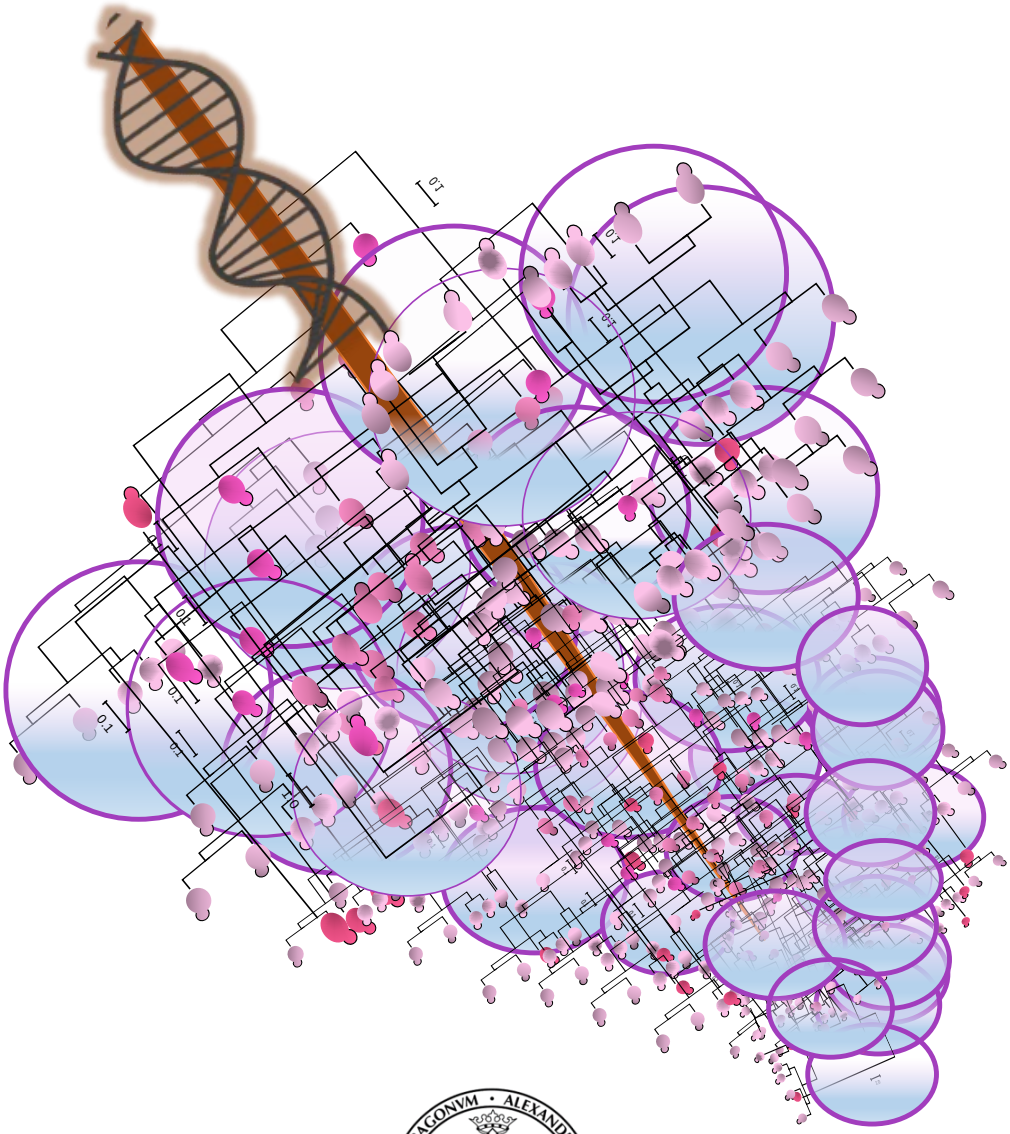
Genome diversity in *Torulaspora microellipsoides* and its contribution to the evolution of the *Saccharomyces* genus



This thesis is presented for the PhD degree of the University of Valencia

Thesis Director: Dr. Eladio Barrio Esparducer

Thesis Supervisor: Dr. Mercedes Costell Roselló

Adriana Mena Romero

Valencia, June 2018

VNIVERSITAT DE VALÈNCIA **Departament de Genètica**

El Dr. Eladio Barrio Esparducer, Profesor Titular del Departamento de Genética de la Universitat de València, adscrito como investigador al Departamento de Biotecnología del Instituto de Agroquímica y Tecnología de los Alimentos, CSIC.

CERTIFICA

Que el presente trabajo titulado **"Genome diversity in *Torulaspora microellipsoides* and its contribution to the evolution of the *Saccharomyces* genus"**, que presenta Dª Adriana Mena Romero para optar al grado de doctor en Biotecnología por la Universitat de València, ha sido realizado bajo su dirección en el Departamento de Genética de la Universidad de Valencia y en el Departamento de Biotecnología del Instituto de Agroquímica y Tecnología de los Alimentos, CSIC.

Y para que conste para los trámites de lectura y defensa de la tesis doctoral, en cumplimiento de la legislación vigente, firma el presente certificado en

Valencia a 15 de Junio de 2018

Fdo. Eladio Barrio Esparducer

Agradecimientos
(Acknowledgements)

Todo este trabajo no tendría sentido sin la gente que lleva años apoyándome para sacarlo adelante. Sin duda, el primer agradecimiento debe ir dirigido a mi familia. Gracias por vuestro apoyo no solo en el sentido económico que me ha permitido acceder a unos estudios universitarios, pero también por aguantarme en todos estos años en los que he antepuesto el trabajo a muchas otras cosas, en definitiva por vuestra paciencia y comprensión. Tanto ha sido el tiempo y tantos los cambios que parece que haya pasado toda una vida y sin embargo siempre habéis estado respaldándome en cada momento. El siguiente agradecimiento a nivel personal me gustaría que fuera para ELLAS, mis amigas del colegio, mis amigas de siempre. Con ellas he compartido la infancia, la adolescencia y tras esto, otra etapa en nuestras vidas: tener que asumir que nos distanciábamos para poder realizarnos cada una profesionalmente. Tenemos muy “buen ojo”, así que poco hemos solapado por las diferentes facultades de Valencia. Ha sido muy dura esta etapa para todas, cada una con profesiones diferentes o trabajos en diferentes ciudades que ha complicado mucho vernos y estar juntas. A pesar de todo ello, siempre hemos econtrado algún momento para vernos, reírnos, salir de fiesta, contarnos nuestras vidas y visitarnos en diferentes ciudades. Les agradezco enormemente que aún sigan ahí a pesar de haber sido absorbida por este mundo que poco tiempo te deja para disfrutar como toca de la gente. Gracias chicas por haber estado y por estar. Seguiremos celebrando todos nuestros logros. En casi todos estos años, otra persona me ha acompañado y me ha comprendido en todo momento. A ti, Juancho, te tengo que agradacer toda la paciencia y toda la ayuda que me has dado sin pensártelo dos veces.

Ya hace mucho que comenzó mi inmersión en este complicado pero fascinante mundo del conocimiento científico. Desde entonces, ya pude comprobar de primera mano cuantísimos grupos científicos desarrollaban trabajos relevantes

para la mejora en el campo de la salud, la alimentación o el conocimiento básico de los procesos biológicos. De esta forma, y en paralelo a la formación que recibía de mis clases universitarias, comencé a realizar prácticas en diferentes laboratorios. Siempre me permito aconsejar a la gente que empieza que dediquen un tiempo importante a la parte práctica, ya que esto nos ayuda a entender mejor todas las materias que estudiamos y verles la aplicabilidad y la importancia de las mismas, además de que necesitamos introducirnos poco a poco en el mundo laboral. Las primeras prácticas que realice en un laboratorio, fueron de forma voluntaria y durante un verano. Esto podría haber sido un arma de doble filo, pues haber tenido una mala experiencia podría haber sido determinante para dirigir mi futuro hacia otro lado que no fuera la ciencia (o quizás no). Sin embargo, la suerte me llevo a vivir una experiencia muy satisfactoria a nivel académico y personal. Por ello, siempre le agradeceré al Doctor Antonio Pineda Lucena y a todo su equipo de entonces, especialmente a Guillermo y Leticia, todos los conocimientos aportados en el campo de la Biología Molecular encaminada al descubrimiento de nuevos fármacos efectivos contra el desarrollo de metástasis por parte de las células cancerosas. Un par de años más tarde realicé otras prácticas dentro del campo de la genética de plantas en el IBMCP. En este nuevo grupo, dirigido por la Doctora Cristina Ferrandiz, me adentre un poco en el fenotipado y genotipado de plantas de *Arabidopsis*. Gracias a Vicente Balanza y especialmente a Marisa Navarrete Gómez por vuestra ayuda durante esta fase.

Una vez comencé mi etapa de Master universitario, aterricé en el grupo dentro del que finalmente realicé mi tesis doctoral, en ese momento situado en el Instituto Cavanilles. Aquí conocí a Noé, un postdoc jienense que me apoyó y codirigió mi trabajo fín de Máster. Gracias Noé por ayudarme a organizar el trabajo experimental y escrito, pero no solo te tengo que agradecer eso. Gracias por

querer transmitirme tu constancia y esfuerzo en el trabajo, cualidades que al final te han llevado a donde estás ahora, al puesto que te mereces como buen ciéntifico que eres.

Muchas más fueron las personas que empezaron a formar parte de mi círculo *pseudoprofesional* y finalmente de mi grupo de amistades. Todas ellas estaban en el mismo lugar, el Instituto Cavanilles. Primero conocí a mi compi, el Doctor David Peris, justo cuando el se iba de estancia a Manchester. Para desgracia de todos, regresó de la estancia, momento en el cual comenzamos a compartir laboratorio, momentos de descanso y terapia. Un día, decidió hacer una PCR y me sustrajo los dNTPs que tenía en mi congelador sin posterior reposición o aviso. Desde entonces supe que me iba a dar mucha guerra y así ha sido hasta el día de hoy (aún compartimos *lab*). Gracias amigo Dino, por todos tus consejos de sabio, por los burofaxes a través del escritorio y por las fiestas de Paiporta, claro. Hablando en serio, gracias por estar ahí en los momentos necesarios como amigo y como compañero.

Un tiempo en el Cavanilles lo compartimos ambos con Laura (laupetra para los amigos). Laura siempre ha estado para dar buenos consejos, a nivel laboral y personal, y siempre pasábamos buenos momentos los tres en el *lab*. Laura, gracias por tu cariño, sobre todo, y por la mano que echas siempre que lo necesitamos, ya sabes que eres una joyita y vales un montón.

Gracias a todos los cavanilleros por los divertidos momentos durante la hora de la comida en el pasillo, por las salidas por el Carmen, por los consejos profesionales, por los debates científicos, económicos, políticos y de cualquier índole que se nos ocurriese. En especial, quiero agradecer todos estos momentos a: Rafeta (o Patiñet o cabeza de viejo, cuerpo de joven y sus aventuras) por su simpatía, a Quelo (Quelífero) por su ayuda también con temas del doctorado ahora

y antes por ser un buen compañero y amigo; a Pilar Domingo por ser tan buena compañera de *lab* cuando lo compartíamos con Rafa Sanjuán y por el viajecito a Málaga con las chicas; a Marianoel también por todo el tiempo vivido compartiendo *lab* y por otros momentos fuera de él en los que nos lo pasábamos genial; gracias a Carmen y Elisa por ser tan buenas consejeras en todo, tan simpáticas, tan amables…recuerdo buenos momentos con vosotras también, el viaje a casa de Carmen y las cervecitas por Valencia; muchas gracias también a Vanesa y a Joanot por nuestras quedadas y por hacerme ver que sois personas con las que podía contar y con las que puedo contar a día de hoy, gracias por ser tan buenos amigos, os quiero un montón.

Ahí, en el cavanilles, fue donde empezó mi etapa predoctoral en 2010. En primer lugar he de agradecer a mi director de tesis, el doctor Eladio Barrio Esparducer, que me diera la oportunidad de realizar la tesis doctoral con él, escogiéndome entre el resto de posibles candidatos. Gracias también a la doctora Amparo Querol Simón, por darme su apoyo y ayuda de forma desinteresada en muchos momentos durante esta etapa de la tesis, tanto a nivel laboral como a nivel personal.

Fue en el ecuador del doctorado cuando nos movimos al IATA, el Instituto de Agroquímica y Tecnología de alimentos del CSIC, y aquí es donde he finalizado la escritura de esta tesis. Los comienzos fueron muy duros, pues traían consigo una nueva adaptación al ambiente. Afortunadamente, muchas fueron las personas que me echaron una mano en este momento. Desde el principio tuve la grata compañía de “los veteranos de guerra”. Entre ellos se encontraban María López, Isaac Córcoles, Alicia Gutierrez, Amparo Gamero y Jordi Tronchoni (señores doctores). Gracias a todos por los buenos momentos, consejos, conversaciones, cenitas y todos los momentos en los que nos hemos echado unas risas. También recuerdo

con mucho cariño a mis compañeras de tesis del 307, a las doctoras Clarita y Eli. Gracias a vosotras por la complicidad, por estar en los momentos más estresantes y también en los más emocionantes como en nuestro viaje a Chicago y el congreso en Madison. No me olvidaré nunca de lo bien que nos lo pasámos y lo chulo que fue viajar a los U.S.A. con vosotras. Gracias también a mis otros compis del 307, en especial a David Lázaro por toda la ayuda recibida. Siempre has estado ahí cuando he necesitado cualquier cosa del *lab*. A Roberto Pérez, por aprovechar un momento para preguntarme qué tal me iba todo y sus ánimos. Gracias a los demás compis con los que he compartido "el 309": a Nerea por estar conmigo demasiados días desde las 9:00h hasta las 11:00h, como si fuera el horario de trabajo de una persona normal; a Roseta por aportar simpatía y consejos y a Carlos por aguantarnos a Anto y a mi y por ser tan buen compi.

Sin duda, tengo que dedicar un parrafazo entero a ELLOS, "LA SECTA". Hemos solapado un buen grupo de personas en periodo predoctoral. Todos ellos son diferentes pero destacan por su calidad como personas y por estar ahí para entenderte en todo. En especial, tengo que destacar la amistad que me une a Antonia María Romero Cuadrado (ya doctora cuando se lea esta tesis). Con ella he pasado muchos momentos de risas y *ll-antos*, pero quedémonos con las risas. Desde los momentos vividos con la Ramona hasta los ataques de risa echadas por los suelos, ha sido una maravilla compartirlo contigo. Y ahora, aquí estamos, intentando defender la tesis las dos en el mismo mes, todas las cosas juntas, pues claro. Gracias por todos esos momentos y por los que vendrán.

El macha-alfísmo ha sido otro de los MUST durante este tiempo, pues, la sincronización de la cohorte es importante. Nosotras lo hemos sabido hacer, eso e *ir cargadas como una mula*, *sudar hasta la cromatografía* en las clases de spinning o "no saber" caminar en la montaña como las cabras. Sí, eres tú, Cecilia. Ya sabes

que te tengo que agradecer consejos, cervezas, pitis y todo lo que nos ha llevado a ser unos verdaderos machos, jajaja. En serio, muchas gracias por estar ahí antes y hoy…bueno, hasta hace muy poquito seguro, jiji. ¡¡Os veo a las dos en "las suecias"!!

Indudablemente, he conocido a más personas cuyos valores y actitudes me han aportado mucho en muchos momentos. Sois sobre todo Aurora, Javi y Jiri. Gracias por eso mismo, por vuestra simpatía que os caracteriza, por ser buenos compañeros, buenos amigos, por las fiestas, viajes, cursos, por ser SECTA SECTA. Gracias a los sectarios bioinformáticos porque sois geniales. Con vosotros me he pasado horas hablando de trabajo (somos bastante "frikis" en ese sentido), en especial con Miguel y luego con Laura. Pero también nos hemos reído mucho juntos, nos hemos cruzado en el tranvía con unas y hemos compartido vestiditos con otras. Mil gracias Miguel, Laura y María, sois increíbles y muy buenos compañeros, unos de categoría.

En este grupo tenemos muchos imprescindibles y que también han contribuido a los grandes momentos, eventos y otros *quereseres* sectarios y por tanto gracias por estar ahí: a Lucín por estar siempre dispuesta a tomar una cerve y pasar un rato agradable dentro del grupo de jinetes; a Alba por momentos vividos entre las dos; a Fani porque contigo *el tután* está asegurado y la entrada a las discos también, jaja… gracias por los momentos divertidos; a Sara por su bondad y por su ayuda siempre que la necesitas, a Rick por su sabiduría (siempre necesaria), por sus momentos de humor y también por su ayuda cuando la he necesitado, a Ying por habernos entendido tan bien a todos desde el principio y por las risas que nos echamos cuando salimos de fiesta y a Romain por ser un tío muy simpático y por presentarnos a Laura, jaja. La mención de honor sería para Josepín, nuestro sectario fugaz que aportó mucho a la secta.

No puedo olvidarme de aquellos que llegaron de fuera, estuvieron unos meses y conquistaron a “LA SECTA”. Hablo de Bene, Migue y Faber. ¡¡Qué rápido os adaptásteis y cuánto os gusto!! jaja. Nos habéis aportado mucho todos, pero en especial quiero agradecerle a Bene su amistad aquí, el kínder bueno que me regaló (no se me olvida) y su amistad en cualquier sitio que nos vemos, sobre todo en Italia (gracias a Marco y a Rita por ser tan majos también).

También a los recién llegados Dolo y Seba. Me ha encantado conoceros porque sois majísimos y siempre os apuntáis a todos los planes que hacemos. Espero que os vaya muy bien durante estos años venideros.

En un momento de ese periodo, realicé una estancia doctoral en Montpellier, Francia, y de esta experiencia aprendí muchísimo, no solo de ciencia, también a sobrevivir fuera de la zona de *confort* y a valorar tanto el apoyo de la gente de allí como todo lo que tenía aquí. Pero ya sabemos que estar fuera siempre es duro y siempre se tiene que agradecer la ayuda desinteresada de la gente. En mi caso tuve mucha suerte de coincidir con mi colega de *lab* Souhir Marsit, siempre trabajando tanto y a la vez siendo un apoyo constante a nivel psicológico y laboral. Gracias por haber estado ahí, por acompañarme y entenderme en los momentos duros y por haberme llevado a otros sitios fuera del *lab* para pasárnoslo bien y olvidarnos un poco de nuestros momentos de agobio. También quiero acordarme del resto de gente que componía el laboratorio en ese momento, empezando por la cabeza de grupo, la doctora Silvie Dequin y el resto de personal técnico, post-docs y pre-docs: Pascale Brial, Valentin Tilloy, Claire Brice, Christian Picou, Erick Casalta, Anna Lisa Coi, Stéphanie Rollero, Pierre De Lobel y muchos más.

De fuera del *lab*, tengo que dar las gracias a Pierre y a Hèléne por invitarme a sus fiestas y por ser tan amables con los españoles que estábamos por allí. Gracias a los chicos de la *Hispanothèque* por su trabajo de integración. A Sebastién

por enseñarnos la ciudad de Marsella a Cristina, a Tamara y a mi. Gracias a todos los españoles en Montpellier, gente genial y que empatiza siempre con cada español recién llegado para enfrentarse al mundo laboral. Gracias a todo el grupo: desde Támara a Virginia hasta Luis, Óscar, Álvaro, Vicente, Fran, Flamen, Eric, Pedro, Dimitri, Carmen, Sofía, Ana, Cristina, Paola, Rebeca, María, Rubén Enrique, Patri, Itziar, etc. Gracias a mi primer compi de piso en Montpellier, Piotr Borowski por las charlas con la casera y mis otros dos compis, Marina y Alexis por ser tan majos y por las noches de *reality* en el sofá partiéndonos de risa.

Durante los últimos años de escritura de la tesis he tenido la suerte de poder trabajar en el grupo de Genómica Funcional de Levaduras (GFL) dirigido por los doctores José Enrique Pérez-Ortín y Paula Alepuz. A los dos tengo mucho que agradecerles. Primero, tengo que dar las gracias a José Enrique por confiar en mí y por contratarme para llevar a cabo todas sus ideas. Gracias por haberme transmitido tantos conocimientos, por tu preocupación constante de sacar el trabajo adelante, por darme la confianza de saber que podía contar con tu ayuda y experiencia para cualquier problema que surgiese en el trabajo y por tu exigencia conmigo y contigo mismo para poder sacar el trabajo adelante. En fín, gracias por ser un buen mentor. Gracias a ti, Paula por transmitirme, a mí y al resto de personas del *lab*, tu constante esfuerzo y dedicación al trabajo y por aportar siempre ideas tan brillantes que hacen que siempre estemos motivados. Gracias también por tu implicación en todo lo que se hace en el *lab* y por reivindicar el trabajo que hacemos las mujeres ciéntificas con todo lo que nos implica por el hecho de ser mujeres y tener que compaginarlo con otras etapas de nuestra vida. A todos los que componen o en algún momento han formado parte del equipo del *lab* y que he conocido personalmente, también os tengo mucho que agradecer. Desde Daniel, quien me enseñó prácticamente todo el funcionamiento del *lab*,

conocimientos acerca de protocolos, programas, bioinformática, etc. Gracias por implicarte tanto y como ya te dije en tu despedida, gracias por dedicarme tanto tiempo, a pesar de estar en los días previos a la lectura de tu tesis. ¡¡Siempre he recibido un montón de ayuda por tu parte y eso no se olvida!! A *nuestro cachorro* Jordi, que ahora será un *crack* de la bioinformática. Gracias por tu simpatía, tus locuras y por aportar al *lab* esa parte de juventud que tanta falta hace de vez en cuando. Gracias también a Tian, por sus risas, los momentos de diversión, la complicidad y por ser un ejemplo a nivel profesional. Gracias a Fani, por demostrar ser tan buena en el trabajo como en la vida y escucharnos siempre que lo necesitábamos. Gracias a Esther y Marta por las risas y los momentos divertidos en el *lab* y por las comidas del *Ademúz*. Gracias a mis compis de ahora, Toni y María, porque sois majísimos, sois unos compis buenísimos, nos reímos mucho y aprendemos mucho juntos. Gracias por toda la ayuda que me habéis dado en estos últimos meses, especialmente tu María, soportando mis cosas en la bancada de al lado.

En lo que concierne a los últimos días previos al depósito de esta tesis, he de agradecer enormemente a las doctoras Carmela Belloch y Christina Toft, su gran ayuda a la hora de realizar correcciones y hacer buenísimas sugerencias para que este trabajo tuviera tanto el fondo y la forma que se merecen. Mil gracias a vosotras dos y al resto de miembros del tribunal por vuestras aportaciones y valoraciones a pesar de tener mil compromisos que atender.

Gracias a ti, que eres mi presente y espero que seas mi futuro. Gracias por ser mi reflejo en el espejo, mi parte simétrica y en difinitiva mi (suena cursi) media naranja. Aunque estas loquito y es muy difícil aguantarte, me hace muy feliz que sea así y espero que continúe siéndolo. Gracias por haberme apoyado durante los últimos meses tanto en la parte profesional como en la parte personal. Aún nos quedan muchas cosas buenas por vivir juntos y muchas cosas malas que superar juntos. Nos vemos en Palermo.

El final de estos agradecimientos se escribe con W…

Thesis contents

Figures and Tables index

Figures and Tables index:

Introduction:

Chapter 1: Active fructose symporter and ammonia exporter acquisition by introgression events in the *Saccharomyces* genus:

Chapter 2: *Torulaspora microellipsoides* as a genus that contributes gene novelties in domesticated *Saccharomyces* strains:

List of supplemental figures and strains tables:

Glossary of abbreviations and definitions

Abbreviations and gene descriptions:

HGT: Horizontal Gene Transfer

HML/HMR: Hidden MAT Left/Hidden MAT Right

Ho: Homothallic (Homothallic switching endonuclease)

MAT: Mating Type

ML: Maximum Likelihood

ORF: Open Reading Frame

WGD: Whole Genome Duplication

***ATO3*:** Ammonia Transport Outward gene. Plasma membrane protein, putative ammonium transporter; regulation pattern suggests a possible role in export of ammonia from the cell; phosphorylated in mitochondria; member of the TC 9.B.33 YaaH family of putative transporters

***FSY1*:** Fructose Symporter 1 gene. Plasma membrane fructose/H(+) symporter that shows high affinity for fructose and is present in some *Saccharomyces* strains.

HXT: Hexose Transporter genes. Gene family that encodes glucose sensors and hexoses (mainly glucose and fructose) membrane transporters.

***HO*:** Site-specific endonuclease; required for gene conversion at the MAT locus (homothallic switching) through the generation of a ds DNA break; expression restricted to mother cells in late G1 as controlled by Swi4p-Swi6p, Swi5p, and Ash1p

Definitions:

Allotetraploidy: the status of a cell or an organism having four full sets of chromosome complements, two derived from one diploid species, the other two from another, different, diploid species.

Assembly: refers to aligning and merging fragments from a longer DNA sequence in order to reconstruct the original sequence. This is needed as DNA sequencing technology cannot read whole genomes in one go, but rather reads small pieces of between 20 and 30000 bases, depending on the technology used.

Biological species concept: defines species as a group of organisms that do or potentially could interbreed with one another indefinitely to the exclusion of other such groups.

BLAST (Basic Local Alignment Search Tool): software suite with programs for searching a sequence against a database in order to find similar sequences, freeware developed at the NCBI.

Bootstrap analysis: in phylogenetic analysis, this technique involves creating replicate data sets of the same size as the original alignment by randomly resampling alignment columns with replacement from the original alignment and reconstruction phylogenetic trees for each. The proportion of each clade among all the bootstrap replicates can be considered as measure of robustness of the monophyly of the taxa subset.

Clean linages (clean populations): lineages that exhibit the same phylogenetic relationship across their entire genomes.

Ecological species concept: defines species as a group of organisms that have a distinct ecological niche that distinguishes them from other species.

