrids in *Saccharomyces* genus for industrial purpose. Int J Food Microbiol 152: 102-111.

Pérez-Través L, Lopes C A, Querol A, & Barrio E (*In preparation*) On the analysis of *S. bayanus* taxon.

Peris D, Belloch C, Lopandic K, Álvarez-Pérez JM, Querol A, & Barrio E (2012a) The molecular characterization of new types of S. *cerevisiae* x S. *kudriavzevii* hybrid yeasts unveils a high genetic diversity. Yeast 29: 81-91.

Peris, D., Lopes, C.A., Belloch, C., Querol, A. & Barrio, E. (2012b) Comparative genomics among *Saccharomyces cerevisiae* x *Saccharomyces kudriavzevii* natural hybrid strains isolated from wine and beer reveals different origins. BMC Genomics online accession. doi:10.1186/1471-2164-13-407.

Peris D, Lopes CA, Arias A, & Barrio E (*In press*). Reconstruction of the evolutionary history of Saccharomyces cerevisiae x S. kudriavzevii hybrids based on multilocus sequence analysis. PLoS ONE. doi: PONE-D-12-10816.

Peris D, Arias A, & Barrio E (*In preparation*) *COX2* sequences of species from *Saccharomyces* genus show a recombination hotspot.

Peris D, Orlic S, Belloch C, & Barrio E (*In preparation*) Introgression in the mitochondrial *ORF1* gene as a footprint of ancestral hybridization events in the genus *Saccharomyces.*

Phaff HJ, Miller MW, & Mrak EM (1978) *The Life of Yeasts*. 2nd ed. (Press HU, ed.). Harvard University Press, Cambridge.

Piskur J & Langkjaer RB (2004) Yeast genome sequencing: the power of comparative genomics. Mol Microbiol 53: 381-389.

Pond SLK & Frost SDW (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21: 2531-2533.

Posada D (2003) Inferring evolutionary relationships: using Modeltest and PAUP* to select a model of nucleotide substitution. In *Current Protocols in Bioinformatics*, (Baxevanis A, Page R & Petsko G, eds.) pp. 6.5.1-6.5.14. Wiley, New York.

Posada D (2008) jModelTest: Phylogenetic model averaging. Mol Biol Evol 25: 1253-1256.

Posada D (2009) Bioinformatics for DNA Sequences Analysis. Methods in Molecular Biology 537: 93-112.

Presser A, Elowitz MB, Kellis M, & Kishony R (2008) The evolutionary dynamics of the *Saccharomyces cerevisiae* protein interaction network after duplication. PNAS 105: 950-954.

Pretorius IS (2000) Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking. Yeast 16: 675-729.

Priest FG & Stewart GG (2006) *Handbook of Brewing*. Taylor & Francis Group, New York.

Procházka E, Franko F, Poláková S, & Sulo P (2012) A complete sequence of *Saccharomyces paradoxus* mitochondrial genome that restores the respiration in *S. cerevisiae*. FEMS Yeast Res.

Pulvirenti A, Nguyen HV, Caggia C, Giudici P, Rainieri S, & Zambonelli C (2000) *Saccharomyces uvarum*, a proper species within *Saccharomyces sensu stricto*. FEMS Microbiol Lett 192: 191-196.

Querol A & Bond U (2009) The complex and dynamic genomes of industrial yeasts. FEMS Microbiol Lett 293: 1-10.

Querol A, Barrio E, & Ramón D (1992) A comparative study of different methods of yeast strain characterization. Syst Appl Microbiol 15: 439-446.

Querol A, Fernandez-Espinar MT, del Olmo Ml, & Barrio E (2003) Adaptive evolution of wine yeast. Int J Food Microbiol 86: 3-10.

Rainieri S, Kodama Y, Nakao Y, Pulvirenti A, & Giudici P (2008) The inheritance of mtDNA in lager brewing strains. FEMS Yeast Res 8: 586-596.

Rainieri S, Kodama Y, Kaneko Y, Mikata K, Nakao Y, & Ashikari T (2006) Pure and mixed genetic lines of *Saccharomyces bayanus* and *Saccharomyces pastorianus* and their contribution to the lager brewing strain genome. Appl Environ Microb 72: 3968-3974.