Horizontal Gene Transfer: the movement of genes from one species to another without successful mating. In contrast to bacteria, this process seems to be fairly rare in yeast, except those undergoing domestication.

Illumina sequencing: method of sequencing based on reversible dye-terminators that enable the identification of single bases as they are introduced into DNA strands. It can also be used for whole-genome and region sequencing, transcriptome analysis, metagenomics, small RNA discovery, methylation profiling, and genome-wide protein-nucleic acid interaction analysis.

Introgression: gene flow from one species to another through a process of successful mating and backcrossing.

Mapping: method that compares the DNA reads of a sequenced sample to one reference sequence to find the corresponding part of that sequence for each read in our sequencing data.

ML: Maximum likelihood. A principle of statistical inference which is essentially generalization of least-squares to non-normal data, and can be shown to lead to optimal estimators, at least for large sample size. Moreover, it is fully automatic

once a model is specified, and allows computing confidence bands by means of the so-called Fisher information.

Mosaic linages (mosaic populations): lineages that exhibit a different phylogenetic relationship from other studied lineages.

NCBI: National Center for Biotechnological Information. A web page that houses a series of databases relevant to biotechnology and biomedicine and is an important resource for bioinformatics tools and services. Major databases include GenBank for DNA sequences and PubMed, a bibliographic database for the biomedical literature.

NJ: Neighbor-joining. A heuristic method for estimating the minimum evolution tree. The principle is to find pairs of operational taxonomic units (OTUs) that minimise the total branch length at each stage of clustering of OTUs starting with a star-like tree.

Orthology: homology between two genes in different species that coalesce to a common ancestral gene without gene duplication or horizontal transmission.

Ohnolog/s: a duplicate gene or paralog pair created by whole genome duplication (after Susumu Ohno).

***PacBio* sequencing:** method based on the Single molecule real time (SMRT) sequencing commercialized by *PacBio* in 2011. A single DNA polymerase enzyme is affixed at the bottom of a nanophotonic confinement structure called ZMW (zero-mode waveguide) with a single molecule of DNA as a template. The ZMW is a structure that creates an illuminated observation volume that is small enough to

observe only a single nucleotide of DNA being incorporated by DNA polymerase. Each of the four DNA bases is attached to one of four different fluorescent dyes. When a nucleotide is incorporated by the DNA polymerase, the fluorescent tag is cleaved off and diffuses out of the observation area of the ZMW where its fluorescence is no longer observable. A detector detects the fluorescent signal of the nucleotide incorporation, and the base call is made according to the corresponding fluorescence of the dye.

Paralogy: homology between two non-allelic genes of the same genome, derived by duplication from a common ancestor.

Phylogenetic species concept: defines species as the smallest group of organisms whose genes (all or nearly all) share a common ancestor that excludes all other species.

Reads: short (method of sequencing size dependent) DNA sequences obtained from sequencing that often come in (or can be converted into) a file format called FASTQ. It is a *plain text* format, containing the sequence and quality scores for every read, where each single read normally occupies four consecutive lines.

***Saccharomyces* 'sensu lato':** *Saccharomyces* in the broad sense; an informal and no longer valid term that historically applied to yeasts from several genera, many of which are now known to have closer non-*Saccharomyces* relatives.

***Saccharomyces* 'sensu stricto':** *Saccharomyces* in the strict sense. There are currently eight natural species in the *Saccharomyces* genus.

Sanger sequencing: method of DNA sequencing first commercialized by Applied Biosystems, based on the selective incorporation of chain-terminating dideoxynucleotides by DNA polymerase during in vitro DNA replication.

sp: Abbreviation referred to a non-identified species belonging to an identified genus.

spp: Abbreviation referred to all the species included inside a genus.

Subtelomere: Genomic fragment whose proximal boundary (towards the centromere) is located in basis of the sudden loss of synteny conservation and whose distal boundary (towards the telomere) is demarcated by the telomere-associated core X and Y' elements.

Synteny: the common presence of genes along a given chromosome or chromosomal segment. The notion generally also implies the order of those genes. Hence, conservation of synteny indicates the conservation of the order of homologous genes between two chromosomes or between chromosomal segments of different species.

`Unware domestication': the process of evolution applied to some organisms (e.g. *Saccharomyces cerevisiae*) that underwent highly specialized on the utilization of the different environments or ecological niches provided by human activity. This evolutionary process is reflected in the genetic characteristics of those organisms.

Brief abstract in Spanish

El estudio de las características moleculares de levaduras de la familia ascomicetes ha sido de gran importancia en los últimos años por su aplicación en diferentes procesos industriales, sobre todo aquellos relacionados con la alimentación y por su uso como organismos modelo de célula eucariota, contribuyendo con multitud de estudios que aportan conocimiento básico acerca del funcionamiento de diferentes procesos a nivel celular. Dentro de este grupo de levaduras, ha sido de gran relevancia la comprensión de los procesos que llevan a cabo las especies del género *Saccharomyces*, directamente implicadas en procesos fermentativos de gran valor en la industria alimentaria. Así la levadura *Saccharomyces cerevisiae* es la predominante en las fermentaciones vínicas y en la producción de pan y las levaduras *Saccharomyces pastorianus* y *Saccharomyces bayanus* poseen características diferenciales que las hacen apropiadas en la fermentación que da lugar a la producción de cerveza. Otras especies de levadura son responsables de las propiedades organolépticas finales de los quesos, como *Kluyveromyces lactis* o el depurado final de los granos de café, que es llevado a cabo por especies de *Candida*, *Saccharomyces* o *Kluyveromyces*. También se han realizado numerosos estudios sobre cómo estas levaduras han llegado a especializarse para dar las características que hoy en día se valoran y pretenden mejorar en estas bebidas y alimentos. En este proceso, ha sido clave la intervención del hombre ya desde la época Mesopotámica, en la que se encontró la jarra de vino más antigua hasta el día de hoy (5400-5000 a. de C.). Se habla de que se llevó a cabo una selección *inconsciente* de las levaduras que mejoraban las propiedades finales, por ejemplo seleccionando los vinos de mejor calidad y desechando los malos. Esta selección se ha conocido como domesticación y su estudio ha sido de gran importancia para saber cómo el genoma de una especie que procedería de un ambiente natural, ha podido ser optimizado a lo largo del tiempo para adaptarse a un ambiente fermentativo y competir con el resto de

especies presentes. La huella del proceso, conocido como “domesticación inconsciente”, se ha visto reflejada en diferentes estudios poblacionales que han demostrado que las cepas de levaduras especializadas en un tipo de fermentación específica están más estrechamente relacionadas entre sí que con las que proceden de una misma área geográfica (Liti G. et al 2011 and Sicard D and Legras JL, 2011).

Según sean las condiciones ambientales, en el ciclo de vida de las levaduras ascomicetes se puede dar reproducción asexual o sexual. Cuando el medio en el que se encuentran es rico en nutrientes y sin ningún tipo de estrés, el crecimiento vegetativo se lleva a cabo por reproducción asexual, de forma que se produce una división asimétrica en la que se genera una célula hija por gemación a partir de una célula progenitora o madre. En condiciones desfavorables, las células haploides conjugan con otra célula haploide de sexo contrario para formar una diploide. Para que se lleve a cabo la conjugación, las células haploides presentan diferentes tipos sexuales que vienen determinados por el gen *MAT* para el que existen dos tipos sexuales o idiomorfos llamados *MATa* y *MATα*. A partir del estado diploide (a/α), se podrá dar la reproducción sexual mediante una división meiótica que produzca esporas resistentes (haploides) que germinarán una vez las condiciones sean favorables.

Según sea el estado predominante, las levaduras se caracterizan por tener un ciclo de vida haplo-biontico o diplo-biontico. En las especies de los géneros *Torulaspora*, *Zygosaccharomyces* o *Zygotorulaspora* predomina el estado haploide mientras que en especies del género *Saccharomyces*, el ciclo de vida predominante es el diplo-bionte o estado diploide.

A partir del estado haploide, dos células de distinto sexo pueden conjugar de varias formas: si el cruce se produce entre dos células procedentes de productos meióticos de diferentes células haploides se denomina amfimixis o cruzamiento externo. Por otro lado, el cruce se puede producir dentro de la ascospora entre dos células de diferente sexo, lo que se conoce como conjugación intra-tétrada. Otra forma de conjugación es la autodiploidización, que implica la fusión de la célula madre con la célula hija, lo que genera un genoma homocigoto excepto para el locus *MAT*. Las levaduras capaces de cambiar el tipo sexual se llaman homotálicas (término referido a la posibilidad de que un individuo presente ambos sexos). Este mecanismo evolucionó dentro de los Saccharomycetes en dos pasos: primero se adquirieron los *cassetes* silenciosos en un ancestro del género *Kluyveromyces* y más tarde se adquirió la endonuclasa Ho antes de la divergencia del género *Zygosaccharomyces*. El *cassete* y la endonucleasa permiten el cambio de sexo por recombinación homóloga.

Un evento clave en la filogenia de la familia de levaduras Saccharomycetaceae fue el que marcó la diferencia entre las levaduras de linajes anteriores a dicho evento y las descendientes: la duplicación del genoma en un ancestro común a varios géneros de la filogenia, entre ellos el género *Saccharomyces*. Hoy en día se propone que esta duplicación genómica se produjo por una hibridación inter-específica que dio como consecuencia la duplicación en la dotación cromosómica. Tras este proceso de hibridación, una estabilización del genoma sería necesaria para la propagación sexual de esta nueva formación, por lo que dos hipótesis se proponen para ello: la primera es que la fusión se produjese entre dos células diploides dando lugar a un tetraploide fértil. Una segunda hipótesis propondría la formación de un híbrido entre dos células haploides, cuya estabilización genómica vendría dada por una subsiguiente autodiploidización para finalmente generar una especie tetraploide.

Este aumento masivo en el número de copias del conjunto de genes del genoma tuvo varias consecuenias inmeditas, que se especula fueron decisivas para la dominancia de estas especies en ambientes fermentativos. En primer lugar, se produjo una pérdida masiva de genes redundantes, con lo que el genoma se redujo drásticamente. Sin embargo, otros muchos genes se mantuvieron duplicados, haciendo que su concentración celular aumentase. Se especula que en este momento fue cuando las levaduras descendientes de esta duplicación adquirieron su metabolismo fermentativo característico. Este metabolismo está basado en el denominado efecto *Cabtree*, que básicamente consiste en que el flujo metabólico esta sesgado hacia el metabolismo fermentativo en condiciones de presencia de oxígeno (aeróbicas).

Se hipotetiza que este comportamiento fermentativo apareció a la vez que surgió el evento de fructificación en las angiospermas, cuyos frutos eran muy ricos en azúcares. A su vez, la duplicación genómica trajo consigo la duplicación y mantenimiento de los genes implicados en el transporte de azúcares al interior celular y su procesamiento para ser incorporados a la ruta metabólica. Este aumento en el flujo de glucosa y fructosa (azúcares mayoritarios en la fermentación) por el propio aumento en el número de permeasas en la membrana celular, permitió a estos organismos crecer más rápidamente que a sus competidores. A pesar de que la fermentación es una estrategia metabólica mucho menos energética que la respiración, ya que produce menos cantidad de ATP (adenosina trifosfato), el aumento en biomasa y la elevada concentración final de etanol a consecuencia de la fermentación alcohólica, hizo que estos organismos, mayoritariamente *S. cerevisiae*, compitiese de manera tan eficiente con el resto de microorganismos presentes en ese ambiente que acabó imponiéndose en las fermentaciones.

En esta tesis tenemos como objetivo general, el estudio de la evolución y los mecanismos implicados en la adquisición de genes relevantes para la preponderancia de estos microorganismos en las fermentaciones. Especialmente, ha sido de nuestro interés el estudio de los transportadores de azúcares, glucosa y fructosa, dada su importancia en la eficiencia de las fermentaciones.

En especial, hemos puesto más atención a un transportador de fructosa de alta afinidad que es específico para el transporte de fructosa y trabaja de forma activa, incorporando la fructosa al interior celular mediante el co-transporte de H^+. Partiendo de trabajos anteriores, ya se conocía que este transportador, cuyo gen que lo codifica fue designado como *FSY1*, se encontraba en las especies *S. eubayanus*, *S. uvarum* y *S. pastorianus*, pero no se había descrito en ninguna especie más del género *Saccharomyces*. Varios años después, se descubrió la presencia de este gen en la cepa vínica EC1118 perteneciente a la especie *S. cerevisiae*. En ese trabajo, se especuló con la idea de que este gen podría haber sido adquirido por transferencia génica horizontal junto a un grupo de genes adyacentes, que en conjunto formaban una región subtelomética de genes no conocidos antes en *S. cerevisiae*. Además de esta región, llamada *Región C*, otras 2 regiones mas conteniendo genes de nueva adquisición fueron descubiertas en ese mismo genoma y en conjunto, las 3 regiones presentaban genes relacionados con el metabolismo fermentativo, cuya adquisición se podría relacionar con una ventaja evolutiva para esta cepa en estos ambientes.

En concreto, el estudio de este gen, que se encarga de transportar fructosa cuando este azúcar se encuentra en concentraciones bajas en el medio, resulta de gran importancia a nivel industrial. Los mostos a partir de los cuales de lleva a cabo la fermentación por parte de las levaduras, poseen unas proporciones equimolares de glucosa y fructosa. Sin embargo, parece ser que *S. cerevisiae* tiene una

preferencia por consumir la glucosa antes que la fructosa, de forma que esta última queda de forma residual al final de las fermentaciones mientras que la glucosa ha sido completamente consumida. Esta situación provoca que los vinos que se producen resulten excesivamente dulces, ya que la fructosa aporta mas sensación dulce que la glucosa, y las cantidades de etanol finalmente producidas no son tan altas, de forma que los vinos son propensos a contaminación por bacterias. Este hecho estimuló la búsqueda de soluciones a nivel industrial para evitar este efecto no deseado de la fructosa residual. Por ello, fue de importancia la identificación de levaduras con un llamado carácter "fructofílico", que ayudase a evitar estas paradas en fermentación y de aquí uno de nuestros objetivos en el estudio de este transportador.

En el trabajo de Novo et al. 2009, se propuso que esta región de genes, llamada Región C y en la que se encuentró el gen *FSY1*, fue adquirida en esta cepa de *S. cerevisiae* por transferencia horizontal. A través de un análisis de la sintenía y homología de los genes de la Región C, los autores especularon que estos genes podrían proceder de una especie desconocida procedente del género *Saccharomyces*.

Con el objetivo de averiguar cual podría haber sido la especie dadora de este gen a *S. cerevisiae*, llevamos a cabo una búsqueda de dicho transportador en diferentes cepas y especies de la familia Saccharomycetaceae. Para ello realizamos una primera búsqueda del gen empleando las secuencias disponibles en las bases de datos de *NCBI* y *Sanger*, de los genomas secuenciados hasta ese momento. Paralelamente diseñamos unos oligonucleótidos generales para llevar a cabo la amplificación por *PCR* de este gen, en el mayor número de especies posible. Una vez obtenidos los amplificados, obtuvimos las secuencias de los mismos mediante amplificación *Sanger*. Durante la visualización de los cromatogramas procedentes

de la secuenciación, observamos que para la cepa CLIB 830T de la especie *Torulaspora microellipsoides*, obtuvimos una representación de dobles picos, lo que sugeria que en esta cepa existía una segunda secuencia que estábamos amplificando accidentalmente. Finalmente, conseguimos obtener esa segunda secuencia y observamos que presentaba alrededor de un 70% de similitud de secuencia, por lo que pensábamos que podría tratarse de genes parálogos y no de copias alélicas.

Nuestro primer resultado reflejaba como se distribuía este gen en la filogenia de la familia Saccharomycetaceae. Observamos que mientras que la mayoría de especies pre-duplicación genómica presentaban este gen en su genoma, tan solo uno de los géneros post-duplicación genómica presentaban este gen. Concretamente, solo se encontraba en las especies en las que ya se había visto: el grupo formado por *S. eubayanus*, *S. uvarum* y *S. pastorianus* en el que todas las cepas analizadas presentaban el gen (100%) y en algunas cepas de la especie *S. cerevisiae*, para las cuales se observó que el gen se presentaba en cepas mayoritariamente vínicas.

Un primer análisis filogenético, reveló que esta especie, *T. microellipsoides*, presentaba una posición filogenética no esperada con respecto a la filogenia obtenida para la familia Saccharomycetaceae en la que como es de esperar, esta especie se encontraba dentro del género *Torulaspora*. Con ello concluimos que esta podría ser la especie responsable de la transferencia de al menos este gen, si no la Region C por completo, a algunas cepas de la especie *Saccharomyces*. Posteriormente, en un trabajo realizado en colaboración con el grupo de la Dra. Sylvie Dequin, se llevó a cabo la secuenciación de la cepa CLIB 830T de la especie *T. microellipsoides*, con lo que se reveló la presencia de esta Region C que presentaba una región adicional de unas 80kb entre el primer y el segundo gen de la región

encontrada en *S. cerevisiae*. En ese mismo trabajo, se realizó una búsqueda de esta región en cepas de *S. cerevisiae* recién secuenciadas. En dos cepas, se encontró nuevamente esta región de genes, pero esta era mas similar a aquella encontrada en *T. microellipsoides*, ya que presentaban algunos de los genes vistos intercalados en la Region C de esta especie.

Estos últimos hallazgos hicieron dudar de la dirección de la transferencia de esta región. En vista de los resultados previos, nuestra hipótesis de trabajo decía que: *T. microellipsoides* habría sido la especie responsable de donar estos genes a aquellas especies del género *Saccharomyces*. Sin embargo, el hecho de haber encontrado estas regiones más parecidas a *T. microellipsoides* en dos cepas de *S. cerevisiae* y que solo tuviéramos la secuencia genómica de una cepa de *T. microellipsoides*, nos llevó a plantearnos la secuenciación del genoma del resto de cepas disponibles de *T. microellipsoides* con el fin de buscar estas regiones en ellas.

A partir de los ensamblajes de la secuenciación, pudimos encontrar estas regiones en tres de las cuatro cepas secuenciadas, lo que apoyaba que la transferencia hubiera sido desde alguna cepa de esta especie hacia aquellas especies del género *Saccharomyces*.

Tras la secuenciación de los genomas de las restantes cepas de *T. microellipsoides* pudimos obtener todas las secuencias del gen *FSY1*. Sorprendentemente, eran tres de las cuatro cepas las que presentaban una secuencia similar a la encontrada en *S. cerevisiae* a la que llamamos *FSY1A*. Una segunda copia paráloga de *FSY1A* a la que llamamos *FSY1B* fue obtenida en todas ellas y adicionalmente se encontró otra copia de *FSY1B* en tres de las cepas, a la que llamamos *FSY1B2*. Una vez obtenidas las secuencias, realizamos un análisis filogenético por el método de Máxima verosimilitud (*ML*) de las mismas. La reconstrucción obtenida reflejaba que aquellas secuencias que obtuvimos para las

cepas de *T. microellipsoides*, ocupaban una posición filogenética diferente a la que fue establecida para la filogenia de la familia por Kurtzman 2003 , siendo una copia de ellas idéntica a la de *S. cerevisiae* y las otras dos muy próximas a las del grupo de *S. eubayanus*, *S. uvarum* y *S. pastorianus*.

Para comprobar el resultado obtenido con el árbol Maxímo Verosímil, llevamos a cabo una comparación de topologías mediante el test de Shimodaira-Hasewaga. De esta forma, comprobábamos estadísticamente que este árbol obtenido era el que mejor que se adecuaba a nuestro alineamiento a pesar de ser incongruente con el árbol esperado para la familia Saccharomycetaceae. De esta comparación observamos que la mejor topología era la obtenida por *ML* para nuestras secuencias de *FSY1*.

Para apoyar la idea de que este gen hubiese sido transferido desde *T. microellipsoides* (una cepa pre-duplicación) a otras especies post-duplicación, inspeccionamos la localización cromosómica de este gen en diferentes especies cuyo genoma estaba anotado, por lo que la sinténia podía obtenerse. Utilizando este análisis, vimos que *FSY1* ocupaba una posición cromosómica central en las especies pre-duplicación *Kluyveromyces lactis* y *Lachancea kluyvery*, *L. waltii* y *L. thermotolerans*. Proponemos que, en un momento dado, *FSY1* pasó a tener una posición subtelomética en especies *Torlaspora delbrueckii*, *Zygosaccharomyces rouxii.* Tras la secuenciación de las cepas de *T. microellipsoides*, también observamos esta misma disposición para todas las copias presentes del gen *FSY1* (*FSY1A*/*FSY1B1*/*FSY1B2*).

Además del gen *FSY1*, encontramos un gen ortólogo en las especies del género *Saccharomyces* a otro de los encontrados en la Región C. Este gen era *ATO3*, cuya función ancestral descrita era el transporte de amonio. Tras realizar un análisis filogenético con las especies de la familia Saccharomycetaceae,

descubrimos que esta copia, que también se encontraba en una región subtelomérica, quedaba mas cerca en la filogenia con las secuencias encontradas en las cepas de *T. microellipsoides* para este gen. Con ello, obtuvimos que otros genes de esta famosa región, habían sido transferido a estas especies, bien siendo en el mismo o en eventos de transferencia diferentes.

Todos estos resultados realizados con tan solo dos genes y con las copias parálogas de los mismos, nos llevaron a observar detenidamente las fiologenias obtenidas con las secuencias de todas las cepas de *T. microellipsoides* disponibles. De ellas, extrajimos nueva información de forma que detectamos que, probablemente, dos de las cepas de esta especie podrían ser especies hibridas. De las mismas reconstrucciones conseguimos inferir cuales podrían ser las posibles especies parentales de aquellas. Para poder confirmar nuestras hipótesis, realizamos nuevos análisis filogenéticos en los que incrementamos el número de genes, obtenidos gracias a la secuenciación de los genomas y nuevamente los resultados reforzaban aquello que ya vimos en los análisis previos. Las cepas NRRL Y-17058 y CBS 6143, serían dos especies híbridas cuyos genomas estaban compuestos por dos subgenomas cuya procedencia estaba probablemente identificada: uno de ellos era muy cercano a la cepa CBS 6641 y el otro similar a la cepa CBS 6762.

De los mismos análisis, observamos que una de las cepas propuesta como posible parental, la CBS 6762, quedaba en una posición filogenética alejada con respecto a las otras cepas de *T. microellipsoides*. Por ello, hipotetizamos que esta cepa podría ser en realidad una especie cercana pero diferente a *T. microellipsoides*. Para reforzar este resultado, llevamos a cabo nuevas reconstrucciones filogenéticas añadiendo todas las secuencias de cepas de *Torulaspora* que aparecían en la bibliografía. Con ello determinamos que la cepa

CBS 6762 no se parecía mas a ninguna de las otras especies, sino que posiblemente se trataba de una nueva especie dentro del grupo de especies no- *Saccharomyces* mas similar a *T. microellipsoides*.

Las principales conclusiones obtenidas de esta investigación durante este periodo de doctorado son: primero, la descripción de la especie *T. microellipsoides* como la responsable de varios eventos de transferencia génica horizontal a especies del género *Saccharomyces*, lo que supone la transferencia de genes entre levaduras de géneros alejados por introgresión y probablemente por algún proceso de hibridación genómica. Segundo, el hallazgo de nuevas especies híbridas dentro de la especie *T. microellipsoides* y por tanto, no pertenecientes a los pocos géneros en los que ya se habían encontrado, como *Saccharomyces* o *Zygosaccharomyces* y de una posible nueva especies dentro de la familia Saccharomycetaceae gracias a la secuenciación de nuevos genomas.

Introduction

1. Yeast ecology and domestication

Yeasts are eukaryotic unicellular microorganisms that belong to the kingdom Fungi and phylums *Ascomycota*, *Basidiomycota* and *Deuteromycota* (fungi imperfecti). Budding yeast, which are the yeasts under study belong to the phylum *Ascomycota*, subphylum *Saccharomycotina*, class *Saccharomycetes* (or *Hemi-ascomycetes: hemi-: half; ascus-: sac-containing spores)*, order *Saccharomycetales*. More than 1,500 species of *Saccharomycetales* are known (Kurtzman et al. 2011) and, according to environmental surveys, it is estimated that large numbers of taxa remain to be discovered that will likely double the known number of species. Although some species are parasites of animals (including humans, e.g. different *Candida* species) and plants (e.g. *Eremothecium gossypii*), most hemiascomycete yeasts are free-living microorganisms present in nutritionally complex substrates rich in simple organic compounds that originate from plants (exudates, flowers, decayed fruits and leaves, etc.). They can also be found in association with animals (e.g. arthropods) to act as gut and intestinal commensals by providing vitamins and enzymes in exchange for efficient habitat and dispersal (Kurtzman and Fell 1998).

The *Saccharomyces cerevisiae* species has also become a good model eukaryotic organism for its simple and short life cycle, its easy culture and manipulation, properties that have favored its use in biochemical and genetic studies. In fact *S. cerevisiae* was the first eukaryotic organism for which the whole genome sequence was obtained (Goffeau 1996) and offers an unparalleled reference source for studying the basic molecular mechanisms of eukaryotic cells. As a result, it has been the subject of pioneering studies in the new *'-omic*' analyses. Sequencing the first *S. cerevisiae* complete genome revealed the presence of 376 duplicated genes in 55 large regions, which led Wolfe and Shields 1997 to postulate the occurrence of an ancient whole genome duplication (**WGD**)

event that occurred in an ancestor of *S. cerevisie* after its divergence from *K. lactis* about 100-200 million years ago. At present, genome sequences are available for many hemiascomycete yeasts and this information has allowed our understanding of several aspects of the evolution of this interesting group of eukaryotic microorganisms (Dujon and Louis 2017).

However, most people know that hemiascomycete yeasts produce CO_2 that makes bread rise and ferments sugary substrates to alcohol because Pasteur recognised the role of yeasts in the production of wine, beer, bread, etc. Some of these yeasts, particularly *S. cerevisiae*, have been responsible for the production of different fermented beverages and foods since the first human societies developed, and are currently used in industrial and biotechnological processes of major economic importance.

Although *S. cerevisiae* is the main yeast used in the production of fermented foods and beverages, other yeasts also make important contributions (Fleet 2007). For example strains of *Debaryomyces hansenii*, *Yarrowia lipolytica* and *Kluyveromyces marxianus* play roles in the development of flavour and texture during cheese maturation. *D. hansenii*, *Y. lipolytica* and some *Candida* species also contribute to the production of meat by products, such as fermented salami, sausages and cured hams. Cocoa beans must be fermented to generate the precursors of chocolate flavour, and this fermentation is carried out by some species of *Hanseniaspora*, *Candida*, *Issatchenkia* and *Pichia*. Coffee beans go through a cleaning process to remove pulp and mucilaginous materials and certain strains of *Candida*, *Saccharomyces*, *Kluyveromyces*, *Saccharomycopsis*, *Hanseniaspora*, *Pichia* and *Arxula* have been associated with this process.

Given its importance in the production of fermented foods and beverages, *S. cerevisiae* studies have focused on it to understand the underlying mechanism that

makes this yeast the best fermenter microorganism (Piškur et al. 2006). This, as well as other related species, are characterised by their ability to ferment simple sugars, mainly mono- and oligo-saccharides, into ethanol and CO_2, even when oxygen is available for aerobic respiration, the so-called Crabtree effect (Hagman et al. 2013).

Although alcohol fermentation is energetically less efficient than aerobic respiration, it provides a selective advantage to these yeasts to out-compete other microorganisms: sugar resources are consumed faster, while the ethanol produced during fermentation, as well as higher levels of heat and CO_2, can be harmful or less tolerated by their competitors. Once competitors are overcome, *S. cerevisiae* can then use the accumulated ethanol as a substrate for aerobic respiration. This ecological strategy, known as (ethanol) 'make-accumulate-consume'(Thomson et al. 2005; Piškur et al. 2006), has been suggested to possibly appear at the time that angiosperms emerge to produce sugar-rich fruits.

At least three yeast lineages, including *S. cerevisiae*, *Dekkera bruxellensis* and *Schizosaccharomyces pombe*, have independently developed the conversion from sugars into ethanol. Two hypotheses for this emerging effect move between 1) the possibility of an early origin in the evolution of Ascomycetes, which could later be lost in different groups; 2) a fermentative life-style appeared and was selected in different lineages (Hagman et al. 2013).

After a WGD, an event some genera underwent before the split of *Saccharomyces* and *Vanderwaltozyma* (see the definition later), novel functions in new duplicated genes can be tested by evolution, and it probably happens for alcohol dehydrogenase genes and hexose transporter genes (HXT). Based on the ability to produce ethanol under aerobic conditions, 40 yeast species can be classified as Cabtree-positive or Cabtree-negative (Hagman et al. 2013). The origin

of the Cabtree effect was determined according to biomass and ethanol production in different species of the Saccharomycetaceae family. It appears to have originated before *S. cerevisiae* separated from the *Kluyveromyces* lineage, thus all the post-WGD genera and the *Zygotorulaspora*, *Torulaspora* and *Lachancea* lineages presented this effect which became even more pronounced from pre-WGD to post-WGD.

The selective advantages of the 'make-accumulate-consume' strategy are only evident in sugar-rich environments, including the artificial fermentation environments created by humans to produce wine, beer, sake and other alcoholic beverages. In this way, *S. cerevisiae* found a new ecological niche, which it successfully occupied, in the crushed grape berries collected by humans to produce the first fermented beverages.

The oldest winemaking evidence is the tartaric acid found in a jar of Mesopotamic origin that dated back to 5400-5000 BC. This arqueological discovery suggests that wine was mainly drunk and offered to gods during religious ceremonies or used as medicine. Wine production started in the Fertile Crescent area and expanded along the Mediterranean Sea by Phoenicians and Greeks before being taken to the northern limit of vine growth by Romans, and then on from the end of the 15th century to the 19th century, it expanded to America, South Africa and Australia. Beer was also a popular drink in ancient times from Egypt to Mesopotamia. Ale beer was prepared from cooked malt with water, and then these mixture was strained free of husks before spontaneous yeast inoculation. Bread making evidence has also been found in ancient Egypt, but more information about its expansion is less known (Sicard and Legras 2011).

The Galatians, a Celtic tribe that colonised Central Anatolia, mediated beer expansion from the ancient Middle East to Central and Northern Europe. Beer became a popular fermented beverage among Celtic and Germanic tribes since the 1st century AD. However lager beer, which is fermented at low temperature, was an innovation from the end of the Middle Ages in Central Europe (Sicard and Legras 2011).

Therefore, it is most interesting to know the molecular mechanisms involved in the evolution of this yeast during its adaptation to the new environments created by humans, a process commonly known as **"unconscious domestication"**. Numerous evidences of this important process in yeast have been detected (Libkind et al. 2011; Sicard and Legras 2011; Almeida et al. 2015; Borneman and Pretorius 2015; Marsit and Dequin 2015).

Since then, wine *S. cerevisiae* yeasts have been exposed to selective pressures due to the fluctuating stresses that occur during wine fermentation, such as osmotic stress due to high sugar concentrations (180–300 gL^{-1}), anaerobic stress, acid stress, nutrient limitations, ethanol toxicity, etc. (Querol et al. 2003) . As a result, wine *S. cerevisiae* yeasts exhibit differential adaptive traits (Marsit and Dequin 2015) and shape a genetically differentiated population (Fay and Benavides 2005; Legras et al. 2007; Liti et al. 2009; Almeida et al. 2015; Peter et al. 2018).

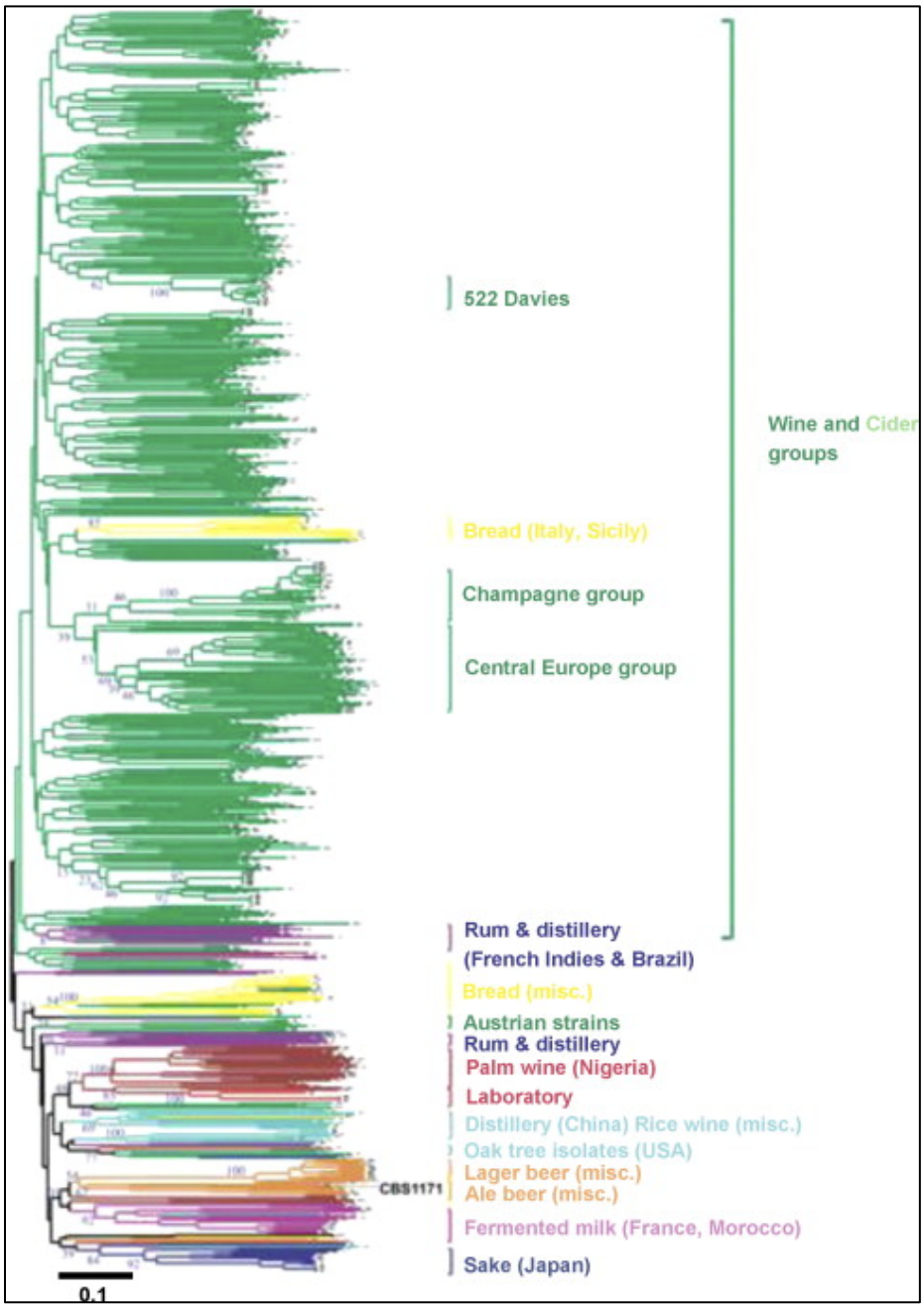


Figure i1. Neighbour-joining tree reconstruction of 651 yeast strains indicating their origin. Green shows that the strains from wine and cider clearly constitute a well-defined group. With less representation, but also clustered by their origin (except for bread from Sicilia, Italy) we see rum, bread, rice wine, oak tree, lager beer, ale beer, fermented milk, sake and lab strains. Extracted from Sicard and Legras 2011.

Numerous yeast species have been isolated in the last twenty years from natural and fermentative environments. With *S. cerevisiae*, only a few strains have been found in soil and bark of oak trees, but an enormous number of them have been found in wine, beer, sake and other fermented beverages. This fact led to two controversial opinions about whether this species was domesticated by humans and then arose in nature, or if it was already in nature and then selected in human activities (Fay and and Benavides 2005). A comparison of both origin strains revealed that while domesticated *S. cerevisiae* strains present less genome variability, natural strains harbour more genetic variability and represent the oldest linages. About this result, a more plausible hypothesis refers to a natural origin of *S. cerevisiae* yeast that was then domesticated by humans. Strong selectivity on domestication has been shown when *S. cerevisiae* strains were clustered according to their ecological or fermentative niche more than their geographical origin (Sicard and Legras 2011). This also suggests independent domestication events for sake, wine, beer and palm wine fermentations (Figure i1). It is noteworthy 95% of the wine yeast isolated around the world cluster together, which means they all of have the same origin.

Domestication has employed distinct selected strategies that we can observe in the yeast genome. First of all, some genes have undergone changes, such as *SSU1*. This gene presents resistance to sulphite, an antiseptic used to clean wine containers since the Egyptians until the present day. A translocation between chromosome VIII and chromosome XVI in the promoter region of this gene is enough for the induction of *SSU1* to expulse sulphite from the cytoplasm, by thus making a more resistant organism (Pérez-Ortín et al. 2002) . A more complex, but frequently selected, strategy consists in the formation of hybrid genomes. The most popular example here refers to the formation of a *Saccharomyces* interspecies hybrid between species *S. cerevisiae* and *S. eubayanus* called *S.*

pastorianus. This hybrid is typically found in low-temperature lager beer fermentation, so the mixing of the two parental strains (*S. cerevisiae* a good fermenter and *S. eubayanus* is more adapted to low temperature) imposed a major advantage that overcame reproductive isolation (Dunn and Sherlock 2008).

2. Evolution of the yeast reproduction strategy

Hemiascomycete yeasts, like all fungi, may have asexual and sexual reproductive cycles (Figure i2). In yeasts, the commonest mode of vegetative growth under nutrient-rich and non-stressful environmental conditions is asexual reproduction by budding. A small bud or daughter cell is formed on the parent cell. The nucleus of the parent cell splits into a daughter nucleus and migrates into the daughter cell. The bud continues to grow until it separates from the parent cell to form a new cell. This mitotic division by budding occurs in both the haploid and diploid phases of the life cycle.Conjugation or mating in yeasts consists of the fusion of two haploid cells of opposite mating types (**a** and **α)** to form a diploid zygote. These mating types are determined by a single locus *MAT* with two alleles, ***MATa*** and ***MATα***. Rather than alleles, they are idiomorphs as they differ in sequence, size and gene content terms. In general, under high-stress conditions such as nutrient starvation, haploid cells die unless they conjugate to generate a diploid cell. Under stressful conditions diploid cells can undergo sporulation by entering sexual reproduction (meiosis) and producing four haploid spores (an ascospore tetrad), which are dormant and resistant to harsh environmental conditions. When returned to a rich medium, ascospores germinate into metabolically active haploid cells of two mating types, *MATα* and *MATa*, which can reproduce asexually by budding or conjugating into diploid cells (Butler et al. 2004).

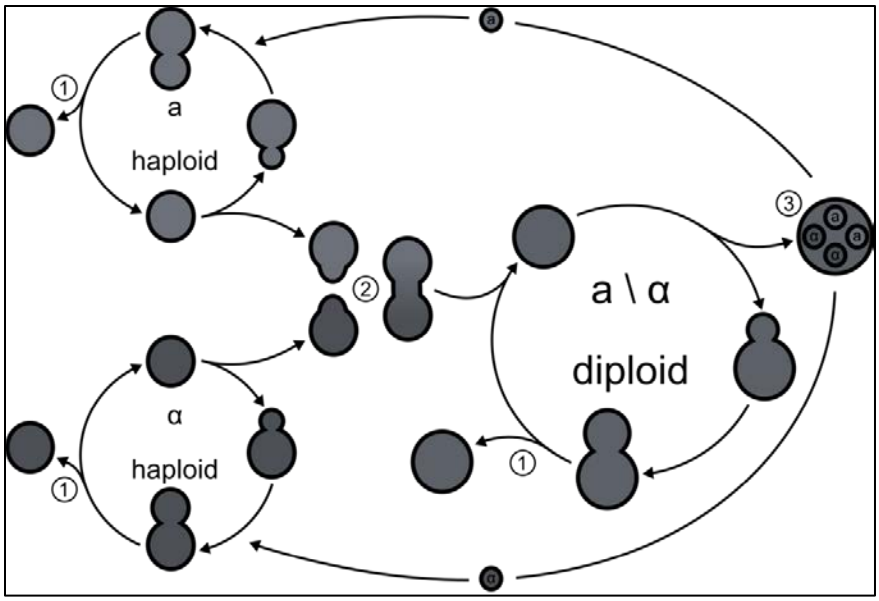


Figure i2. Hemiascomycete yeast life cycle. 1. Diploid or haploid yeast asexual reproduction by budding. 2. Conjugation or mating of two haploid cells of different mating types (**a** and **α**) to generate a diploid cell. 3. Sporulation, generation of haploid cells by the meiotic division of a diploid cell. Image from https://en.wikipedia.org/wiki/Yeast.

On the basis of their sexual reproduction, hemiascomycete lifecycles can be classified into haplobiontic, diplobiontic and haplo-diplobiontic according to the predominance the haploid or diploid phases, or to the equal presence of both. In yeasts with a haplobiontic life cyle, such as most species of the *Torulaspora, Zygosaccharomyces and Zygotorulaspora* genera, the haploid phase is predominant. Under harsh environmental conditions, the haploid cells of the two mating types conjugate to form a diploid zygote, which immediately enters meiosis to generate resistant haploid ascospores. Under appropriate environmental conditions these ascospores will germinate to yield haploid cells, which will grow asexually by budding.

In diplobiontic yeasts, such as *Saccharomyces* genus species, the diploid cycle is predominant. Diploid yeasts reproduce asexually by budding in rich nutrient medium. However, if placed in a medium lacking sufficient nitrogen to maintain mitosis, diploids can undergo meiosis to produce haploid ascospores. When returned to a rich medium, ascospores germinate into metabolically active haploid gametes of the two mating types to conjugate and recover the diploid state.

In species with haploid mating types, e.g. *S. cerevisiae*, three different mating behaviours are possible: outcrossing (amphimixis), intra-tetrad conjugation (automixis or auto-fecundation) and autodiploidisation (or haplo-selfing) (Figure i3.). Outcrossing is the mating of the haploid cells that derive from the meiotic products of unrelated diploid cells. Intra-tetrad conjugation is the mating of the haploid cells that derive from the meiotic products of a single cell. Autodiploidisation or haplo-selfing involves the conjugation of the haploid cells that derive from the mitotic division of the same haploid cell (Knop 2006). Autodiploidisation is possible only when one of the haploid cells (mother), which derives from a mitotic division by budding, is able to change its mating type and conjugate with the other cell (daughter). This mother-daughter mating produces a homozygous diploid cell for all genes except for the *MAT* locus. The yeasts able to switch mating type are called homothallic, and those that never had or lost this capability are heterothallic.

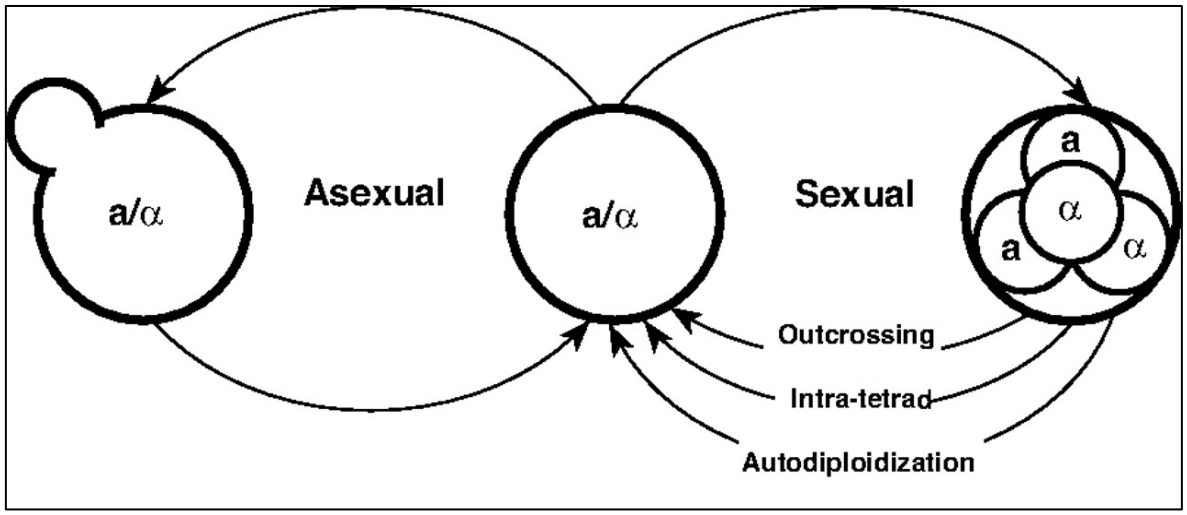


Figure i3. Sexual and asexual reproduction in hemiascomycete yeasts. A diploid yeast asexual life cycle is based on growing in size until budding, the time when a daughter cell (or bud) emerge to generate a new younger individual from an identical one (mother cell). Under special conditions, cells can sporulate by meiosis to generate four haploid spores, each with a single MAT idiomorphical type (a or α). While inside the ascus, two spores of different sexual types can conjugate to form a new diploid adult. After sporulation, each spore can intercross with another haploid from outside the tetrad (outcrossing). Otherwise when it grows and starts budding, a diploid adult cell can arise from conjugation with the daughter cell. Modified from Tsai et al. 2008.

Mating-type genetic structures have appeared or have been lost across different yeast lineages (Figure i4B). A single *MAT* locus is observed in the "basal" subgroup of this subphylum, yet a duplicated *MAT* locus is present in Methylotrophs. This second locus is located in the same chromosome, but in a **subtelomeric** position, and consists in a silenced *α1α2 MAT* locus that can be flipped with the *α1α2* expressed one to change the sexual type. Hence, these species are heterotallic. In contrast across the CTG subgroup, a single *MAT* locus called MTL, is retained that bears either *MATa* or *MATα*, but in some species it is presented as a mixture of them and does not change the mating type. Finally, a two-step evolutionary process is proposed for the current system of the Saccharomycetaceae subgroup (refer to Figure i4A). The first event was to acquire one α and one a *silent cassette* or **HML** (*MATα1* and *MATα2*) and **HMR** (*MATa1* gene) cassettes. Both are subtelomeric sider non-expressed copies of the MAT locus, which are involved in the mating type switch by a homologous recombination with the expressed locus. The *MATa2* locus is present in some non-WGD species of the *Kluyveromyces*, *Lachancea* and *Zygosaccharomyces* genera, but a posterior loss of this gene, which codifies for a transcription factor is seen after WGD. An endonuclease and an endo-nucleotide site were necessary to carry out the mate type change. The second event was to acquire the **Ho** endonuclease (derived from a homing endonuclease encoded in an intein) and its endo-nucleolitic site in *MATα1*. This site-specific endonuclease promotes the switching of mating type by a recombination via a DNA double-strand break and a posterior repair with the proper silent cassette. The ***HO*** gene present is the *Z. rouxii* and *T. delbrueckii* species (defined later as the ZT clade), but species of the *Kluyveromyces*, *Lachancea* and *Eremothecium* genera are absent (defined later as the KLE clade). It is known that in *K. lactis*, transposases α3 and Kat1 perform this activity by changing *MATα* into *MATa* and *MATa* into *MATα*, respectively. However,

due to the independent acquisition of the silent cassettes and the Ho endonuclease (Figure i4A), some Saccharomycetaceae species like those from the *Zygosaccharomyces*, *Saccharomyces* and related genera, possess an efficient mating type, while others, like those from the *Kluyveromyces* and *Lachancea* genera, have an imperfect system that depends on the spontaneous recombination of silent cassettes and the *MAT* locus (Butler et al. 2004; Dujon and Louis 2017).

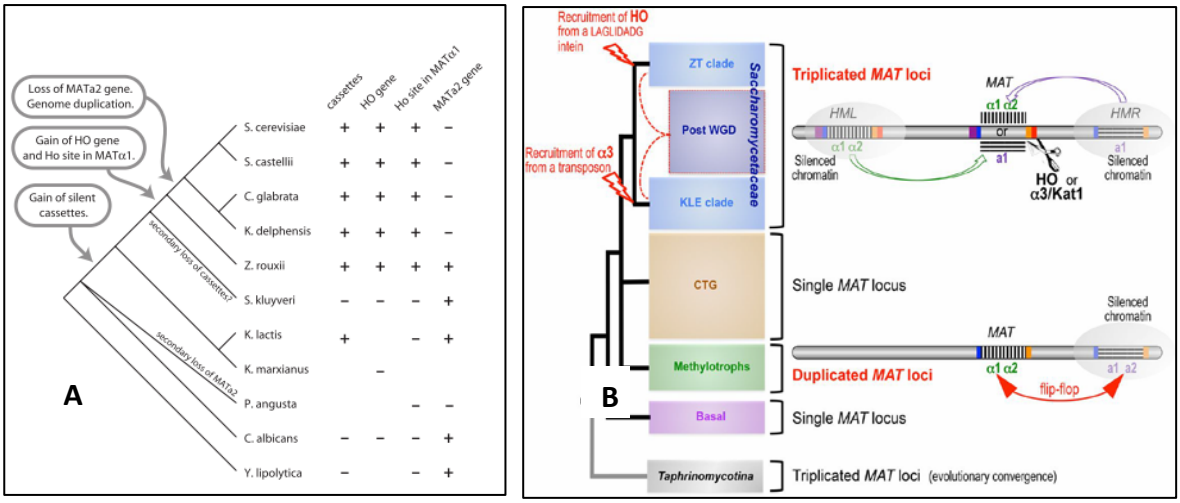


Figure i4. *Saccharomycotina* and hemiascomycete reduced phylogenies (not in scale) showing the evolution of a mating-type switch system. **A**: Gain of silent cassettes appeared before *Kluyveromyces* divergence and the *HO* gene is present in all the species that diverged after *Z. rouxii*. Adapted from Butler et al. 2004; **B**: Gain or loss of *MAT* locus copies are shown for the different hemiascomycete linages.

3. Hemiascomycete yeast systematics and phylogeny

Detection, identification and classification of yeasts have greatly transformed in the last few decades following the application of gene sequence analyses and genome comparisons. The use of phylogenetic analyses of gene sequences has led to a major review of yeast systematics, which has resulted in a redefinition of nearly all genera. In a recent review, (Kurtzman et al. 2015) authors proposed subdividing the Saccharomycotina species into 12 clades (Figure i5). Four of them englobe more species than others. Clade 1 corresponds to the Saccharomycetaceae family and includes most yeast of biotechnological interest. Some characteristics of this group are that they present well-defined centromeres and triplicated MAT loci, but lack the genes for complex I subunits of the respiratory chain in their mitochondrial DNA. A single rDNA locus is frequently presented and the 35S precursor and the 5S gene define the structure.

Genome size ranges from 9-14Mb and present a high density of coding sequences, which come close to 4500-5900 protein-coding genes, but only a few need to be processed by splicing (3-5% of CDS). Clade 6 belongs to the “CTG clade”. Long centromeres (~3-4 kb) with no conserved sequence motif can be observed in the species of this group, and they also present a single MAT locus. DNA content is slightly larger than in Saccharomycetaceae (10-16 Mb), as is also the density of protein coding genes (~5600-6400) and spliced sequences (6-7% of CDS). Moreover, they present all the subunits of complex I of the respiratory chain. They are distinguished from others for the alternative use of the genetic code by using CUG for serine instead of leucine or for both serine and leucine. The Methylotrophs group is marked as clade 5 and comprises species with a similar genome size to those species from the Saccharomycetaceae cluster. Switching the sexual type is possible as they present duplicated MAT loci. They share some other signatures

with the “CTG clade” as regards similar regional centromeres and compact genomes, and they present all the complex I subunits. Finally, clade 9 corresponds to the cluster of the “basal linages”, which show wide variability among the species inside. Common signatures refer to many protein-coding genes (more than 6000 in sequenced genomes), but genome size and spliced genes differ between species.