Ramsey J & Schemske DW (1998) Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annu Rev Ecol Syst 29: 467-501.

Redón M, Guillamón J, Mas A, & Rozès N (2011) Effect of growth temperature on yeast lipid composition and alcoholic fermentation at low temperature. Euro Food Res Technol 232: 517-527.

Redzepovic S, Orlic S, Sikora S, Majdak A, & Pretorius IS (2002) Identification and characterization of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* strains isolated from Croatian vineyards. Lett Appl Microbiol 35: 305-310.

Rementeria A, Rodriguez JA, Cadaval A, Amenabar R, Muguruza JR, *et al.,* (2003) Yeast associated with spontaneous fermentations of white wines from the "Txakoli de Bizkaia" region (Basque Country, North Spain). Int J Food Microbiol 86: 201-207.

Reuter M, Bell G, & Greig D (2007) Increased outbreeding in yeast in response to dispersal by an insect vector. Curr Biol 17: R81-R83.

Rieseberg L (1997) Hybrid origins of plant species. Annu Rev Ecol Syst 28: 359-389.

Rokas A, Williams BL, King N, & Carroll SB (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425: 798-804.

Salminen MO, Carr JK, Burke DS, McCutchan FE, & Garcia-Martinez J (1995) Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. AIDS Res Hum Retroviruses 11: 1423-1425.

Salvadó Z, Arroyo-López FN, Barrio E, Querol A, & Guillamón JM (2011) Quantifying the individual effects of ethanol and temperature on the fitness advantage of *Saccharomyces cerevisiae*. Food Microbiol 28: 1155-1161.

Sampaio JP & Gonçalves P (2008) Natural populations of *Saccharomyces kudriavzevii* in Portugal are associated with oak bark and are sympatric with S. *cerevisiae* and S. *paradoxus*. Appl Environ Microb 74: 2144-2152.

Samuel D (1996) Investigation of ancient Egyptian baking and brewing methods by correlative microscopy. Science 273: 488-490.

Sargueil B, Delahodde A, Hatat D, Tian GL, Lazowska J, & Jacg C (1991) A new specific DNA endonuclease activity in yeast mitochondria. Mol Gen Genet 225: 340-341.

Sato M, Kishimoto M, Watari J, & Takashio M (2002) Breeding of brewer's yeast by hybridization between a top-fermenting yeast *Saccharomyces cerevisiae* and a cryophilic yeast *Saccharomyces bayanus*. J Biosci Bioeng 93: 509-511.

Scannell DR, Zill OA, Rokas A, Payen C, Dunham MJ, Eisen MB, *et al.,* (2011) The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. G3: Genes, Genomes, Genetics 1: 11-25.

Schacherer J, Shapiro JA, Ruderfer DM, & Kruglyak L (2009) Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. Nature 458: 342-345.

Schacherer J, Tourrette Y, Souciet JL, Potier S, & de Montigny J (2004) Recovery of a function involving gene duplication by retroposition in *Saccharomyces cerevisiae*. Genome Res 14: 1291-1297.

Schäfer B, Kaulich K, & Wolf K (1998) Mosaic structure of the *COX2* gene in the petite negative yeast *Schizosaccharomyces* pombe: a group II intron is inserted at the same location as the otherwise unrelated group II introns in the mitochondria of higher plants. Gene 214: 101-112.

Schmidt HA, Strimmer K, Vingron M, & von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18: 502-504.

Schütz M & Gafner J (1994) Dynamics of the yeast strain population during spontaneous alcoholic fermentation determined by CHEF gel electrophoresis. Lett Appl Microbiol 19: 253-257.

Séraphin B, Simon M, & Faye G (1987) The mitochondrial reading frame RF3 is a functional gene in *Saccharomyces uvarum*. J Biol Chem 262: 10146-10153.

Séraphin B, Faye G, Hatat D, & Jacq C (1992) The yeast mitochondrial intron aI5[alpha]: associated endonuclease activity and in vivo mobility. Gene 113: 1-8.