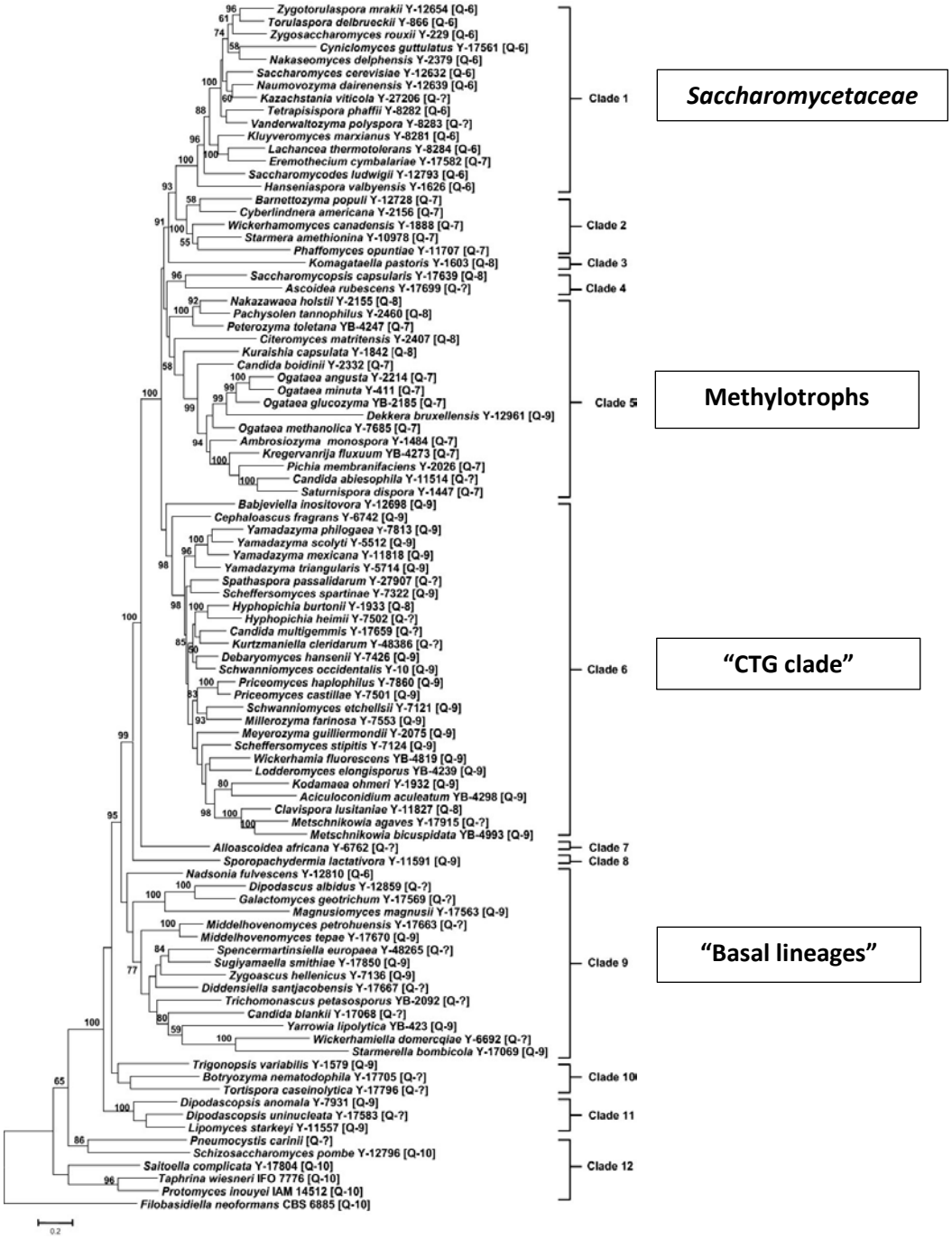


Figure i5. Phylogenetic relationships among the type species of hemiascomycete yeast. Genera were determined from a maximum likelihood analysis of the concatenated gene sequences for LSU rDNA, SSU rDNA, translation elongation factor-1α, and RNA polymerase II, subunits B1 and B2. From Kurtzman et al. 2015.

The phylogenetic relationships among species of the Saccharomycetaceae family, formerly included in four main genera, *Kluyveromyces, Torulaspora, Saccharomyces* and *Zygosaccharomyces*, were also obtained by Kurtzman 2003, who observed that these genera were not monophyletic groups, and hence proposed a major review and redefinition of the genera as depicted in Figure i6.

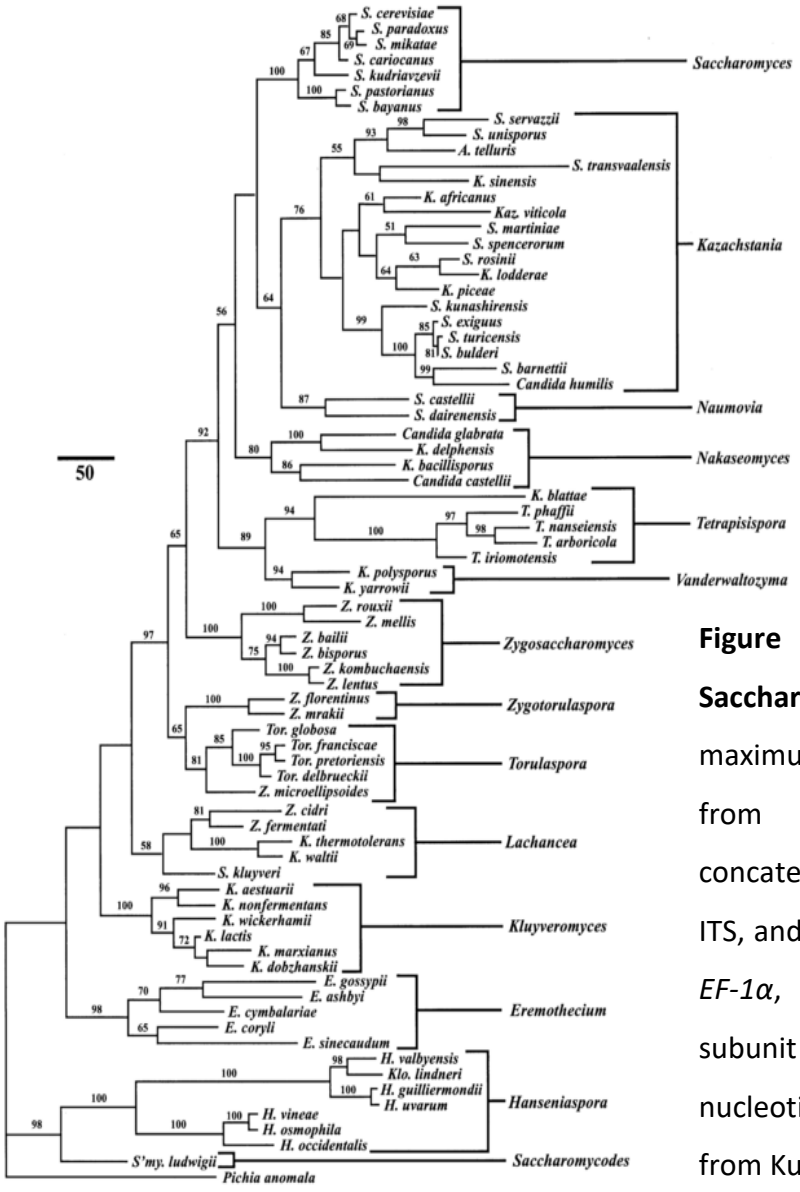


Figure i6. The Saccharomycetaceae complex. A maximum-parsimony tree derived from an alignment of the concatenated 18S, 5.8S/ alignable ITS, and 26S (three regions) rDNA, *EF-1α*, mitochondrial small-subunit rDNA and *COXII* nucleotide sequences. Adapted from Kurtzman 2003.

Accordingly, the former *Saccharomyces* genus contained the species in the ***Saccharomyces sensu stricto*** group (*S. paradoxus*, *S. cerevisiae* and *S. bayanus*) and the new discovered species. The species included in the older ***sensu lato*** group (*S. martiniae*, *S. exiguus*, *S. castellii*, *S. barnettii,* etc.) were transferred to the *Kazastania* and *Naumovia* genera.

So this new *Saccharomyces* genus, described by Kurtzman 2003, included *S. cerevisiae*, its sister species *S. paradoxus* (only isolated in natural environments) and *S. bayanus* (found in natural environments, and in beer and wine fermentations). Three new species, *S. cariocanus* from Brazilian forest and *S. kudriavzevii* and *S. milkatae* both isolated from Japan, have been described by Naumov et al. 2000. Genetic hybridisation analysis provided species status to these isolates because a postzygotic barrier was seen according to low spore viability (0% of viable spores, only some of the crosses formed one to four microcolonies), when intercrosses were obtained with reference strains from each species. For *S. cariocanus*, a previous work (reference not available but indicated in Naumov et al. 2000) indicated that the Brazilian strains were genetically related to *S. paradoxus*. In order to clarify if they could be treated as separate species, interbreeding with *S. paradoxus* strains from different populations (European, North American, Far- East Asian and Hawaiian strains) and the resulting spores were tested here. As obtained before, viability was 0%, which indicates that it was a different species. This assignation was based on the ***biological species concept*** after establishing that *S. cariocanus* was reproductively isolated. As for the other two species, the currently available *S.mikatae* strains proceed them all from Japan (Borneman and Pretorius 2015), while the new *S. kudriavzevii* has been discovered in Europe (Sampaio and Gonçalves 2008). “Cryotlerant” strains have been found in the bark of *Quercus* spp. in Portugal that differed from the Japanese population in terms of possessing of an active galactose utilisation pathway. Later *S. kudriavzevii* was also isolated in oak

trees from different parts of Spain and, together with the Portuguese ones, a European origin was derived for *S. cerevisiae* by *S. kudriavzevii* hybrids (Lopes et al. 2010). In addition *S. pastorianus* (syn. *S. carlbergensis*), responsible for lager beer fermentation, was included in the "*Saccharomyces* complex", but a few years before, some researchers had already demonstrated that it was a hybrid between *S. cerevisiae* and another species related to the ancient *S. bayanus* species (now *S. eubayanus*). The last incorporation into the genus is *S. arboricolus,* which is isolated from the bark of broadleaf *Fagaceae* trees from different regions of China (Wang and Bai 2008).

About the *S.bayanus* species arose considerable controversy with competing groups arguing that two varieties of the same species, *S. bayanus* var. *uvarum* and var. *bayanus* or two natural species *S. bayanus* and *S. uvarum*, were included (Nguyen and Gaillardin 1997). The recent discovery of a new species, *S. eubayanus*, from natural environments in Patagonia (Libkind et al. 2011) clarified the status of this former *S. bayanus* species complex, as well as the origin of *S. pastorianus*. *S. uvarum* is described here as a natural species and *S. bayanus* a hybrid lineage generated by the hybridisation between *S. uvarum* and *S. eubayanus*, which is the donor of the non-*S. cerevisiae* portion of the hybrid *S. pastorianus* (Figura i7).

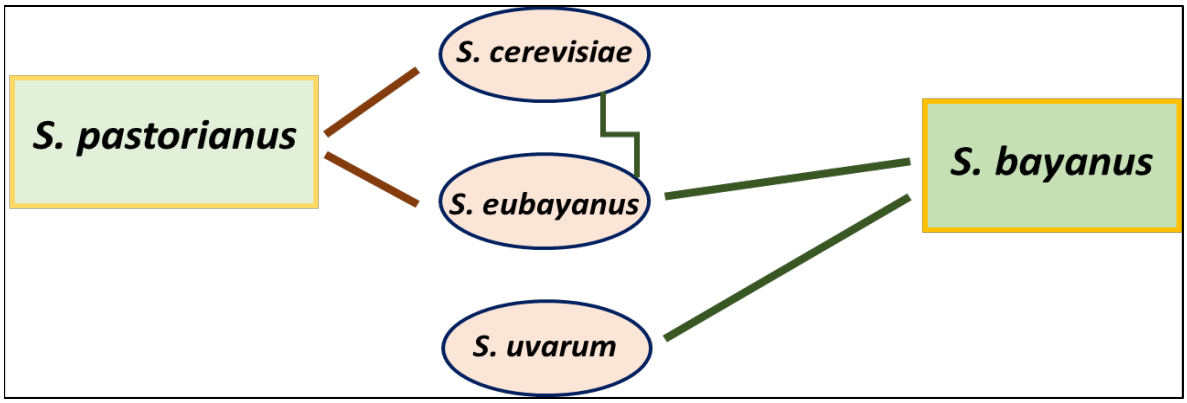


Figure i7. **The origin of the *S. pastorianus* and *S. bayanus* hybrids**. Hybrid *S. pastorianus* is composed by the genomes of *S. cerevisiae* and *S. eubayanus* while the complex hybrid species *S. bayanus* is mostly composed of *S. eubayanus* and *S. uvarum* species and some traces of *S. cerevisiae.*

In recent years, a more proper redefinition of *Saccharomyces* genus has appeared, mainly with the progress made in genome sequences in more group species. Since then, numerous works have appeared to attempt to clarify the inclusion of some species, or if they can be treated as varieties of species, as real species or as hybrids of them. Controversy about the *S. bayanus sensu stricto* species arose from the molecular characterisation of this species (Nguyen and Gaillardin 1997). PCR amplification and restriction polymorphism of the internal transcribed spacer 1, ITS 1 (authors wrongly referred to the second non-transcribed spacer) in ribosomal DNA, in addition to DNA karyotyping, showed the existence of two different subgroups inside the *S. bayanus* species. To distinguish them, one was called ***S. bayanus var. uvarum*** as it was found mainly in wine, and the other ***S. bayanus var. bayanus***. The variety characteristic was first proposed because of the semisterility of the hybrids between them (Naumov 2000). A short time afterwards, one of these varieties was proposed as a new species inside *Saccharomyces sensu stricto*: *S. uvarum* (Pulvirenti et al. 2006). The main differences with *S. bayanus* were the physiological characteristics described before by Naumov 2000, such as

the ability to ferment melibiose, but the inability to grow above 37^0C, and molecular aspects like karyotyping profile and the sequence of NTS2 (non-transcribed spacer 2) and partial 26S. Nowadays the strains from *S. uvarum* exist from natural environments than from low-temperature industrial fermentations. However, only type strain CBS 7001, from an insect found in Spain, is its genome sequenced. It is noteworthy that in previous papers, *S. bayanus* has been presented as a distinct species with a fermentation profile in grape juice and an electrophoretic karyotype typical of hybrid cultures, which does not sporulate.

The complete solution for these species came in 2011 when 123 isolates of "cryotolerant" *Saccharomyces* from the *Notofagus* species from Patagonia were identified as strains related to either *S. bayanus* or *S. uvarum* (Libkind et al. 2011). Each group of strains showed wide intra-spore viability, but hybrids were semisterile as seen before (7.3% of viability). Thus postzygotic isolation was enough to conclude that they were two different species, but each one was associated with different *Notofagus* tree species. Finally, the genome sequencing of isolated individuals showed their purity or hybrid nature. While the group related to *S. uvarum* indeed had a very low degree of divergence (0.52%) to the *S. uvarum* reference strains and also to the type strain, the group that appeared to be related to *S. bayanus* was in fact closely related to the part of the *S. pastorianus* genome non-*S. cerevisiae*. When the genome of the *S. bayanus* type strain (CBS 380T) was sequenced (Okuno et al. 2016), a hybrid strain was found to be composed (over 60%) of *S. uvarum*, and the second group was apparently related to *S. bayanus*. Some traces of the *S. cerevisiae* genome, mainly the genes involved in maltose and maltotriose metabolism, were detected in the type strain. All these conclusions were determinant to propose the new pure and ecologically isolated wild species ***S. eubayanus*** as the non-*S. cerevisiae* contributor of *S. pastorianus* and the non-*S. uvarum*/non-*S. cerevisiae* part of the *S. bayanus* hybrid species (Libkind et al. 2011).

The genes related with the better utilisation of maltose (MAL family), which is the most abundant sugar in wort, were found to be **horizontally transferred** from the *S. cerevisiae* to *S. bayanus* hybrids, as seen in *S. pastorianus*. Moreover, the overexpression of *S. eubayanus* allele *SUL2* of *S. bayanus* correlated with an improved sulphite production, which is also important in brewing.

An examination of the expected differences between brewing domesticated *S. pastorianus* and wild *S. eubayanus* was carried out to reconstruct a model (Figure i8) about the formation of human employed hybrids in the brewery industry: firstly, the *S. eubayanus* wild strains would fuse with ale-type brewer *S. cerevisiae* to form **allotetraploid** hybrids, which would be the *S. pastorianus* ancestors. At this point, the chromosomal translocations between both genomes produce some of the characteristics seen in the modern *S. pastorianus*, such as aneuploidy fragments or loss of heterozygosity in some regions where the *S. cerevisiae* genome is visible. The contribution of *S. cerevisiae* regions to the *S. bayanus* hybrid genomes was possible by mixing *S. eubayanus* chromosomes, which were able to recombine with those of an *S. uvarum* hybrid content. Hypothetically prior to *S. bayanus* formation, *S. pastorianus* DNA fragments would transform the *S. eubayanus* strains. Surprisingly, no *S. eubayanus* strain has been isolated in Europe. So it is hypothesised that they had to be imported from Patagonia. Here two separate *S. eubayanus* populations were found, and both exist in the industrial *S. pastorianus* and *S. bayanus* hybrids from Europe. In addition, an admixture of these Patagonian populations is located in North America.

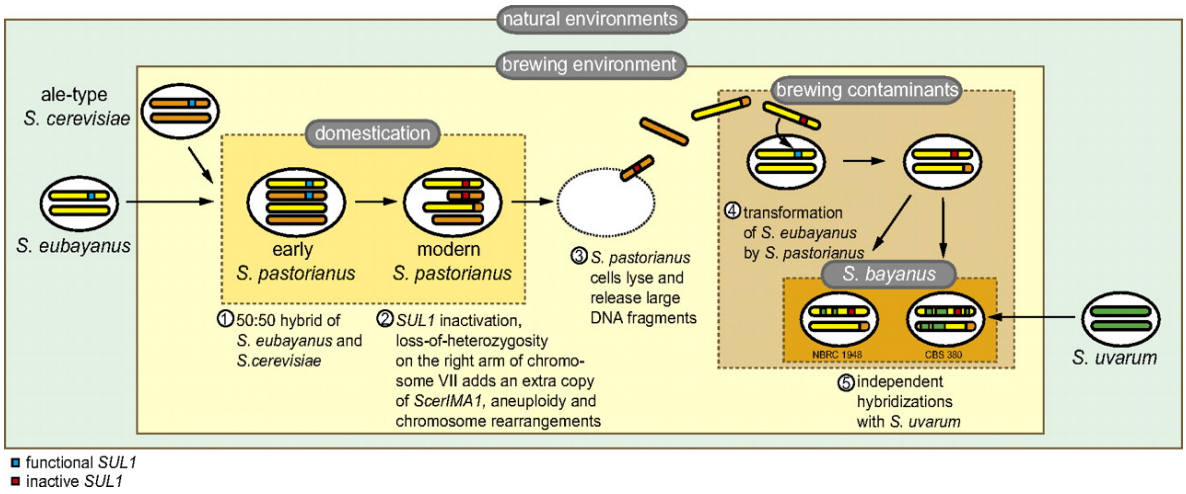


Figure i8. **A schematic representation of the *S. pastorianus* and *S. bayanus* hybrids' parental species composition**. *S. pastorianus* is an interspecies hybrid constituted by the parental genomes of *S. cerevisiae* and *S. eubayanus*. Genome sequencing has shown different quantitative contribution of each parental strains according to the *S. pastorianus* type of fermentation (Saaz-type and Frohberg-type strains). The more infringing *S. bayanus* hybrid is definitively a mixed *S. eubayanus* and *S. uvarum* genome with strains specific traces of *S. cerevisiae*, which correspond mainly to subtelomeric fragments. Adapted from Libkind et al. 2011.

The reproductive isolation of the *Saccharomyces* species arises in nucleotide divergence rather than in translocation that prevents spore viability (Hittinger 2013). Scape of those barriers has been possible for stable interspecies hybrid formation (by the polyploidisation of the final genome), whose evolutionary selectivity has been reflected in preponderance in wine or beer fermentations. Each parental strain contributes with particular and selected genes to bring the desired property of such species: **the hybrid vigour** (Shull 1948)**.**

As mentioned before, human yeast "domestication" has imposed important hybrid strains for the brewery industry, such as *S. pastorianus* or *S. bayanus.* Concerning the former, two sublinages have been finally determined by genome sequencing and divergence between them might be explained as different cryotolerance behaviour (Nakao et al. 2009; Walther et al. 2014): one is Saaz-type strains (*S. carlbergensis*), generally a 2:1 *S. eubayanus*: *S. cerevisiae* triploid genome, and the perfect tetraploid Frohberg type (Weihenstephan strain WS34/70). The shared molecular footprints between these subgroups, such as the degree of nucleotide variation, and in *S. eubayanus* fragments the reciprocal translocations between subgenomes, suggest that they differred after common ancestral hybridisation (Peris et al. 2014). The *S. cerevisiae* and *S. kudriavzevii* hybrid species (***Sc x Sk*)** were found in Belgian-style brewing (Hittinger 2013).

In winemaking, low-temperature fermentation has been dominated by hybrid species ***Sc x Sk*** and *S. bayanus* hybrid strains (Borneman and Pretorius 2015). The complete genome sequencing of the ***Sc x Sk*** VIN7 wine strain (Borneman et al. 2012) showed a genome of an allotriploid formed by a ***Sc*** heterozygous diploid and a haploid ***Sk*** from Europe. The contributions of the ***Sk*** subgenome concerned not only cold adaptation, but also the production of aromatic compounds. It is noteworthy that, the comparative genomics on ***Sc x Sk*** hybrids from wine and beer fermentative environments have revealed different origins from those interspecies hybrids (Peris et al. 2012), which suggests the selectively of pure species properties in only one genome.

By taking the last reorganisation from the *S. eubayanus*/*uvarum*/*bayanus* sequencing project (Libkind et al. 2011), the ***Saccharomyces* genus** formed a monophyletic group composed of seven natural (some domesticated) pure species (Hittinger 2013): *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, *S.*

arboricola, *S. eubayanus* and *S. uvarum* and historically selected industrial yeast hybrids *S. cerevisiae* x *S. kudriavzevii*, *S. bayanus* and *S. pastorianus* (Figure i9). Not only in vineyards or breweries, but also in nature, can we find strains of all species that have been isolated from particular tree species like *Quercus spp*. (or *Notophagus spp.* if they were found in the Southern hemisphere) in many tree parts, like leaves, bark or exudates, and in soil circumscription (Libkind et al. 2011).

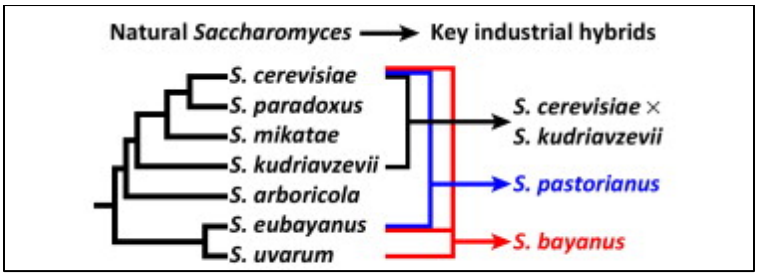


Figure i9. **The currently accepted *Saccharomyces* genus phylogeny, supported by genome sequencing**. On the left hand side of the panel we can see the relationship among the seven species isolated from natural niches. The interspecies hybrids between the genera species are appreciated for their key role in winemaking and brewing, and are shown on the right of the figure. Adapted from Hittinger 2013.

For its prevalence in oak trees, and overall in oak leaf litter rather than in oak bark (Kowallik and Greig 2016), *S. paradoxus* has been supposed as the natural predecessor of *S. cerevisiae* domesticated yeast. Common properties, such as having the same profiles of assimilation and fermentation of organic compounds, shared habitat and phylogenetic clade, make them seem very similar species. Some known differences about optimal growing temperatures are that *S. cerevisiae* moves closely to 30^0C, but *S. paradoxus* accommodates a wide range. Unlike *S. cerevisiae*, which has been found mainly in fermented beverages and discretely in wild ones, *S. paradoxus* is not assumed involved in fermentative processes. The former hypothesis of a domesticated *S. paradoxus* to give rise to *S. cerevisiae* has

been rejected as more genomic studies appeared and revealed that differentiated populations exist between these species (Liti et al. 2009).

A very recent work has incorporated a new species into the genus "*S. jurei*", a sister species of *S. mikatae* (Figure i10). Two strains have been isolated from a 1000-metre altitude on oak tree bark and surrounding soil from Saint Auban in France, and classified as different species from the other *Saccharomyces* in genetic hybridisation analysis terms (biological species concept) (Naseeb et al. 2017).

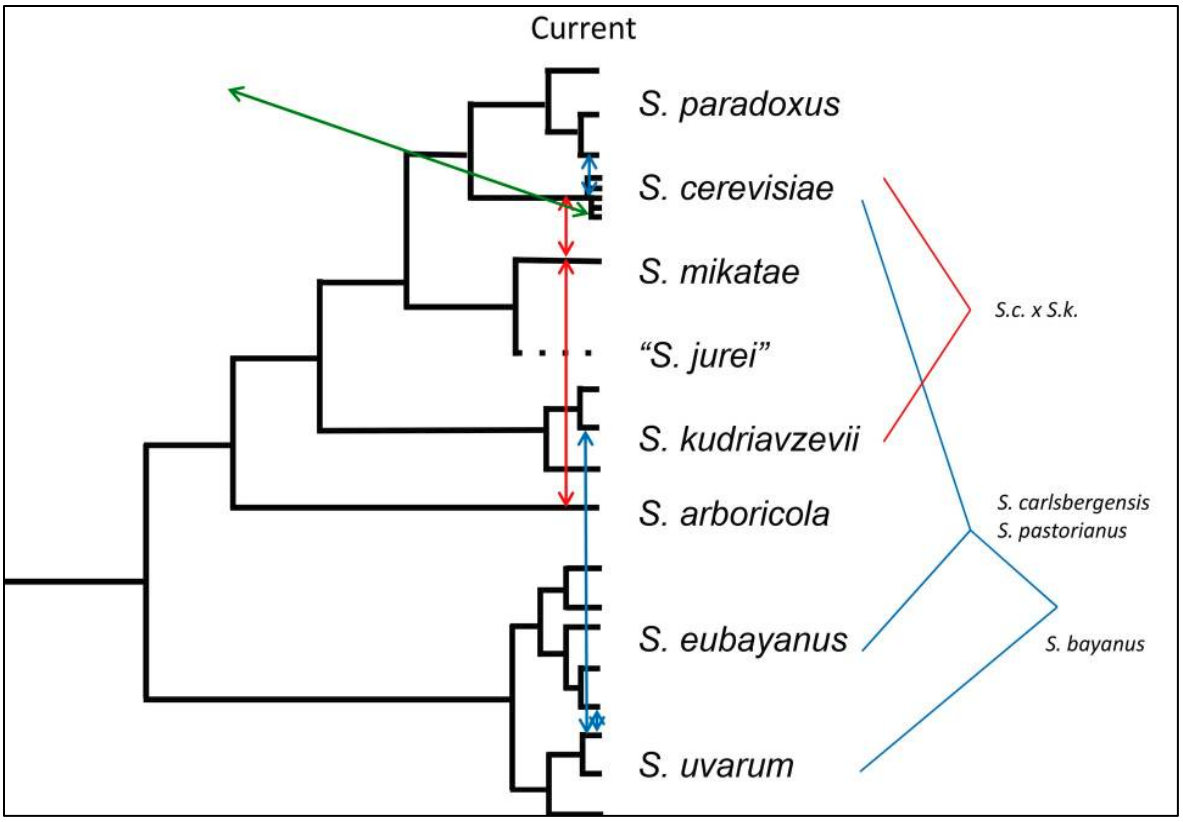


Figure i10. **The last *Saccharomyces* phylogeny released.** The new species *S. jurei*, a sister species of *S. mikatae* was recently incorporated to the genus.

4. Evolution of the physiological properties of yeasts that derived after Whole Genome Duplication (WGD)

During the evolution of the Saccharomycetaceae family, a historical and determinant event was **Whole Genome Duplication (WGD)**, which occurred in an ancestor lineage that diverged after the split of the *Zygosaccharomyces* genus (Figure i11A). Among the descendent linages, some species, including those of the *Saccharomyces* genus, went from having 16 chromosomes in a haploid state instead of 7 or 8 chromosomes of precedent yeasts lineages. The pre-WGD species belonged to the *Hanseniaspora*, *Eremothecium*, *Kluyveromyces*, *Lachancea*, *Torulaspora*, *Zygotorulaspora* and *Zygosaccharomyces* genera. The post-WGD components were *Vanderwaltozyma*, *Tetrapisispora*, *Nakaseomyces*, *Naumovia*, *Kazachstania* and *Saccharomyces*. Most *sensu lato* species were included in these last groups and were, thus, renamed according to the new ones (Kurtzman 2003). In recent years, evidence for WGD has been reported and explanations about the ancestor responsible for doubling have been hypothesized (Wolfe 2015; Marcet-Houben and Gabaldón 2015). The WGD hypothesis was strongly supported by the genome sequencing of some pre-WGD species, such as *Kluyveromyces waltii* (Kellis et al. 2004), where a correspondence was found for each region of this yeast to two *S. cerevisiae* regions. Older studies had already proposed a **polyploid** origin for post-WGD yeast based on histone gene loci studies (Smith 1987) by arguing that at least three (improbably) isolated duplication events for the core histone gene pairs would be necessary to give up them up. The point to elucidate this hypothesis was the genome sequencing of *S. cerevisiae* S288C in the 1990s (Goffeau 1996) from which up to 55 blocks of syntenic **paralogue** genes were detected (Wolfe and Shields 1997). Due to the technical consideration made to detect these regions (as a minimal of three pair of homologues, **BLAST**-P scores of ≥200, and the conservation of gene order and orientation), the total size of blocks was assumed

to be 50% of the genome. The mosaicism in the blocks distribution was explained by several reciprocal translocations, which blurred the original chromosomes size and, once again, **tetraploidy** was the supported hypothesis.

The model described tetraploid formation by the fusion of two diploid ancestral yeasts that led to the loss of the majority of the genome to become yet again a diploid cell (so the sequence identity and about 85% of duplicates were lost), followed by reciprocal translocations, and finally by maintaining the current diploid/haploid life styles of *Saccharomyces*. The question as to whether they were from the same or different species was left open. **Autopolyploidisation** happens when two diploid yeast cells from the same species are homologous for the MAT locus (α/α or a/a). Employing gene content and the gene order of the genome sequenced species, Gordon et al. 2009 showed the reconstruction of an ancient yeast genome that they assumed had undergone WGD, which thus suggested a model of autopolyploidisation from an eight-chromosome genome species (Figure i11B-C).

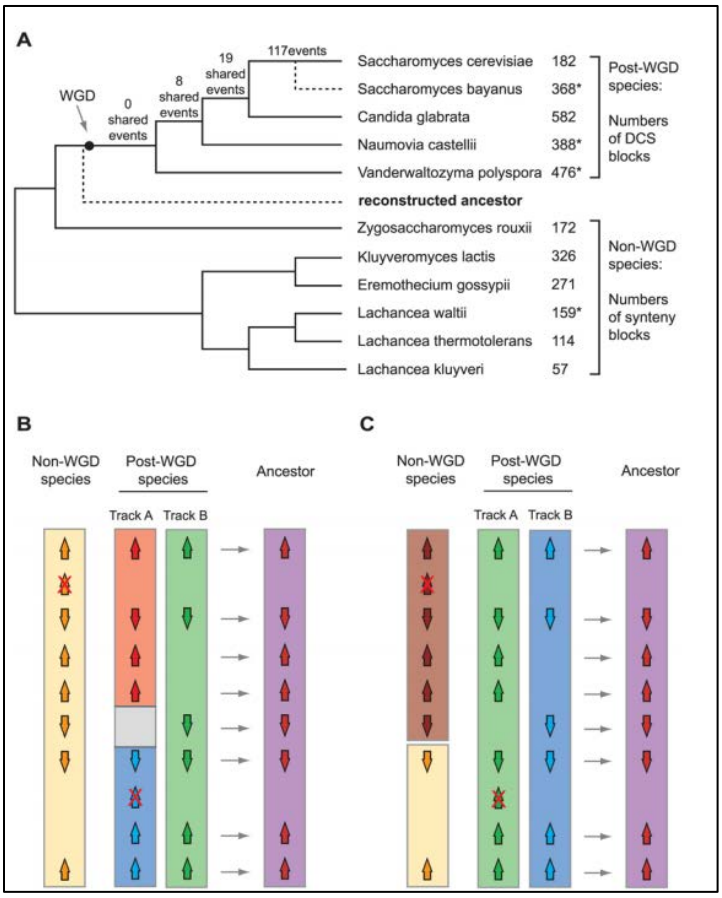


Figure i11. Saccharomycetaceae gene content and order. According to the autopolyploidisation hypothesis, an ancestor was inferred from gene content and the pre-WGD and post-WGD order (with two tracks of syntenic blocks or double conserved synteny (DCS). The reconstructed ancestor was predicted for the phylogeny branch before the *Vanderwaltozyma* genus linage. Two different gene rearrangement strategies were followed. In part B of the figure, a single break of synteny is on one track of the post-WGD species by arrangement after WGD, which thus infers an ancestor is carried out by the non-WGD track and the other track of the post-WGD. If a double break of synteny appears (C part of the figure) in the pre-WGD species, it therefore infers that it must be done by the two tracks of the post-WGD species. Inferring is possible only when a homologous gene appears on both post-WGD tracks or at least on one of them and in the pre-WGD track. Adapted from Gordon et al. 2009.

The case of **allopolyploidisation** was attributed to *S. pastorianus* (previously mentioned) as a possible hybrid between *S. cerevisiae* and *S. bayanus*, finally confirmed by the genome sequencing by Nakao et al. 2009. One possibility was the autopolyploidisation of a common ancestor of the post-WGD species.

In 2015, Marcet-Houben and Gabaldón employed a large-scale phylogenetic approach to show that **interspecies hybridisation** could probably be the underlying mechanism that triggered genomic doubling (a way to recovery fertility after hybrid formation). As a result, genome rearrangements by reciprocal translocations and massive gene losses reduced the genome to a minimal number of **ohnologous** genes (paralogous genes derived from WGD). The reconstruction of the evolutionary history for each gene in a given genome or **phylome** was employed for *S. cerevisiae* using two different strategies. One was for all the genes in the genome to include all the paralogous genes, and a second one using only *ohnologous* genes to avoid any bias shown in the whole phylome mainly for recently duplicated genes not related to WGD.

Based on the algorithm that detects duplication nodes (the two daughter branches sharing at least one species are defined as nodes), the highest duplication density was found in a branch (Figure i12A n3 branch) before the divergence between *Saccharomyces* and a group composed of *Kluyveromyces*, *Lachancea* and *Eremothecium* (KLE). Most of the duplications were mapped on this branch for both the whole phylome and the *ohnologous* phylome (Figure i12B) but also a second branch (Figure i12A n4 branch) corresponding to the *Zygosaccharomyces rouxii* and *Torulaspora delbrueckii* group (ZT) managed to accumulate some more duplications. Unexpectedly, a few duplications mapped on the expected WGD branch (Figure i12B n5 branch). At this point, the interspecies hybridisation hypothesis may be supported by performing the same analysis to the known *S.*

cerevisiae-S. eubayanus hybrid *S. pastorianus.* In this case, a duplication peak was mapped on the branch before the divergence of the two parental species, and not on the branch of the same species, as expected (Marcet-Houben and Gabaldón 2015). In the same way, for the *S. cerevisiae* linage formation, the species close to clades KLE and ZT could be the potential parental for interspecies hybridization (Figure i12C). In order to know which species could be the contributing ones, the authors decided to review all the trees that shaped the *S. cerevisiae* phylome to observe from which clade they came close to (ZT or Parental A and KLE or Parental B) and to generate new reduced phylomes to infer phylogenetic affiliations more accurately (Figure i12D).

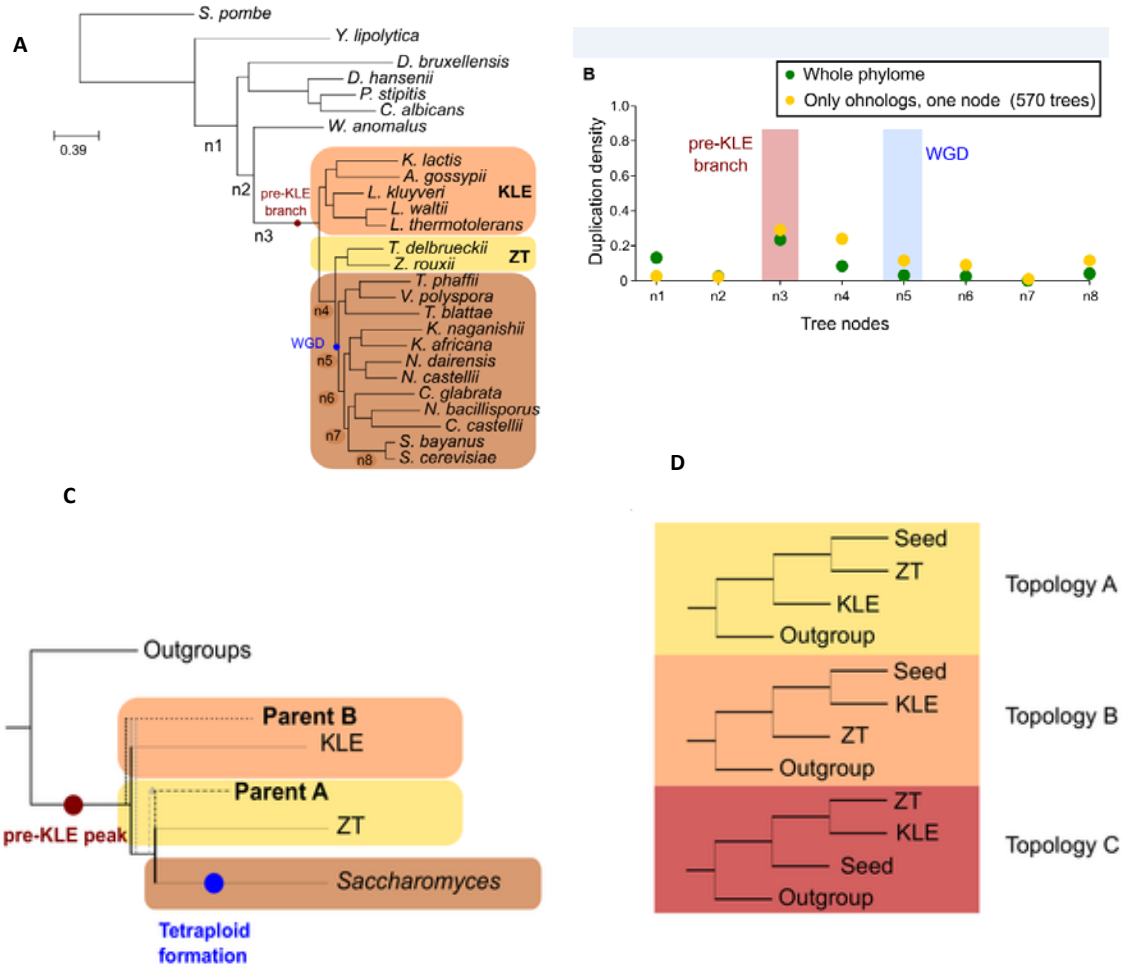


Figure i12. The phylome reconstruction for *S. cerevisiae*. A *S. pombe* rooted maximum likelihood tree (**A**) was built from 516 single-copy orthologous genes of the *Saccharomycotine* genome sequenced species. Various nodes were analysed to detect accumulation of duplications or duplication peaks (from n1 to n8). This duplication density is defined by duplication nodes, which are those nodes where the two daughter branches share at least one species. These duplication densities were calculated for all the duplicated genes, and also for genes assumed to come from WGD (ohnologuos pairs). In both analyses (**B**), a higher density was obtained for the pre-**KLE** branch (*Kluyveromyces*, *Lachancea* and *Eremothecium*) and the second one was located in the **ZT clade** (formed by *Zygosaccharomyces* and *Torulaspora*). Low duplication density was observed at the WGD node. Therefore, this suggests that the ancestors of WGD come close to these clades. From the results in A and B, a hypothesis of interspecies hybridisation between two parental species descending from the pre-KLE peak and corresponding to the KLE and ZT clades is resumed in part C of the figure above. Hybridisation is assumed to undergo a tetraploid formation to immediately recover fertility. In D, the possible topologies were compared to deduce the placement of post-WGD by taking the whole proteome or *ohnologous* only. The results of this comparison revealed that topology A was predominant for the *S. cerevisiae* proteins. Adapted from Marcet-Houben and Gabaldón 2015.

Schematically three different topologies are observed from the inspected trees: two are the parental and can come close to ZT or KLE (A or B, respectively) and one is a *S. cerevisiae* seed form a sister group (Figure i12D). The resulting analysis indicated that most topologies (60-82%) belonged to topology A, which agrees with the current *Saccharomycetaceae* family tree reconstruction (Figure i6) and this high percentage seems due to the possibility of partial or total gene conversion between *ohnologs*. Thus, a related lineage of the ZT clade would be involved in hybridisation (Parent A). For clade KLE, only between 4-14% genes were found and 14-28% of the genes formed a sister group in topology C. The conclusion drawn for Parental B was that it could have diverged before the KLE clade, thus no large association was found. Finally, collinearity was tested to ask the question about autopolyploidisation or hybridization, but the levels of similarity were too high to reject one of the possibilities.

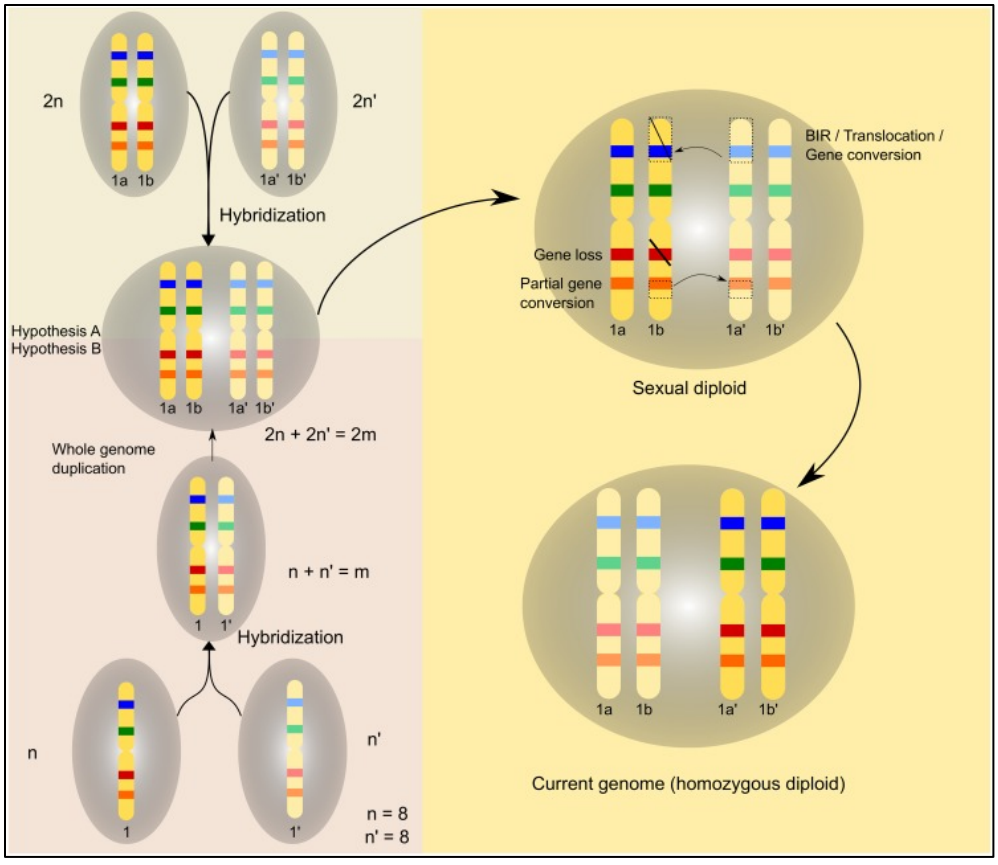


Figure i13. **The two interspecies hybridization hypothesis**. Two possible scenarios presented to explain the results of the observed peaks. With hypothesis **A**, two diploid species would form an allotetraploid that would undergo genome rearrangements and gene losses to finally render a linage with double the chromosomes. According to a second hypothesis (**B**), two haploid cells from different species would fuse and form an unstable allodiploid that can reproduce only asexually. Therefore, an additional duplication is needed to stabilise this formation, which is favoured by autopolyploidisation. Adapted from Marcet-Houben and Gabaldón 2015.

In conclusion, the presented model was a very ancient hybridisation between two different species. This hybridisation is difficult to detect because of the partial or total gene conversion, gene losses or translocations that have blurred parental genomes. So a phylogenetic analysis can slightly, but with good support, show from which clades these parental species could come close to. For the hybridisation event, two possible scenarios have been proposed (Figure i13): in the first one, the formation of an allotetraploid by two diploid different species happened which, after recombination processes and gene losses, gave rise to a genome with the double chromosome dotation. The second scenario implied two haploid species to form an allodiploid, and must necessarily be followed by autopolyploidisation to acquire a stable sexual cell cycle. The immediate consequences of whole genome duplication undergone in some yeast lineages concern the amplification over the genes presented in precedent or parental genomes.

As this genome increment is still observed in different degrees for the post-WGD species (Dujon et al. 2004), an immediate advantage of doubling is assumed to result. At least 55 blocks of the duplicated genes were maintained in *S. cerevisiae*, but after WGD many redundant genes were lost. As mentioned in the Cabtree effect, fermentative metabolism appears to be temporally related to WGD and a characteristic of all post-WGD species. Moreover, sugar signalling, transport and first metabolic processing genes are maintained in duplicate in important fermenter strains. Based on all these data, a hypothesis about the selective advantage that WGD could confer referred to the rapid growth of strains by an incremented glycolytic flux as a result of gene dose increases in genes like hexose transporters (Conant and Wolfe 2007). From this general hypothesis, three specific hypotheses were tested. The first was, after WGD gene loss, the glycolytic enzymes concentration raised, to become less diluted enzymes. The consequence of this led to the second hypothesis, which suggests that an increase in concentration means an increase in the fermentation flux. Finally, the third hypothesis was, once again, the result of the second one about the selective advantage of faster fermentation. Even in the presence of oxygen, *S. cerevisiae* prefers to transform glucose into ethanol via fermentation, despite this pathway being less efficient in producing ATP than respiration. However in the presence of large amounts of glucose, fermentation ensures that this species can grow faster than its competitors, and also adapts to support high ethanol concentrations, which is inhibitory for the majority of the microorganism.

The instant effect of increasing dose may not compare with the slow mutation accumulation in duplicated genes, from which gene novelties arose through sub-functionalisation and neo-functionalisation. Examples of the imposed fermentative metabolism are the alcohol dehydrogenase gene and the glucose signalling pathway. In the former, *ADH1* codifies for the enzyme that produces ethanol from acetaldehyde and the product of its WGD paralogue, *ADH2*, which is responsible for converting ethanol by an inverse reaction. Functional divergence is presented for the glucose-sensing proteins as *SNF3* senses low glucose concentrations while *RGT2* senses high concentrations.

5. From lab strains to industrial strains genome sequencing

The sequencing of *S. cerevisiae* laboratory strain S288c by Goffeau 1996 was the first eukaryote genome to be released and the largest at that time. They found over 5,885 codifying genes along a size of 12 Megabases (Mb) and only 4% of them contained introns. As mentioned on previous pages, the complete genome sequence revealed the existence of groups of genes that were strikingly similar to one another (what they called cluster homology regions), which was the key discovery to later hypothesise about possible **WGD**.

After this first yeast genome publication, more work on genome sequencing was carried out in later years. Genomes were obtained not only from strains from laboratories but from other sources. In this way, pathogenic yeast strains, such as *Candida albicans*, were also sequenced in 2004 by Jones et al.

Inside the *Saccharomyces* genus, Liti et al. 2009 did genetic variation association studies with phenotypical variation. A first sequencing approach, based on the **Sanger** methodology (First-generation sequencing), was run to obtain variability by SNP recompilation between *S. cerevisiae* and *S. paradoxus* strains of different origins. In order to easily analyse this nucleotide variation, monosporic strains were used to remove heterozygosity (which makes the analysis complicated) and genomes were assembled without having to perform any annotation.

A phylogeny analysis (Figure i14) revealed that *S. cerevisiae* strains presented some homogeneous or **clean populations** that were equidistant from one another, named by their source of isolation or origin as Wine/European, West African, Malaysian, Sake and North American. Apart from them, other strains were assigned as mosaics as they showed an admixture of the former populations because they presented mixed genomes as a result of outbreeding. No defined

structure population was found for the tested *S. paradoxus* strains and they all clearly diverged from *S. cerevisiae*, except for some Hawaiian strains. The ecological niches of most were reduced to *Quercus spp.*, so genome variations were due to geographical origins and some were reproductively isolated. Moreover, the *S. cariocanus* (previously considered another species of the *Saccharomyces* genus) strains were related to North American strains. These results correlated well with the phenotypic variation assays for each species; e.g. inside *S. cerevisiae* species, two groups were differenciated concerning to their growth rate. The *S. cerevisiae* Wine/European lineages and the **mosaic strains** presented a rapid growth compared to the other groups and an advantage in fermentation was suggested for them.

Recently, some strains under study by Liti et al. 2009, have been a whole genome sequenced by Second-generation sequencing (2nd generation) **Illumina** system (https://www.illumina.com/) and Third- generation sequencing (3rd generation) Pacific Bioscience (Pac-Bio, http://www.pacb.com/), with the latter providing some important advantages. Seven *S. cerevisiae* and five *S. paradoxus* strains of wild and domesticated origins were chosen for deep sequencing (Yue et al. 2017). The reliable **assembly** of producing long fragments was obtained (sequences continuously read by the sequencing machine), so complex genome regions, like those that carry repetitions or gene duplications, can be resolved. This is particularly important for telomeric and subtelomeric regions, for which anterior sequencing approaches have not been solved due to the presence of repetitive elements. Basically, two elements are observed in a strain-dependent copy number on telomeres: the core X and Y' elements. Albeit only one core X element is located close the chromosome end, it could be alone or followed by 1-4 Y' elements. Despite this assumption, new non-canonical chromosome ends were discovered thanks to this sequencing technology. In tandem repetitions of the core X element

(up to six copies) in either in *S. cerevisiae* and in *S. paradoxus*, and surprisingly some chromosome ends, carried one copy or more of the Y' element, but lacked the core X element, which is known to be important in genome stability.