Serra A, Strehaiano P, & Taillandier P (2005) Influence of temperature and pH on *Saccharomyces bayanus* var. *uvarum* growth; impact of a wine yeast interspecific hybridization on these parameters. Int J Food Microbiol 104: 257-265.

Shibata T & Ling F (2007) DNA recombination protein-dependent mechanism of homoplasmy and its proposed functions. Mitochondrion 7: 17-23.

Shimodaira H & Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol 16: 1114-1116.

Sicard D & Legras JL (2011) Bread, beer and wine: Yeast domestication in the *Saccharomyces sensu stricto* complex. C R Biol 334: 229-236.

Sickmann A, Reinders J, Wagner Y, Joppich C, Zahedi R, *et al.,* (2003) The proteome of *Saccharomyces cerevisiae* mitochondria. PNAS 100: 13207-13212.

Sipiczki M (2008) Interspecies hybridization and recombination in *Saccharomyces* wine yeasts. FEMS Yeast Res 8: 996-1007.

Skelly PJ & Clark-Walker GD (1991) Polymorphisms in tandemly repeated sequences of *Saccharomyces cerevisiae* mitochondrial DNA. J Mol Evol 32: 396-404.

Smith JM (1992) Analyzing the mosaic structure of genes. J Mol Evol 34: 126-129.

Sniegowski PD, Dombrowski PG, & Fingerman E (2002) *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. FEMS Yeast Res 1: 299-306.

Solieri L, Antúnez O, Pérez-Ortín JE, Barrio E, & Giudici P (2008) Mitochondrial inheritance and fermentative : oxidative balance in hybrids between *Saccharomyces cerevisiae* and *Saccharomyces uvarum*. Yeast 25: 485-500.

Souciet JL, Dujon B, Gaillardin C, Johnston M, Baret PV, *et al.,* (2009) Comparative genomics of protoploid Saccharomycetaceae. Genome Res 19: 1696-1709.

Staden R, Beal KF, & Bonfield JK (2000) The Staden Package, 1998. In *Methods in Molecular* Biology. (Cliften NJ, ed.) pp. 115-130.

Strimmer K & Rambaut A (2002) Inferring confidence sets of possibly misspecified gene trees. Proc Biol Sci 269: 137-142.

Suh SO, Blackwell M, Kurtzman CP, & Lachance MA (2006) Phylogenetics of *Saccharomycetales*, the ascomycete yeasts. Mycologia 98: 1006-1017.

Tamura K, Dudley J, Nei M, & Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596-1599.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, & Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol 28: 2731-2739.

Taxis C, Keller P, Kavagiou Z, Jensen LJ, Colombelli J, *et al.,* (2005) Spore number control and breeding in *Saccharomyces cerevisiae*. J Cell Biol 171: 627-640.

Taylor JW (1986) Fungal evolutionary biology and mitochondrial DNA. Exp Mycol 10: 259-269.

Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, *et al.,* (2000) Phylogenetic Species Recognition and Species Concepts in Fungi. Fungal Genet Biol 31: 21-32.

This P, Lacombe T, & Thomas MR (2006) Historical origins and genetic diversity of wine grapes. Trends Gen 22: 511-519.

Thompson JD, Higgins DG, & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through

sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680.

Torija MJ, Beltran G, Novo M, Poblet M, Guillamón JM, *et al.,* (2003) Effects of fermentation temperature and *Saccharomyces* species on the cell fatty acid composition and presence of volatile compounds in wine. Int J Food Microbiol 85: 127-136.

Torres EM, Williams BR, & Amon A (2008) Aneuploidy: cells losing their balance. Genetics 179: 737-746.

Torriani S, Zapparoli G, & Suzzi G (1999) Genetic and phenotypic diversity of *Saccharomyces sensu stricto* strains isolated from Amarone wine. A van Leeuw 75: 207-215.

Tronchoni J, Gamero A, Arroyo-Lopez FN, Barrio E, & Querol A (2009) Differences in the glucose and fructose consumption profiles in diverse *Saccharomyces* wine species and their hybrids during grape juice fermentation. Int J Food Microbiol 134: 237-243.

Tronchoni J, Rozás N, Querol A, & Guillamón JM (2012) Lipid composition of wine strains of *Saccharomyces kudriavzevii* and *Saccharomyces cerevisiae* grown at low temperature. Int J Food Microbiol 155: 191-198.

Tronchoni J, Combina M, Medina V, Pérez-Torrado R, Guillamón JM, & Querol A (*In preparation*) Transcriptome analysis of *S. kudriavzevii* and their hybrids during wine fermentation at low temperature.

Van Oevelen D, Spaepen M, Timmermans P, & Verachtert H (1977) Microbiological aspects of spontaneous wort fermentation in production of lambic and gueuze. J Ins Brew 83: 356-360.

Vaughan-Martini A & Kurtzman CP (1985) Deoxyribonucleic acid relatedness among species of the genus *Saccharomyces sensu stricto*. Int J Syst Bacteriol 35: 508-511.

Vaughan-Martini A & Martini A (2011) *Saccharomyces* Meyen ex. Rees (1870). 5th: 733-746.

Vaughan-Martini A & Martini A (1995) Facts, myths and legends on the prime industrial microorganism. J Ind Microbiol Biot 14: 514-522.

Vinogradov AE & Anatskaya OV (2009) Loss of protein interactions and regulatory divergence in yeast whole-genome duplicates. Genomics 93: 534-542.

Vivier MA & Pretorius IS (2002) Genetically tailored grapevines for the wine industry. Trends in Biotechnol 20: 472-478.

Wang SA & Bai FY (2008) *Saccharomyces arboricolus* sp. nov., a yeast species from tree bark. Int J Syst Evol Micr 58: 510-514.

Waterhouse AM, Procter JB, Martin DMA, Clamp M, & Barton GJ (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25: 1189-1191.

Wei W, McCusker JH, Hyman RW, Jones T, Ning Y, *et al.,* (2007) Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. PNAS 104: 12825-12830.

Weiller G, Schueller CME, & Schweyen RJ (1989) Putative target sites for mobile G+C rich clusters in yeast mitochondrial DNA: Single elements and tandem arrays. Mol Gen Genet 218: 272-283.

Wendel JF, Schnabel A, & Seelanan T (1995) Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). PNAS 92: 280-284.

Whitfield JB, Cameron SA, Huson DH, & Steel MA (2008) Filtered Z-Closure Supernetworks for extracting and visualizing recurrent signal from incongruent gene trees. Syst Biol 57: 939-947.

Wilkie D & Thomas DY (1973) Mitochondrial genetic analysis by zygote cell lineages in *Saccharomyces cerevisiae*. Genetics 73: 367-377.

Wolfe K (2004) Evolutionary Genomics: Yeasts accelerate beyond BLAST. Curr Biol 14: R392-R394.

Wolfe KH & Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387: 708-713.