In the same work, a first definition of a subtelomeric region is provided to define the chromosomal regions where more gene variability is observed which is related to an assumed advantage under some environmental conditions. Since then, sequencing did not provide enough resolution to be certain about where a subtelomeric region started and ended. Here, and based on a multi-genome comparison, it was suggested that "for each subtelomere, we located its proximal boundary on the basis of the sudden loss of synteny conservation and demarcated its distal boundary by the telomere-associated core X and Y' elements". Then by refusing the anterior definition of subtelomers that located them in an arbitrary 20-30 kb from the chromosome ends with no other criteria, this new approach to detect them was based on a gene synteny conservation profiles comparison of each chromosome end among 12 strains. Resolving subtelomeric regions has generated interest because it informs about the variability presented in a specific strain, which is related to the environments in which it lives. The genes found in these close-to-end zones are involved in response to stress and toxins, the metabolism of different compounds, and metal, amino acid and carbohydrate transporters (Brown et al. 2010). Concerning this last group, some genes belonging to different family genes like those from the **HXT**, **MAL** and **MEL** families, are frequently found. This is due mainly to the high frequency of recombination favoured by the repetitive elements in the aforementioned telomeres (which represents recent paralogy events). Sugar transport is a common function along these family genes and depends on the preferred source of carbon, and we can found more or less copies inside these groups. Members of the **HXT** (**H**e**x**ose **t**ransporter family) family are involved in the transport and signalling of glucose

and fructose (oligosaccharides found in wine fermentations), and at least 20 genes have been described in *S. cerevisiae* (Boles and Hollenberg 1997) of which eight are within or close to a subtelomeric position (*HXT15*, *HXT13*, *HXT12*, *HXT16*, *HXT8*, *HXT9*, *HXT17* and *HXT11*).

However, the most important *HXT* transporters that contribute to glucose and fructose ingestion are found in the core region of the chromosomes, a secondary transport of different sugars has been found in some subtelomeric transporters. For example, *HXT13*, *HXT15*, *HXT16* and *HXT17* transport mannitol and sorbitol with moderate affinity, and xylitol transport is mediated by *HXT11* and *HXT15* (Jordan et al. 2016). The transport of maltose and melibiose implies MAL and MEL families, respectively. Numerous copies of MAL family genes are found in brewing strains because maltose is the predominant sugar in beer. Inside this family, three families are differentiated according to their function (Brown et al. 2010): MALR (regulators), MALT (permeases) and MALS (maltases). From each one, there are gene subfamilies which contain duplicated genes in some strains (paralogous) that have evolved for a particular substrate specify distinct to maltose. One example is in the MALR family, where the MAL63-like clade is composed of gene copies that enable growing, not only in maltose, but also in other carbon sources, such as maltotriose, turanose, isomaltose, sucrose, palatinose and me-alpha-G. Nevertheless, the *MAL13*-like clade lacks the former activities except the one for palatinose as it seems that this copy has evolved to process the sugar found in sugarcane and honey.

Although it was assumed that wine yeasts genomes would harbour genetic content that could explain their adaptation to the fermentative environment since 2009, only comparative genomic hybridization information in wine yeast was available to suspect it. The complete genome sequencing of the *S. cerevisiae* EC1118 wine strain finally revealed particular traits of wine yeast that had never been seen before. The most relevant result of the paper by Novo et al. 2009 was the discovery of three new gene regions (Figure i15A): Region A (38kb), Region B (17kb) and Region C (65kb). They found that the genes that contained these regions were involved in winemaking-related functions, such as carbon and nitrogen metabolism (Figure i15B).

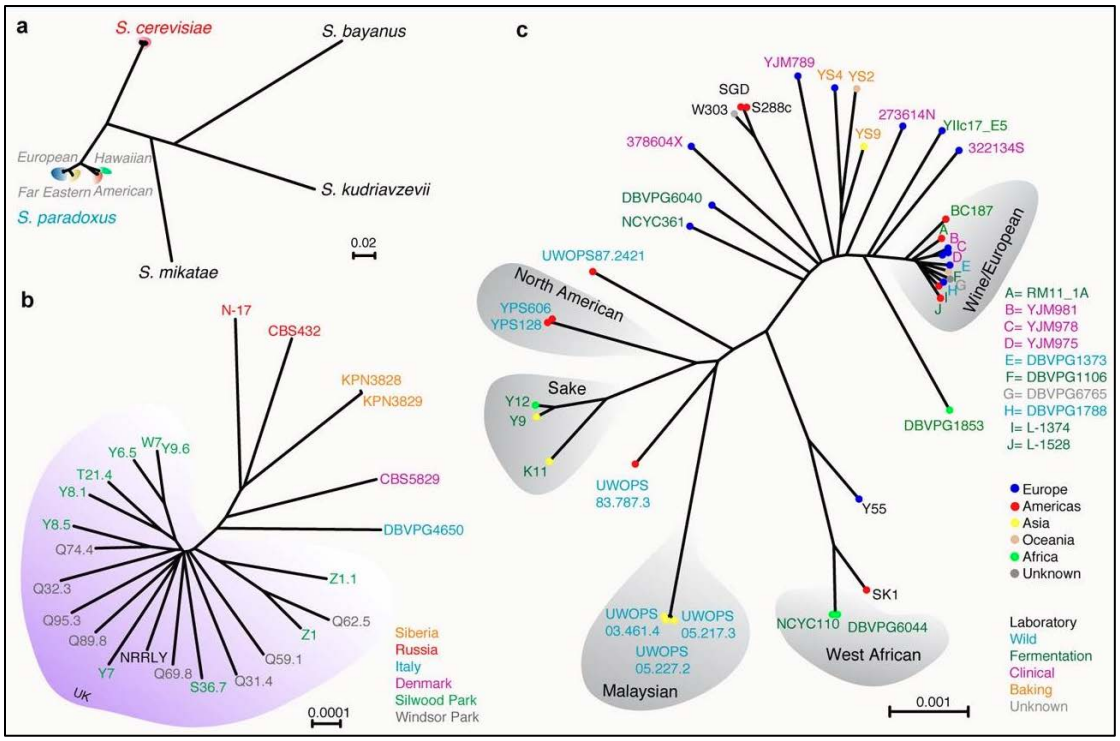


Figure i14. **Populations of *S. cerevisiae* and *S. paradoxus* strains of different origins.** *S. bayanus*, *S. mikatae* and *S.kudriavzevii* are the output species. The trees obtained from the SNP differences clearly separate the *S. cerevisiae* and *S. paradoxus* populations. The *S. cerevisiae* subpopulations, like that from Sake or Wine European, reflects the idea of the heavy weight of domestication by clustering strains into their particular fermentation more than their geographical origin. Adapted from Liti et al. 2009.

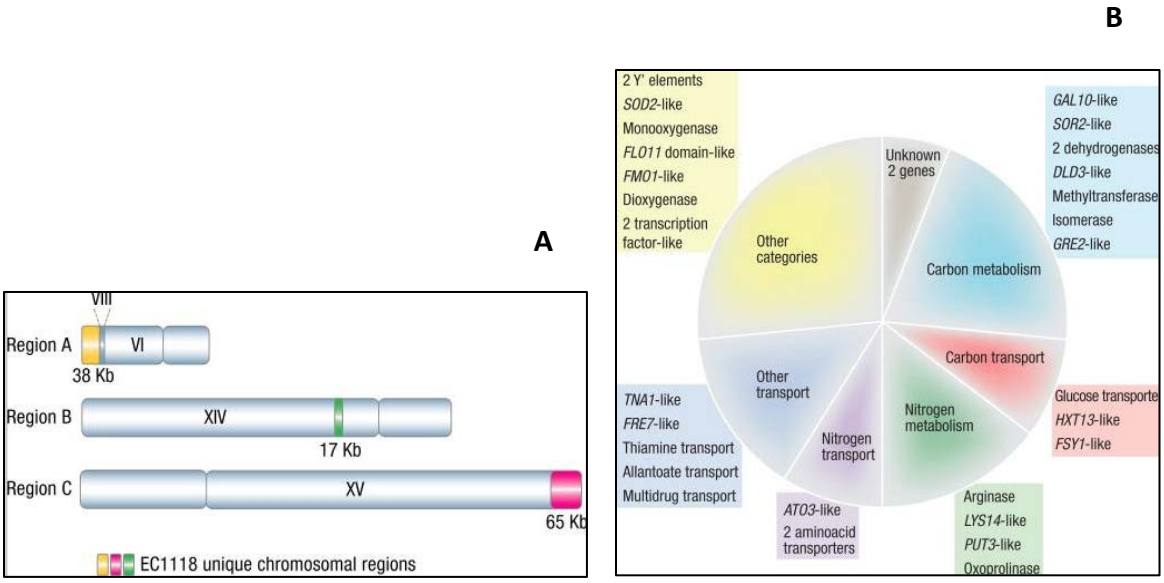


Figure i15 .Discovery of three new genomic regions never seen before in *S. cerevisiae*, found specifically in the *S. cerevisiae* wine strain EC1118. These two subtelomeric and other central regions were rich in the genes related to fermentative processes like those involved in sugar transport (*HXT13* or *FSY1*) or metabolism flow (*GAL10* or *SOR2*), nitrogen assimilation (*PUT3* or *ATO3*) and other important genes for the cell surveillance. Adapted from Novo et al. 2009.

To find out the possible species donor of these three regions, PCR amplification was performed among 77 *Saccharomycetaceae* species. Amplification was only obtained only for Region B in the *Zygosaccharomyces bailii* CBS 680 type strain. For Region A and Region C, it was concluded that an unknown species, which differed for each region, would be the contributor of the region. A synteny analysis concluded that Region C could come from a *Saccharomyces* species.

6. Particularities of the *Saccharomyces* species during fermentation and their importance in industry

As described before, the metabolic processes carried out by *S. cerevisiae* during fermentations are settled in an aerobic-fermentative metabolism. Both glucose and fructose sugars are present in equal quantities in grape must, and could be incorporated by multiple hexose transporters (HXT) that act as passive diffusors, but with a higher affinity for glucose. However, preference for glucose is presented by wine yeast (known as the *glucophylic* character of *Saccharomyces* wine yeast), which means that fructose is consumed less during fermentation by creating a discrepancy between both saccharides (Berthels et al. 2004). This inherent preference could imply a problem for the final wine product because fructose could lie residually and then produce excessive sweet wines with the additional problem of more final ethanol, which would expose wine to the risk of bacterial contamination.

It has been shown that fructose utilisation is inhibited more severely than glucose by high ethanol conditions, and increases in nitrogen-supplied must. It has been proposed that the distinction of consuming glucose more than fructose is determined mainly by the first sugars fermentation stages. In this way, hexose transporters' affinities and, after being inside the cell, the phosphorylation of sugar by hexokinases seems mainly responsible. These sugars consumption differences appear as a problem for wineries, so it is important to look forwards for yeasts with "fructophilic" behavior. Of all the new genes found in the named Region C of *S. cerevisiae* EC1118, one gene was found that codified for a high-affinity active fructose transporter: ***FSY1***. *S. cerevisiae* possess on their surface a huge number of hexoses transporters, which work only by facilitated diffusion, and no active transport was already known in it until the work of Novo et al. 2009 was published. Knowledge about *FSY1* emerged in 1990 when Rodrigues de Sousa H et al.

investigated fructose/H^+ symport activity in some strains of the ancient *Saccharomyces sensu stricto* classification. They tested four strains of *S. pastorianus* and five of *S. bayanus* (currently invalid assignment), and they all presented activity for transport. Five *S. cerevisiae* strains and three strains of *S. paradoxus* did not show this phenotype. In 1994, Tornai-Lehoczki J. *et al.* redid this analysis, but with more strains of those species, and exactly 66 strains were isolated from Hungarian wine and beer. As expected from the results of Rodrigues de Sousa *et al.* 1990, *S. cerevisiae* (50 of the total) and *S. paradoxus* did not present any activity and, as expected, all the tested *S. bayanus* and *S. pastorianus* were positive. This shows that these authors probably found a physiological trait to differentiate *Saccharomyces* species. In fact *FSY1* sequences are employed for current *S. eubayanus* and *S. uvarum* pure and hybrid species differentiation. The first isolation of the *FSY1* gene was done from a genomic library construction of *S. pastorianus* PYCC 4457. Activity with fructose was assayed and gene sequence was firstly described. It consisted in a high-affinity fructose/H^+ symporter (K_m=0.2 mM) of a 570 amino acids-coding sequence (Rodrigues de Sousa et al. 2004). A homology (Figure i16) analysis revealed that it was a transporter with no striking similarity with others known from the Major Facilitator Superfamily (Gonçalves et al. 2000).

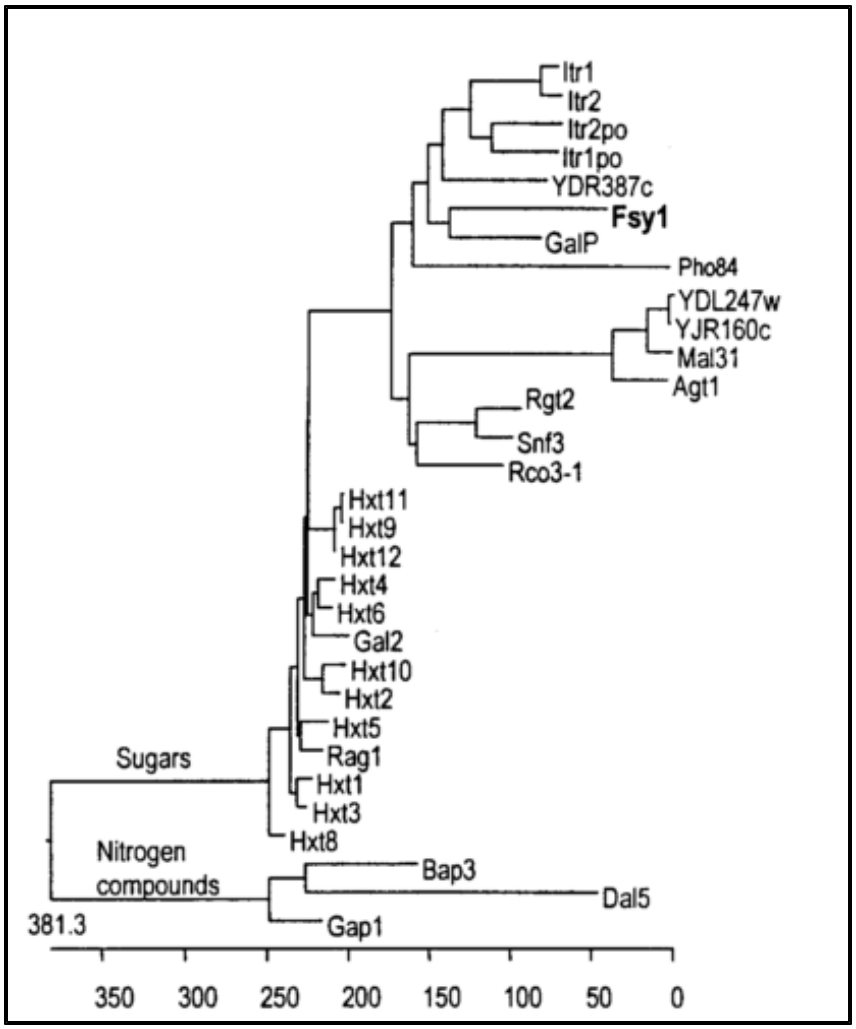


Figure i16. The Fsy1 and sugar permeases relationships. Dendrogram showing the relationships between the permeases of the Major Facilitator Superfamily (MFS) and Fsy1 of *S. carlbergensis* (the *S. pastorianus* Saaz-type strain), the first sequence obtained for this gene. *S. cerevisiae* hexoses transporters cluster far from that of Fsy1, whose amino acidic sequence is similar to *S. cerevisiae* Itr1, Itr2, YDR387c and Pho84 and those from other species (Itr1po and Itr2po from *S. pombe*; GalP from *E.coli*). Adapted from Gonçalves et al. 2000.

Fructose symporter homologous sequences were discovered in other non-*Saccharomyces* yeasts. The *Kluyveromyces lactis FRT1* gene also presented also high affinity to fructose transport (K_m=0.16 mM) and was related directly to the

sequence of *S. pastorianus* (Diezemann and Boles 2003). Another specific fructose transporter was found in *Zygossacharomyces bailii* (*FFZ1*) but consisted, in this case, in a facilitator protein with K_m=80.4 mM, and with specificity for fructose, but not for glucose (Pina et al. 2004). Two fructose uptake systems, which formed a phylogenetic cluster together with *Z. bailli FFZ1*, were discovered in *Zygosaccharomyces rouxii*: ZrFFZ1, a high-capacity fructose facilitator (K_m=400 mM) and ZrFFZ2, a second facilitator that transports both glucose and fructose with similar capacity and affinity (Leandro et al. 2011). Interestingly in the grey mold *Botrytis cinerea* which is frequently found in vineyards, a high-affinity fructose symporter (K_m=1.1 mM) was found. Clustering revealed that it was related to those of *S. pastorianus* and *K. lactis* (Doehlemann et al. 2005).

The gene regulation of *S. pastorianus* PYCC 4457 and *S. bayanus* PYCC 4565 *FSY1* was analysed. *FSY1* expression was higher in the presence of small amounts of fructose and was repressed in the cells inoculated in glucose media. Catabolite repression by glucose was suggested as the regulator mechanism, as supported by the presence of Mig1 (a transcriptional repressor) binding sequences in the *FSY1* promotor. The relative expression of the *FSY1*-GFP constructions of both strains revealed that the *S. bayanus* transporter was equally repressed at 0.1% of glucose as *S. pastorianus* at 0.5% of the same sugar, with more resistance shown for the latter. Finally, the de-repression of the system was carried out by shifting in media with low glucose or fructose concentrations, which also induced the expression of the system (Rodrigues de Sousa et al. 2004). In contrast, the regulation of *K. lactis FRT1* was not influenced by catabolite repression (Diezemann and Boles 2003), but by a large amount of glucose or fructose, where glucose was more efficient in repression.

The *FSY1* kinetics of found in the EC1118 genome were characterised in Galeote et al. 2010. It also presented high affinity for fructose (K_m=0.24 mM) and regulation showed repression by high glucose or fructose concentrations and by high expression on ethanol as a sole carbon source. The phylogenetic reconstruction with the other fructose symporter clustered it with the monophyletic group formed by *S. pastorianus* and *S. uvarum FSY1* (Figure i17).

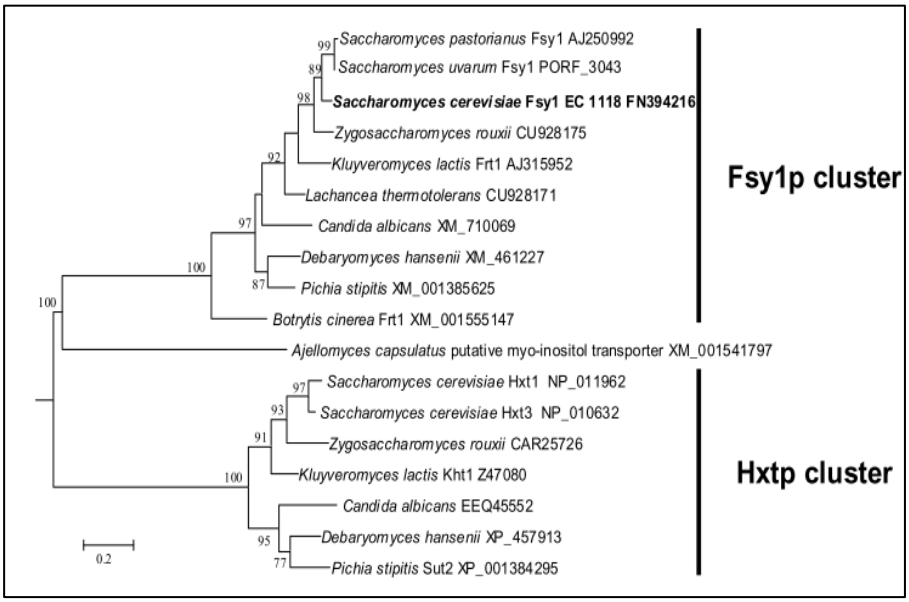


Figure i17 .**The Fsy1p and Hxt phylogenetic reconstruction**. A ML tree of fungal transporters showed two separate cluster, one of those protein sequences of Fsy1 and the other formed by the some hexose transporters from the same species. The position *of S. cerevisiae* EC1118 Fsy1 suggests a similar function to the other *Saccharomyces* strains. Adapted from Galeote et al. 2010.

Finally, a search of *FSY1* along the *Pezizomycotina* and *Saccharomycotina* sub-phyla of Ascomicota resulted in 109 homologous and 10 independent sequences, probably the horizontal gene transfer (**HGT**) of this gene where shown between different species (Coelho et al. 2013).

Objectives

The origin of the *FSY1* gene and the remaining Region C genes was revealed in the work of Marsit et al. 2015 in collaboration with our group thanks to the analysis presented in this thesis work. The genome sequencing of *Torulaspora microellipsoides*, a pre-WGD species, showed that this species harboured not only the *FSY1* gene but a strenght of genes similar to the Region C genes with an additional 80 kb of genes between the *ARB1* and the *PUT3* genes. Besides, two *S. cerevisiae* strains sequenced there one from food and another one from beer presented a region of genes more similar to *Torulaspora microellipsoides* than to the Region C found in the EC1118 strain and others. The discovery of this **introgression** denote the relevancy of a pre-WGD species, *T. microellipsoides*, on its contribution to the genome evolution of a *S. cerevisiae* lineage, a post-WGD species. Concerning to the genes that were probably acquired from this species a role in adaptation to fermentative environments was suggested. Additionally, an interspecific hybridization between two distantly yeast species was proposed to explain the introgression event what suggests that this kind of breeding is more frequent than it was though. Moreover, the WGD hypothesis presented in Marcet-Houben and Gabaldón 2015 supports this kind of hybridization.

In light of the results presented in the Introduction section, *T. microellipsoides* seems to be an important contributor to, at least, *S. cerevisiae* domesticated species, so we decide to carry out a deeper study on this species which was poorly studied before and only the genome sequences of the type strain was available. In this way, we state the following objectives for this thesis work while it is important to remark that new objectives have appeared during the arising of the results and they are stated in each chapter:

1) Dilucidate the evolutionary history of the *FSY1* gene mainly due to their importance in the possession of a “fructophilic” behaviour in some industrial yeast.

2) Figure out the probably donor species of the *FSY1* gene to *S. cerevisiae* wine strains.Once we deduced that *T. microellipsoides* was the possible species responsible of the transference of the complete Region C:

3) Clarify the Region C direction of transfer, it means from *T. microellipsoides* to *S. cerevisiae* or in the other direction by performing the genome sequencing of the available *T. microellipsoides* strains.

Chapter 1:

Active fructose symporter and ammonia exporter acquisition by introgression events in the *Saccharomyces* genus

1.1 Abstract:

Acquisition of *S. cerevisiae* EC1118 Region C was already unrevealed in Marsit et al. 2015, in a collaboration with us. Here we pointed out the non-*Saccharomyces* species, *T. microellipsoides* as the donor of this novel region. Although all newly discovered genes were probably related to improve wine fermentation (Novo et al. 2009), we especially focused on the origin of *FSY1*: a gene that codifies for a fructose/H^+ symporter expressed at low fructose concentrations. Harbouring it may imply an important industrial trait if the problem of unfinished wine fermentations, due to incomplete fructose consumption in some cases, could be reduced. *S. cerevisiae* possess on their surface a huge number of hexoses transporters that work by only by facilitated diffusion and no active sugar transport was described until the work of Novo et al. 2009 was published. Previous publications have shown that this gene was present in the genome of the cluster composed of *S. eubayanus*, *S. uvarum* and *S. pastorianus* but was absent in the other *Saccharomyces* species. Such energy-consuming transport is typically found in species with a high respiro/fermentative ratio, which is a characteristic metabolism of some non-*Saccharomyces* yeasts.

In order to find out the donor species of *FSY1* (all Region C) to *S. cerevisiae*, we searched for this transporter in numerous *Saccharomyces* and non-*Saccharomyces* strains along the Saccharomycetaceae complex (C. Kurtzman 2003). After retrieving the known *FSY1* sequences from databases, we started to amplify newly sequences from available species. Sequencing the *T. microellipsoides* amplicon showed that it possessed two divergent copies of this gene in its genome. The phylogenetic reconstruction revealed that each homologous copy was related to one *Saccharomyces* species, one to *S. cerevisiae* and the other one to the *S.*

eubayanus, *S. uvarum* and *S. pastorianus* group. Then the topology obtained for the tree did not correspond to the topology expected for the yeast family, so we tested the significance of this tree topology using a Simodaira-Hasewaga comparison. The distribution of the *FSY1* gene inside the genome sequenced species, and events occurred until its transfer was described to support the obtained results. For the *FSY1* gene, we proposed two independent gene transfer events to each *Saccharomyces* species.

The PCR amplification of the Region C genes and posterior genome sequencing of *T.microellipsoides* confirmed the presence of the complete region in this species. As with *FSY1*, we wondered if these genes (or copies of them) could be transferred to other *Saccharomyces* genus species. We performed a blast search for yeast sequenced genomes from databases and our lab. As a result, we found a homologous copy of the *ATO3* gene in five *Saccharomyces* species and more than one copy was detected in some strains thanks to new ***PacBio*** technology assemblies. The ancestral function described for its homologous YDR384C gene was an ammonia transporter to outside of the cell, but a specific function is yet to be described. The phylogenetic reconstruction showed a monophyletic cluster composed of the new copy found in the *Saccharomyces* species that diverged first from the group of *T. microellipsoides* and *S. cerevisiae* Region C-related sequences. We hypothesised that if no *T. microellipsoides*, a close unknown species, could be the donor of this gene to these *Saccharomyces* species. It would probably be an ancient transfer as the majority of the species present this gene with a high nucleotide divergence from *T. microellipsoides*.

1.2 Results:

1.2.1 Presence or absence of the *FSY1* gene in Saccharomycetaceae family: the Horizontal Gene Transfer (HGT) hypothesis

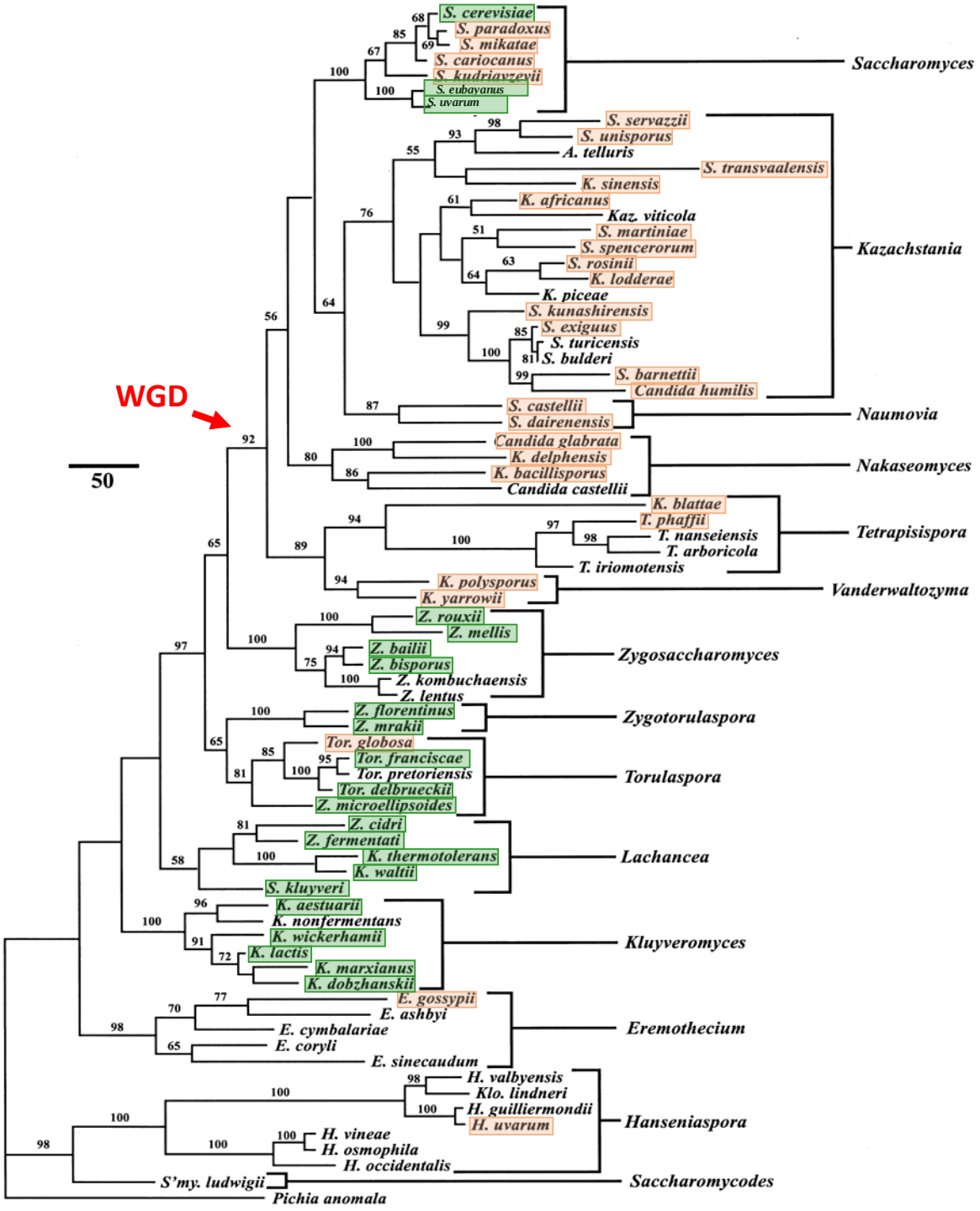


Figure 1.1. The PCR amplification of a 903 bp central fragment of *FSY1* in the *Saccharomycotina* complex. The positive and negative amplification is respectively shown in the green or red box and the species not included in the analysis were not color-squared. The *FSY1* gene was present in the majority of the pre-WGD species tested for genera *Kluyveromyces*, *Lachancea*, *Torulaspora*, *Zygotorulaspora*, *Zygosaccharomyces*. However, no amplification resulted from the post-WGD species tested from genera *Vanderwaltozyma*, *Tetrapisispora*, *Nakaseomyces*, *Naumovia*, *Kazachstania* except inside the *Saccharomyces* genus. The figure is adapted from Kurtzman *et al*., 2003 with the *Saccharomyces* genus updated according to the current classification. The WGD event is denoted with a red arrow.

The *FSY1* gene was searched in the "*Saccharomyces* yeast complex" to find out the donor species of this gene to *Saccharomyces cerevisiae* strains. The complete sequence of this gene (1794 bp) was first obtained from the available genomes in databases (see Methodology 1.3.2). From those sequences, we designed a pair of general primers that amplified all the species included in this yeast family (see Methodology 1.3.3). As the Carboxi and N-terminus of the gene were quite variable, amplicon size was reduced to 903 bp (which meant about 50 % of the ORF). A screening of 206 strains was performed for this study, but we obtained only amplification of 67 of them (Strains Table 1). A schema of the positive and negative results is shown in Figure 1.1.

The majority of the pre-WGD genera strains presented this gene in their genome, except *Torulaspora globose* that did not present any amplification. For some strains of species *Zygosaccharomyces rouxii*, *Kluyveromyces lactis*, *Lachancea thermotolerans*, the complete gene sequence was previously found in databases. PCR amplification was obtained in all, or in some, species of the *Kluyveromyces*, *Lachancea*, *Torulaspora*, *Zygotorulaspora*, *Zygosaccharomyces* genera. Negative results were obtained only for three tested species (*T. globose*, *E. gossypii* and *H. uvarum*) and some species were not included in the analysis (Figure 1.1). The sources of isolation of the 31 positive pre-WGD strains were soil, *Quecus spp.* trees, in symbiosis with *Drosophila,* and from a fermentative lifestyle.

Concerning the post-WGD genera, only 16% of the tested strains (26 of the 162 strains) presented this central sequence of the *FSY1* gene (Strains Table 1). No amplification was obtained for the *Vanderwaltozyma*, *Tetrapisispora*, *Nakaseomyces*, *Naumovia* and *Kazachstania* species. Nevertheless, positive results were obtained for some *Saccharomyces* species: inside the *S. eubayanus*/*S. uvarum* and *S. pastorianus* clade, 100% of the tested strains presented the gene, but we looked for the fructose transporter in *S. cerevisiae*, only ten of 61 were positive and most were wine strains (Strains Table 1).

1.2.2 *FSY1* phylogeny I: **maximum likelihood** (**ML**) tree reconstruction of the *FSY1* partial gene sequence

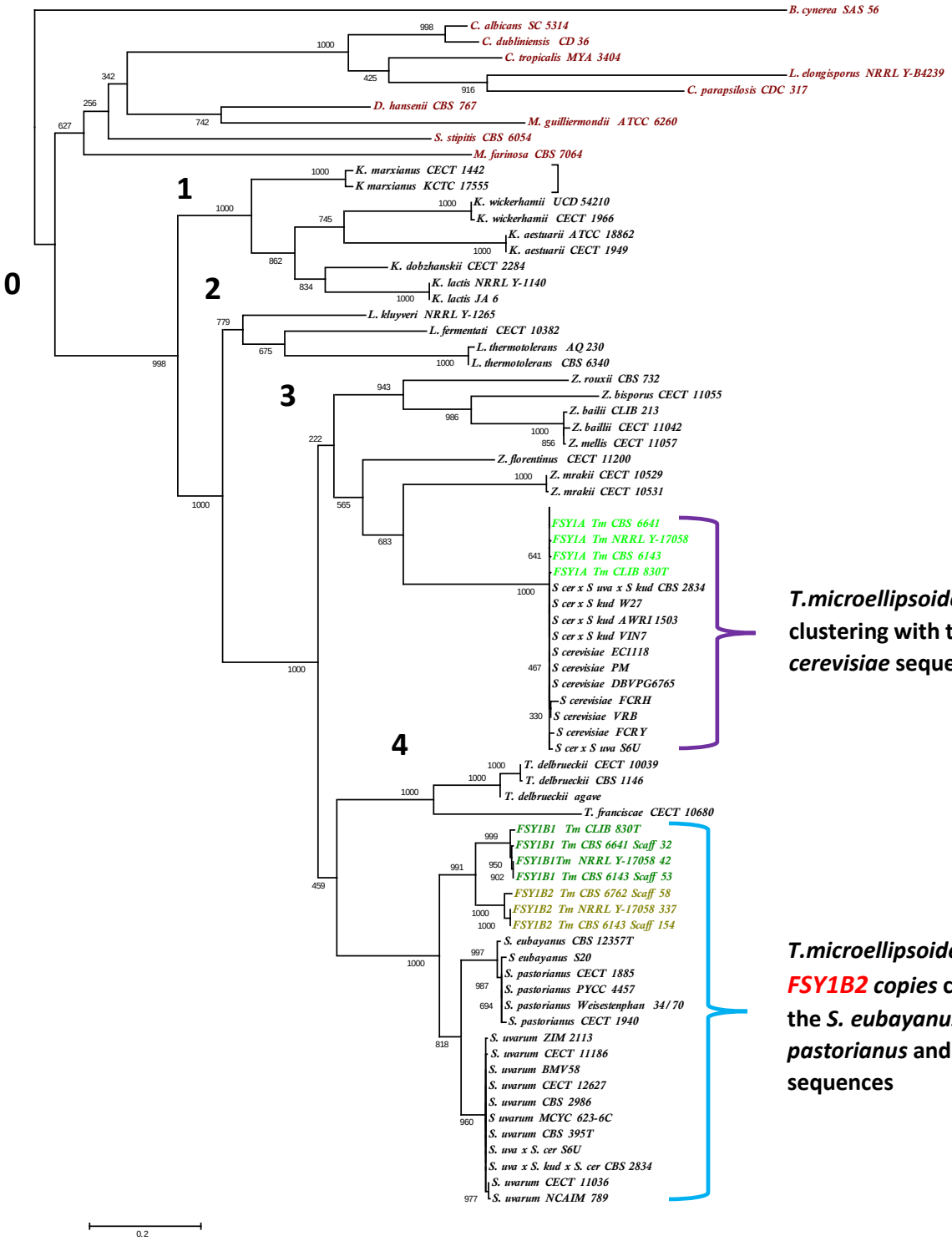


Figure 1.2. The ML phylogeny obtained with a 903 bp *FSY1* sequence of *Saccharomycetaceae* species. Outgroup strains are shown (in red) at the top of the tree and clustered from the **0** Species: name and strains numbers refer to their collection or origin precedence. Genera are represented by: **1**-*Kluyveromyces*; **2**-*Lachancea*; **3**-*Zygosaccharomyces* and **4**-*Torulaspora*. Different tones of green highlighted the *FSY1* copies found in the *Torulaspora microellipsoides* strains (*FSY1A*, *FSY1B* and *FSY1B2*). A purple key, indicate the group formed by the *FSY1A* and *S. cerevisiae* sequences, and a blue key denotes that composed of *FSY1B*/*FSY1B2* with the other *Saccharomyces* species. The numbers at nodes correspond to the bootstrap values over 1000 bootstrap pseudo-replicates. The scale is given in nucleotide substitutions per site.

The phylogenetic reconstruction with a central region of the *FSY1* gene (Figure 1.2) was performed to discover possible donor species. The outgroup species seen at the top of the tree were the grey mold *Botrytis cinerea* and *Candida*, *Debaryomyces*, *Meyerozyma,* and other yeast species. The outcome reconstruction showed that species *Kluyveromyces*, *Lachancea*, *Zygosaccharomyces* and *Zygotorulaspora*, maintained their phylogenetic position inside the yeast family tree. Therefore, the sequence of this gene fragment seems well conserved, after differentiating each genus with **bootstrap** values of 1000.

When we looked at the *Torulaspora* species in the resulting tree, we immediately see that *Torulaspora microellipsoides* ocuppies an unexpected position compared to the other pre-WGD genera. Given these results, we thought that the phylogeny incongruences showed that *T. microellipsoides* could probably be the donor of at least the *FSY1* gene, and perhaps of the complete Region C (confirmed in Marsit et al. 2015). Before the tree reconstruction, we found that the *T. microellipsoides* species presented more than one sequence for this gene. We also found these copies first by the PCR amplification and Sanger sequencing of the

FSY1 central region in the type strain and then by the genome sequencing of the available *T. microellipsoides* strains (see Chapter 2).

The phylogenetic analysis revealed that one of the *FSY1* copies, which we called *FSY1A*, grouped together with those sequences found in some *S. cerevisiae* strains inside Region C with a bootstrap value of 1000. The *Zygotorulaspora* species included in this analysis (*Z. mrakii* and *Z. florentinus*) grouped at with this monophyletic group the same time.

Two additional copies of *FSY1A* were found in *T. microellipsoides*. The one we called *FSY1B* was in the genome of CLIB830T, NRRL Y-17058, CBS 6143 and CBS 6641. The other one was probably a recent paralogous copy of *FSY1B* as it was located in a homologous chromosomic region. We named *FSY1B2* at this second sequence as we first found *FSY1B* in the type strain. It was present in the NRRL Y-17058, CBS 6143 and CBS 6762 strains. Details of the chromosomic context of *FSY1B*/*FSY1B2* are provided in Supplemental Figure 2 and better discussed in the next Chapter. Both sequences clustered with *S. eubayanus*/*S. pastorianus* and the *S. uvarum* clade (bootstrap value of 1000). The other *Torulaspora* species included in the phylogeny (*T. delbrueckii* and *T.franciscae*) were the closest to this group.

The immediate conclusion of these analyses was the discovery of what was probably donor species of the *FSY1* gene, *T. microellipsoides*. Hence, this species was responsible for the transfer to some *S. cerevisiae* strains, but we accidentally found that, probably, another copy of this gene could be transferred to the *S. eubayanus*/*S. pastorianus* and the *S. uvarum* clade by the same species.

When we looked at the nucleotide level of each *FSY1* copy in the *Saccharomyces* species (Figure 1.3), we determined that they were quite different for *FSY1A* and *FSY1B*/*FSY1B2*. *S. cerevisiae FSY1* and its homologous copy *FSY1A* from *T. microellipsoides* CLIB830T, presented a 100% homology and only a few

changes in the other strains (three non-synonymous amino acid changes). However, between brewery species, the *FSY1* sequences and *FSY1B*/*FSY1B2* paralogous copies with more nucleotide changes accumulated compared to one another (14.8% on average).

		1	2	3	4	5	6	7
1	***S. cerevisiae*** _EC1118							
2	***S. cerevisiae*** _CBS_7957	0						
3	***S. cerevisiae*** _CLIB_382	0	0					
4	***FSY1A*** _CLIB_830T	0	0	0				
5	***FSY1A*** _NRRLY-17058	8	8	8	8			
6	***FSY1A*** _CBS_6143	8	8	8	8	0		
7	***FSY1A*** _CBS_6641	9	9	9	9	3	3	

		1	2	3	4	5	6	7	8	9	10
1	***S. uvarum*** _CBS_395T										
2	***S. eubayanus*** _CBS_12357T	130									
3	***S. pastorianus*** _Weisestenphan 34/70	134	20								
4	***FSY1B*** _CLIB_830T	268	275	281							
5	***FSY1B*** _NRRLY-17058	268	272	278	14						
6	***FSY1B*** _CBS_6143	268	272	278	14	0					
7	***FSY1B*** _CBS_6641	268	272	278	13	5	5				
8	***FSY1B2*** _NRRLY-17058	**248**	257	261	154	156	156	155			
9	***FSY1B2*** _CBS_6143	**248**	257	261	154	156	156	155	0		
10	***FSY1B2*** _CBS_6762	**252**	265	269	158	160	160	159	36	36	

Figure 1.3. The nucleotide distance between *Saccharomyces* and the *T. microellipsoides FSY1* complete gene sequence. A distance matrix for aligned 1794 bp was obtained with the Mega *v*6.0 software (Tamura et al. 2013). **Upper panel**: three *Saccharomyces cerevisiae* strains compared with FSY1A from the *T. microellipsoides* strains. **Bottom panel**: one representative of *S. eubayanus*/*S. pastorianus* and *S. uvarum* compared to the *FSY1B*/*FSY1B2 T. microellipsoides* sequences. Relevant changes in both panels are squared or highlighted in capitals.

1.2.3 *FSY1* phylogeny II: maximum likelihood (ML) tree reconstruction of an *FSY1* complete gene sequence

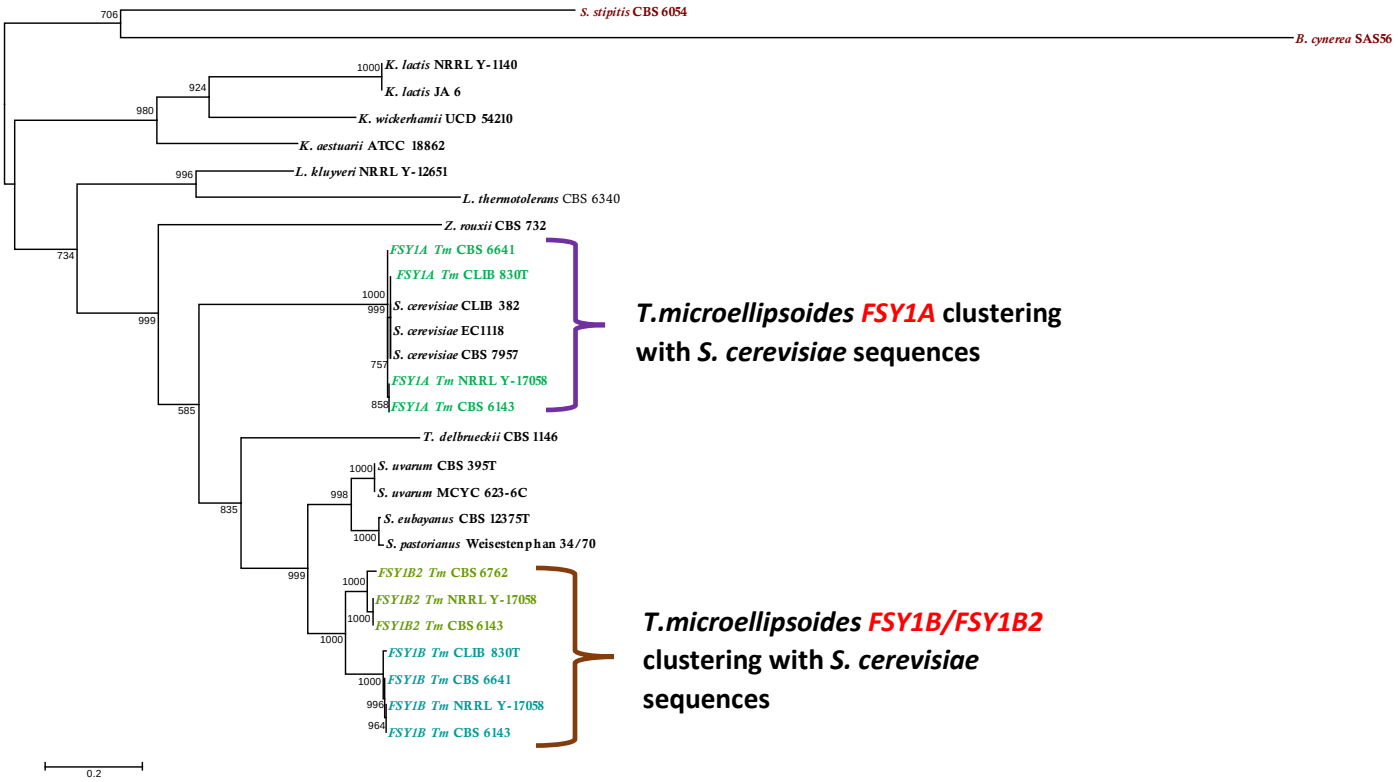


Figure 1.4. The ML phylogeny obtained with a 1794 bp the *FSY1* complete sequence of the *Saccharomycetaceae* species. *S. stipites* and *B. cynerea* are the outgroups. Species' name and strains number refer to a yeast collection or origin. Different tones of green highlight the *FSY1* copies found in the *T. microellipsoides* strains. Numbers at nodes correspond to the bootstraps values over 1000 bootstrap pseudo-replicates. The scale is given in nucleotide substitutions per site.

The phylogeny with a complete sequence of the *FSY1* gene was carried out to gain more support for the previous reconstruction with a partial gene sequence (Figure 1.4). However, fewer sequences were used for this study given the aforementioned problems.The outgroups species were *S. stipites* and *B. cinerea*, and the latter had a longer branch length. The pre-WGD genera included in the tree, *Kluyveromyces*, *Lachancea* and *Zygosaccharomyces*, maintained their expected phylogenetic position, and *Z. rouxii* was the closest species to group with the *FSY1 S. cerevisiae*/*FSY1A T. microellipsoides* sequences observed as a monophyletic group as in the partial sequence reconstruction. The *T. microellipsoides FSY1B*/*FSY1B2* duplicates formed a group with *S. uvarum*/*S. eubayanus*/*S. pastorianus*, and *T. delbrueckii* was the closest species to join this group. Despite the observed wide variability in the 5' and 3' nucleotide extremes of the gene, we obtained similar results using the complete or partial sequence. This finding suggests the robustness of our results.

1.2.4 Support of *FSY1* introgression by a tree topology comparison

The reconstruction of the *FSY1* compiled sequences (Figure 1.2 and 1.4) revealed that those from *T. microellipsoides* and some *Saccharomyces* species ocuppied in an unexpected position compared to their phylogenetic position established for the *Saccharomycotina* complex (Figure i6). As mentioned before, this result suggest that probably an introgression of this fructose symporter gene could be directed from a *T. microellipsoides* strain towards some *Saccharomyces* species in two independent gene transfer events. In order to assess if the obtained ML phylogeny was statistically acceptable, we followed a method that compares different tree topologies and selects which is the best based on likelihood calculations (Figure.1.5 Panel A and B.). As explained in the 1.3.5 Methodology

section, three hypothetical comparisons with four trees each, were tested to statistically support the introgression. In all the comparisons made of both the topologies obtained with the partial and complete *FSY1* sequence phylogenies, the best topology was always that for the *FSY1* ML tree.