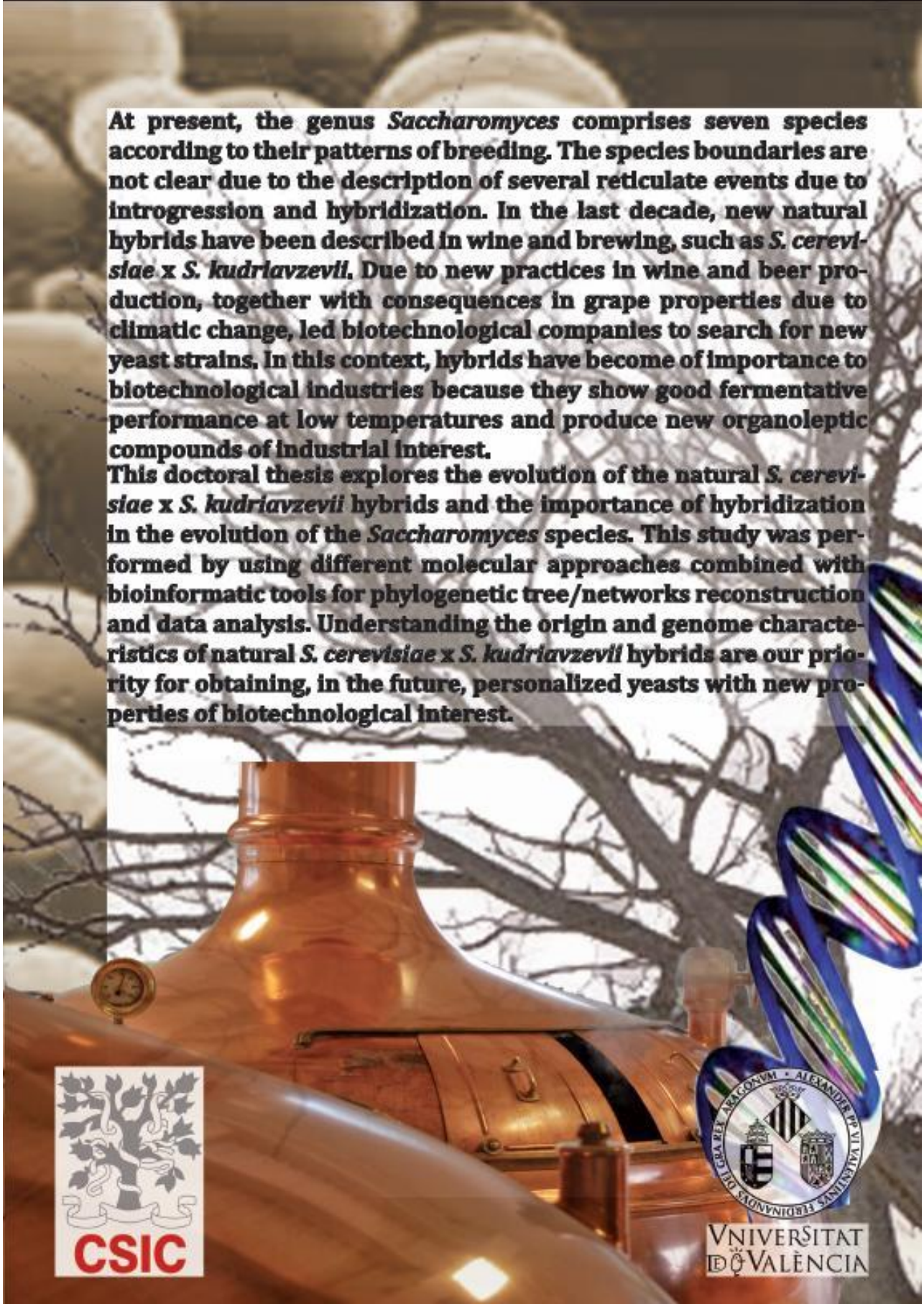
Yang X & Griffiths AJF (1993) Male transmission of linear plasmids and mitochondrial DNA in the fungus Neurospora. Genetics 134: 1055-1062.

Zambonelli C, Passarelli P, Rainieri S, Bertolini L, Giudici P, & Castellari L (1997) Technological properties and temperature response of interspecific *Saccharomyces* hybrids. J Sci Food Agric 74: 7-12.

Zamora F (2009) Biochemistry of alcoholic fermentation. In *Wine Chemistry and Biochemestry* (Moreno-Arribas MV, Polo MC, eds.) pp. 3-26. Springer, New York.

At present, the genus *Saccharomyces* comprises seven species according to their patterns of breeding. The species boundaries are not clear due to the description of several reticulate events due to introgression and hybridization. In the last decade, new natural hybrids have been described in wine and brewing, such as *S. cerevisiae* x *S. kudriavzevii*. Due to new practices in wine and beer production, together with consequences in grape properties due to climatic change, led biotechnological companies to search for new yeast strains. In this context, hybrids have become of importance to biotechnological industries because they show good fermentative performance at low temperatures and produce new organoleptic compounds of industrial interest.

This doctoral thesis explores the evolution of the natural *S. cerevisiae* x *S. kudriavzevii* hybrids and the importance of hybridization in the evolution of the *Saccharomyces* species. This study was performed by using different molecular approaches combined with bioinformatic tools for phylogenetic tree/networks reconstruction and data analysis. Understanding the origin and genome characteristics of natural *S. cerevisiae* x *S. kudriavzevii* hybrids are our priority for obtaining, in the future, personalized yeasts with new properties of biotechnological interest.



CSIC

VNIVERSITAT DE VALÈNCIA