FSY1A_FSY1B/FSY1B2 **PANEL A**

tree	lnL	ΔlnL	±SE	pKH	pSH	pRELL
1 (best)	-17464.722	0.000	0.000			1.000
2	-18009.193	-544.470	53.120	0.000	0.000	0.000
3	-17738.035	-273.312	28.074	0.000	0.000	0.000
4	-17832.444	-367.722	37.247	0.000	0.000	0.000

tree	lnL	ΔlnL	±SE	pKH	pSH	c-ELW
1 (best)	-17838.6	0	0.000	1.000	1.000	1.000
2	-18411.41	-572.81	50.921	0.000	0.000	0.000
3	-18123.43	-284.83	28.087	0.000	0.000	0.000
4	-18395.15	-556.55	30.479	0.000	0.000	0.000

tree	lnL	ΔlnL	±SE	pKH	pSH	pRELL
1 (best)	-16758.778	0.000	0.000			0.995
2	-17089.279	-330.501	45.319	0.000	0.000	0.000
3	-16789.650	-30.872	12.442	0.007	0.262	0.005
4	-16893.779	-135.001	22.001	0.000	0.000	0.000

tree	lnL	ΔlnL	±SE	pKH	pSH	c-ELW
1 (best)	-17108.180	0.000	0.000	1.000	1.000	0.994
2	-17453.510	-345.330	45.666	0.000	0.000	0.000
3	-17138.680	-30.500	12.268	0.011	0.267	0.006
4	-17248.320	-140.140	22.065	0.000	0.000	0.000

tree	lnL	ΔlnL	±SE	pKH	pSH	pRELL
1 (best)	-17436.638	0.000	0.000			0.992
2	-17565.812	-129.174	22.347	0.000	0.000	0.000
3	-17511.895	-75.258	17.362	0.000	0.000	0.000
4	-17465.600	-28.962	12.284	0.009	0.076	0.008

tree	lnL	ΔlnL	±SE	pKH	pSH	c-ELW
1 (best)	-17802.160	0.000	0.000	1.000	1.000	0.984
2	-17934.630	-132.470	22.305	0.000	0.000	0.000
3	-17878.600	-76.440	17.463	0.000	0.000	0.000
4	-17830.520	-28.360	12.136	0.013	0.077	0.016

FSY1A_FSY1B/FSY1B2 **PANEL B**

tree	lnL	ΔlnL	±SE	pKH	pSH	pRELL
1 (best)	-17158.980	0.000	0.000			1.000
2	-17858.095	-699.116	55.179	0.000	0.000	0.000
3	-17588.558	-429.578	31.262	0.000	0.000	0.000
4	-18150.020	-991.040	36.309	0.000	0.000	0.000

tree	lnL	ΔlnL	±SE	pKH	pSH	c-ELW
1 (best)	-18306.650	0.000	0.000	1.000	1.000	1.000
2	-19080.210	773.560	55.304	0.000	0.000	0.000
3	-18775.940	469.290	31.752	0.000	0.000	0.000
4	-19363.210	1056.550	38.004	0.000	0.000	0.000

tree	lnL	ΔlnL	±SE	pKH	pSH	pRELL
1 (best)	-15744.664	0.000	0.000			1.000
2	-16087.871	-343.206	38.808	0.000	0.000	0.000
3	-15788.053	-43.389	16.092	0.004	0.125	0.000
4	-16054.771	-310.107	35.228	0.000	0.000	0.000

tree	lnL	ΔlnL	±SE	pKH	pSH	c-ELW
1 (best)	-16795.430	0.000	0.000	1.000	1.000	1.000
2	-17141.340	345.920	37.969	0.000	0.000	0.000
3	-16844.320	48.890	16.039	0.005	0.085	0.000
4	-17103.770	308.340	34.360	0.000	0.000	0.000

tree	lnL	ΔlnL	±SE	pKH	pSH	pRELL
1 (best)	-17091.680	0.000	0.000			1.000
2	-17154.979	-63.299	17.250	0.000	0.000	0.000
3	-17139.648	-47.968	13.336	0.000	0.001	0.000

tree	lnL	ΔlnL	±SE	pKH	pSH	c-ELW
1 (best)	-18233.260	0.000	0.000	1.000	1.000	1.000
2	-18303.820	70.550	17.607	0.000	0.000	0.000
3	-18287.670	54.400	13.746	0.000	0.000	0.000

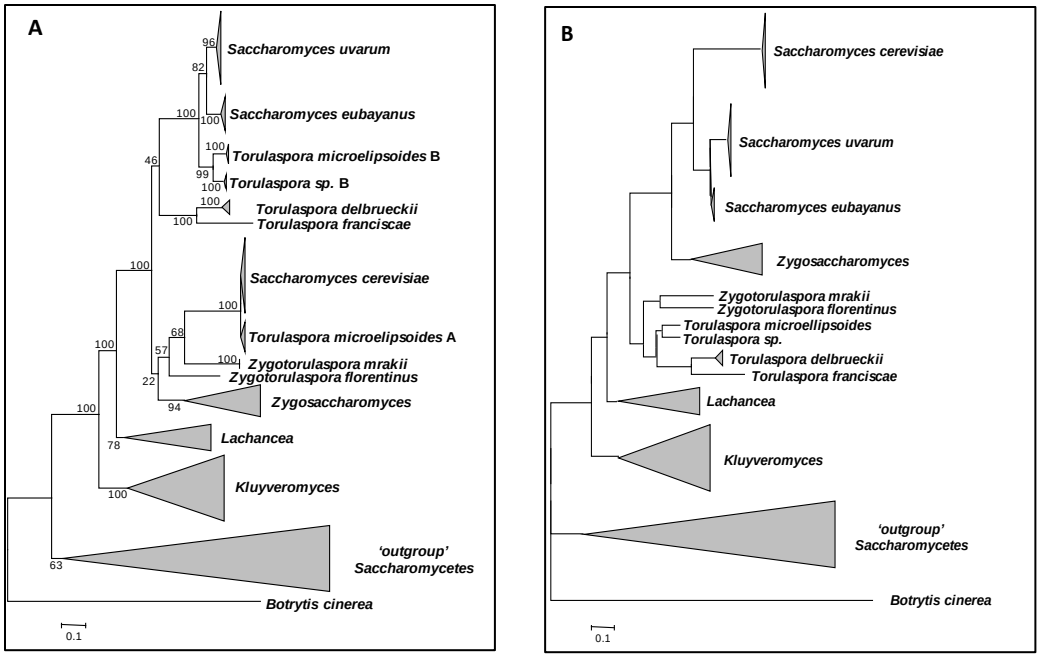
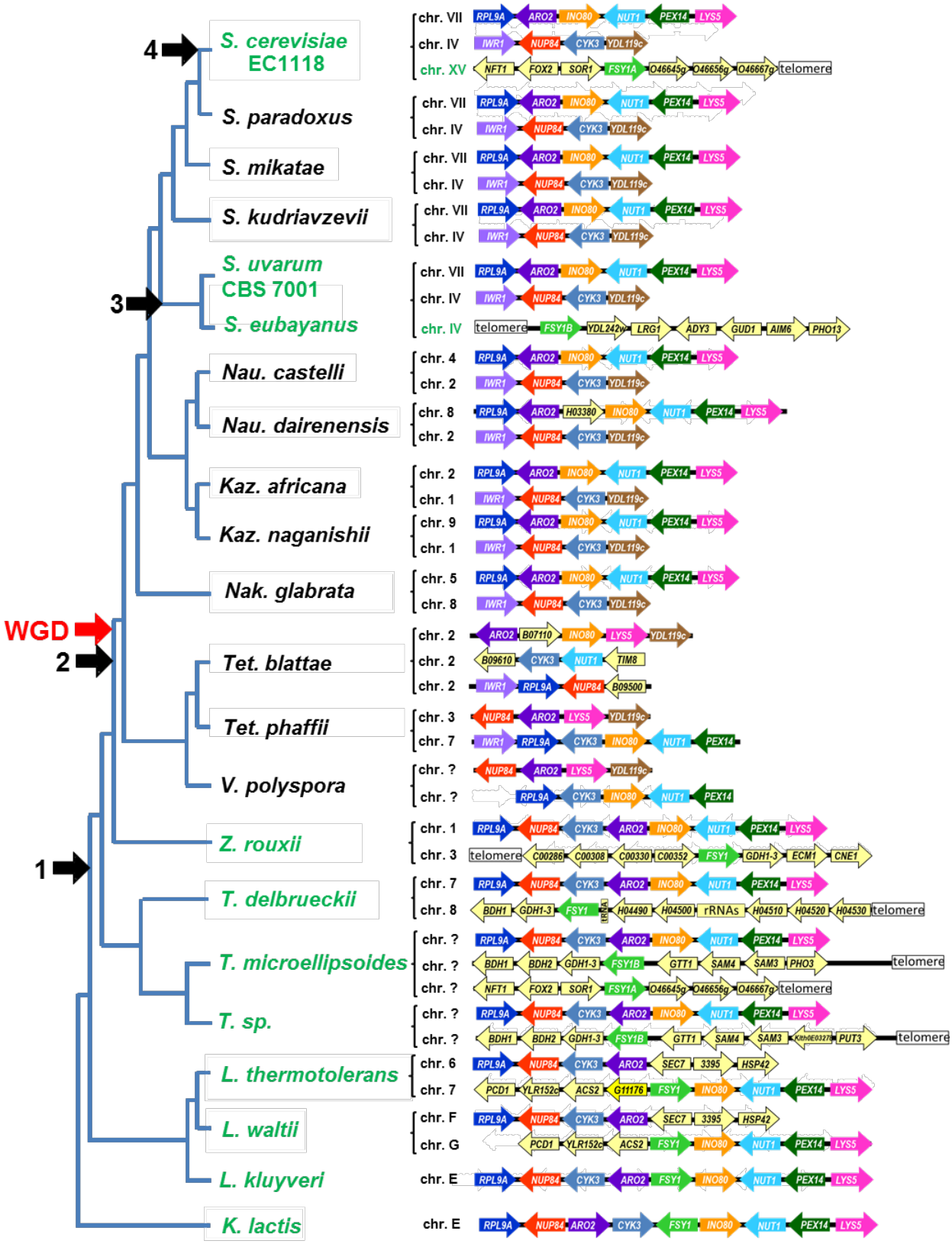


Figure 1.5. Topology comparison through the Shimodaira-Hasewaga test implemented in PAML 4.9c. Three different comparisons (*FSY1A* and *FSY1B/FSY1B2*; only *FSY1A* and only *FSY1B/FSY1B2*), including four or three topologies, are shown in upper panels A and B. The tested tree topologies were: **tree 1 (corresponding to tree A on the left)**: the maximum likelihood (ML) tree obtained by the PhyML software with the evolutionary model chosen in jModeltest; **tree 2 (corresponding to tree B on the left)**: the *Saccharomycetaceae* phylogeny from Kurtzman *et al*., 2003; **tree 3**: the ML tree forcing the *T.microellipsoides* sequences to the predicted position for *FSY1A*; **tree 4**: the ML tree forcing the *T.microellipsoides* sequences to the predicted position for *FSY1B*/*FSY1B2*. **PANEL A**: topology testing by taking the *FSY1* partial sequence reconstruction. **PANEL B**: topology testing by taking the *FSY1* complete sequence tree. **lnL**: ln likelihood, SE: Standard Error, **pKH**: P-value for the KH normal test (Kishino and Hasewaga 1989), **pSH**: P-value with multiple-comparison correction, **pRELL**: RELL bootstrap proportions (Kishino and Hasegawa 1989) and **c-ELW**: Expected Likelihood Weights (Strimmer and Rambaut 2002).

1.2.5 Synteny and position of *FSY1* among *Saccharomycetaceae* phylogeny: a proposed model



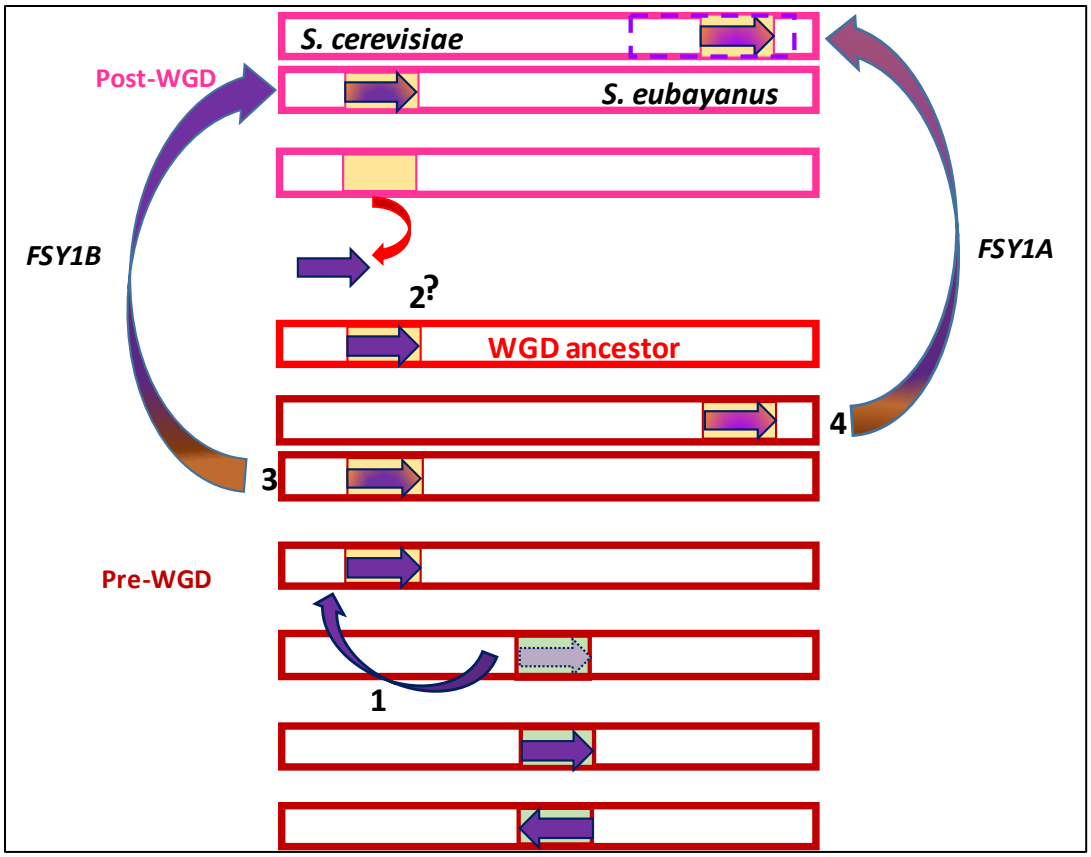


Figure 1.6. Distribution of the *FSY1* gene in the "*Saccharomyces* complex". On the left, we show the chromosome position and the synteny view (extracted partly from YGOB) in the Saccharomycetaceae phylogeny. On the right, we draw the hypothetical model of transfer from a pre-WGD species to the *Saccharomyces* species: **1**: Movement of *FSY1* towards the subtelomeric position; **2**: *FSY1* was lost in all or in some post-WGD genera. Its presence in the WGD ancestor is uncertain (symbolised as "**?**"). If the *FSY1* acquisition in the *Saccharomyces* genus was done through two independent gene transfer events, we hypothesise 3 and 4: **3**: acquiring *FSY1B* (or *FSY1B2*) by the ancestor of *S. uvarum*, *S. eubayanus* and *S. pastorianus* and 4: acquiring *FSY1A* differently by a wine lineage of *S. cerevisiae*. The arrows inside boxes correspond to the *FSY1* gene and the longer arrows represent the transfer event.

We took the non-*Saccharomyces* whole genome sequenced species to reconstruct the gene context of *FSY1* and to propose a model of a possible transfer to discuss it (Figure 1.6). As we previously detected *T. microellipsoides* to be the possible donor species, we selected some genera that diverged before. From the *Kluyveromyces* genus, we had the complete sequence of *Kluyveromyces lactis* and the three species from *Lachancea*: *L. thermotolerans*, *L. waltii* and *L. kluyveri* available. Inside these genomes, the *FSY1* gene was located at a central position. In *K. lactis* it was surrounded by *CYK3* and *INO80* and then an inversion of *FSY1* and *CYK3* and *ARO2* was observed in *L. kluyvery*. A translocation may have occurred in *L. waltii* and *L. termotolerans* at a point before starting *FSY1* ORF. In species *Torulaspora*, *T. microellipsoides* and *T. delbrueckii*, we observed that *FSY1* moved from the *CIK3*/I*NO80* context to a different chromosome and in a subtelomeric position. We draw the chromosomes for each copy that we found, *FSY1A* and *FSY1B*/*FSY1B2,* as we cannot deduce the exact ancestral and duplicated copy, but we discuss this aspect in the next results section. In *Zygosaccharomyces rouxii*, we also detected this gene at the end of the chromosome. The same genes with the same order are arranged in this species, as *FSY1* is the only gene that moved from its ancestral loci and was implicated in a translocation. When we move up in the phylogeny, we start to find the post-WGD genera: *Vanderwaltozyma*, *Tetrapisispora*, *Nakaseomyces*, *Naumovia* and *Kazachstania*, in whose genomes *FSY1* was not found. So we show the ancestral gene position, which is split and modified as a result of WGD. Inside the *Saccharomyces* genus, we redraw the *FSY1* gene for the resulting species *S. eubayanus* and *S. uvarum*, and for *S. cerevisiae* again in a chromosome terminal position.

We present a model for the gene transfer that simplifies the first figure, where we wish to emphasise that *FSY1* moves from a central to a subtelomeric position, and firstly in the *Torulaspora* genus. We later discuss the ancestral and

duplicated copy of *FSY1*, but in any case *FSY1A* or/and *FSY1B*/*FSY1B2* probably moved from the *T. microellipsoides* to the *Saccharomyces* species thanks to the recombinational properties of rich-in-repeats subtelomeric positions which facilitates the interchange of genetic material. We especially wish to remark with the broken lined square that the *FSY1* from *S. cerevisiae* was transferred together with other genes that also proceed from *T. microellipsoides*, which is called Region C. However, the *FSY1* from *S. eubayanus* is surrounded by its own genes and none proceeded from a horizontal gene transfer.

1.2.6 *ATO3* phylogeny: maximum likelihood (ML) tree reconstruction of an *ATO3* partial gene sequence

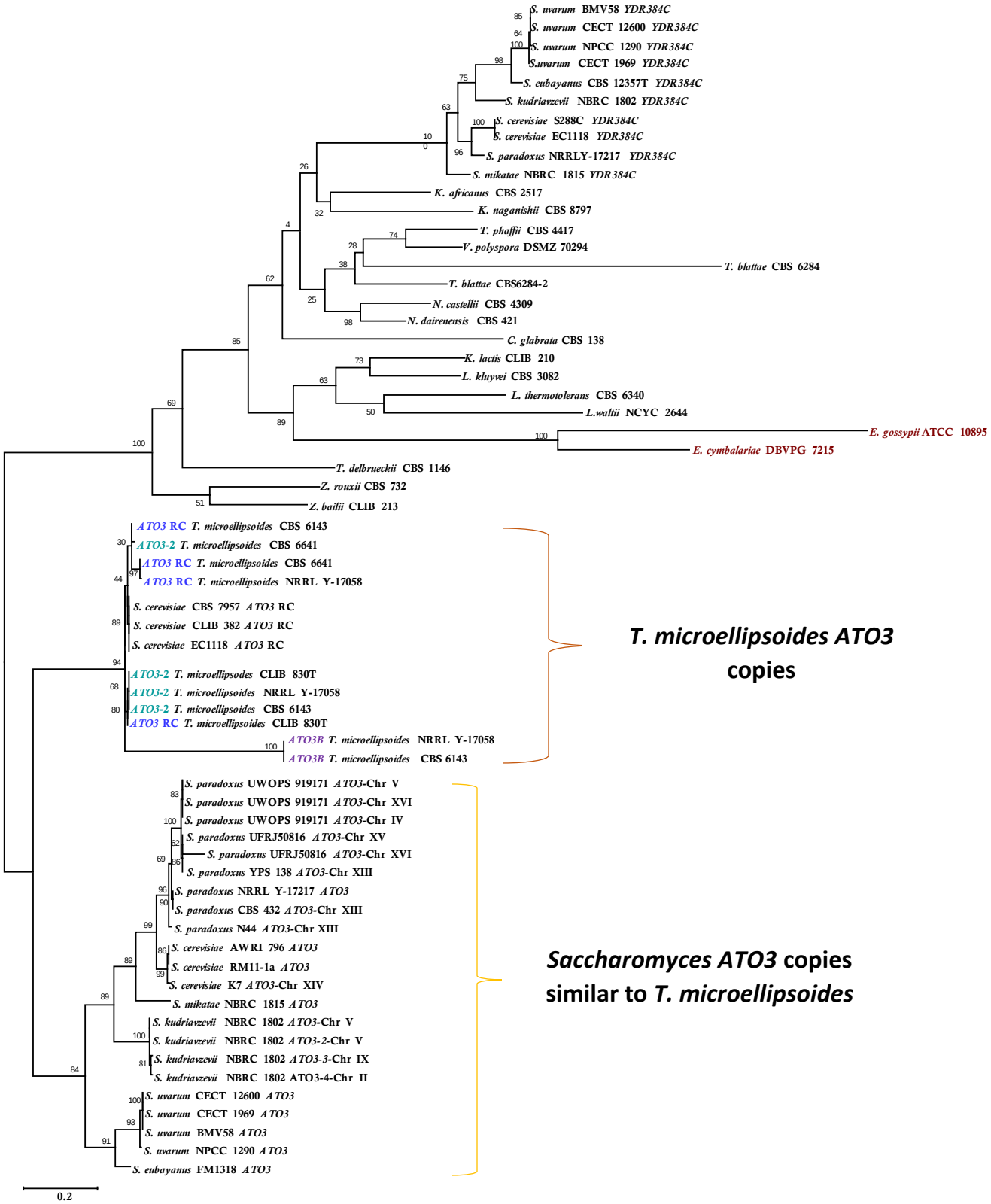


Figure 1.7. The ML phylogeny obtained from a 543 bp *ATO3* partial sequence among *Saccharomycetaceae* species. The outgroup species were *E. gossypii* and *E. cymbalariae* (drawn in red). Species' name and strains number refer to a yeast collection or a source of origin. Different tones of blue highlight the *ATO3* copies found in the *T. microellipsoides* strains. The numbers at nodes correspond to the bootstraps values over 100 bootstrap pseudo-replicates. The scale is given in nucleotide substitutions per site.

In order to detect any possible new introgressions from *T. microellipsoides*, we searched for the homologous genes of the *S. cerevisiae*/*T. microellipsoides* Region C genes in the genomes of the *Saccharomyces* species. Of them all, we only detected a copy of the *ATO3* gene (whose inferred function is ammonia export). We performed a phylogenetic reconstruction with the newfound sequences, together with those of Region C and the ancestral gene *ATO3* present in the Saccharomycetaceae genera (Figure 1.7). It is noteworthy that the size of the sequences in the reconstruction went up to 543 bp, but that of a complete sequence for the ancestral gene and the new copies was up to 831 bp. This was because a codon stop was found in the *S. cerevisiae* Region C *ATO3* gene at the 56 amino acid position when aligned against the *ATO3* gene in *T. microellipsoides*. Thus, the ORF in Region C of *S. cerevisiae* was shorter than the **orthologous** sequence in other species. By taking the corresponding nucleotide sequence before this stop codon, and translating it into amino acids and aligning it against the new copies found in the *Saccharomyces* species, it still conserved some amino acidic positions and others were probably lost by a genetic drift (Supplemental Figure 1.3) We included *E. gossypii* and *E. cymbalariae* as outgroup species. Interestingly inside the same chromosome of the *T. blattae* CBS 6284 strain, we detected two copies of the ancestral homologous gene.

The new copy was found in the majority of the *Saccharomyces* species, except in *S. arboriculus*, and no sequence or DNA was available for *S. jurei*. Although *S. cariocanus* was not included in the new *Saccharomyces* phylogeny, a novel *ATO3* sequence was present in strain *S. paradoxus* UFRJ50816, regarded before as *S. cariocanus*. The new *Pacbio* technology assemblies from *S. cerevisiae* and *S. paradoxus* (Yue et al. 2017) allowed us to better resolve subtelomere sequences and more than one copy of *ATO3* was found in strains *S. paradoxus* UWOPS 919171. Within it, gene copies were located in chromosomes V, XVI and IV and two subtelomeric copies in each extreme of chromosome XV. In four additional *S. paradoxus*, the same copy was present and three of them harboured it in chromosome XII. Sequences of *S. paradoxus* were joined to those of *S. cerevisiae* in a monophyletic group. For this species, we obtained from the three strain sequences for the tree reconstruction and in one of them that the gene was present in chromosome XIV. We included one sequence of a *S. mikatae* strain and four distributed copies of the gene in *S. kudriavzevii* NBRC 1802 (chromosomes V, a second copy in chromosome V, chromosome IX and II). Four *S. uvarum* strains and one *S. eubayanus* presented each one, one copy of the gene.

We searched these novel sequences in the genome of the hybrid of *S. pastorianus*, but no results were obtained .In the *S. cerevisiae* and *S. kudriavzevii* parental genomes from the VIN7 strain, we found an *ATO3* gene only in the *S. kudriavzevii* subgenome. The *S. cerevisiae* origins of the strains where we found this new gene were Australia, EEUU and Hawaii. No evidence for a European origin was obtained. Moreover, the inferred VIN 7 *S. cerevisiae* subgenome origin was Europe and the closest strains found was EC1118 (Borneman et al. 2012).

1.3. Methodology:

1.3.1 The yeast strains used in this chapter

Partial or complete *FSY1* gene sequences were obtained from yeast databases (Strains Table 1) or by PCR amplification (Strains Table 2). The complete sequences of *FSY1 Torulaspora microellipsoides* strains were extracted from genome sequencing in our lab group (see Chapter 2). Yeast cultures were performed in YPD media: yeast peptone dextrose broth (YPD) rich media (2% w/v yeast extract, 2% w/v peptone and 2% glucose, pH = 6.5). We grew *S. cerevisiae* species at 30^0C and 190 rpm, and different optimal temperatures were used for other *Saccharomyces* and *non-Saccharomyces* species. Sequences of *ATO3* homologous genes were obtained from databases and publications, and the species used in this work are listed in Strains Table 3.

1.3.2 *In silico* search for *FSY1* and the *ATO3* gene

For some strains listed in Strains Table 1**,** we found complete or partial fructose transporter sequences in the databases of the National Center for Biotechnology Information, **N.C.B.I**.,U.S. National Library of Medicine 8600 Rockville Pike, Bethesda MD, 20894 USA and Welcome Trust Sanger Institute, Hinxton, Cambridge,UK. Homologous sequences of *ATO3* gene were retrieved from **N.C.B.I.**, and from the high-resolution genome sequencing of *S. cerevisiae* and *S. paradoxus* published by Peter et al. 2018 and Yue et al. 2017.

1.3.3 Isolation of yeast genomic DNA, PCR amplification and Sanger sequencing

In order to obtain genomic DNA from the yeast described in **Strains Table 2**, an overnight culture was placed in YPD-rich media at 25-30^{0}C and 190 rpm. Then genomic DNA was extracted according to the procedure of Querol *et al.,* (1992). A pair of general primers was designed for the PCR amplification of a 903-bp central region of *FSY1* by taking the complete sequences available in databases. No complete gene sequence amplification was possible as a sequence of the extremes was too variable. PCR amplification was performed among the *Saccharomycetaceae* species by taking ***FSY1g*F**: 5'-GAAGGTGGTGGTRTTGGTGT-3' and ***FSY1g*R:** 5'-GCRATACCACCGTARAAGCC-3' and one different forward primer to specifically amplify the *FSY1 S. cerevisiae* sequence inside the hybrid genomes: ***FSY1*$_{EC}$F:** 5'GAAGGAGGCGGTRTTGGCGT-3'. We employed *Phusion High-Fidelity DNA Polymerase* (Thermo-Fisher Scientific) with the suggested protocol and PCR conditions were as follows: an initial denaturing at 98^{0}C for 30s; 30 cycles of denaturing at 98^{0}C for 10s, annealing at 62^{0}C for 30s and an activity step at 72^{0}C for 30s. The programme was completed with a final step at 72^{0}C for 5 minutes. The obtained fragments were purified with the *High Pure PCR Product Purification Kit* (Roche Molecular Systems). For Sanger sequencing, we labelled samples with *BigDye Terminator v3.1 Cycle* (Thermo-Fisher Scientific) and the next cycling programmme was as follows: initial denaturing at 94^{0}C for 3 minutes; and 99 cycles consisting in: 1) 96^{0}C for 10 s, 2) 50^{0}C during 5 s, and 3) 60^{0}C for 4 minutes. The labelled fragments were red with ABI 3730 (*Applied Biosystems*, Foster City, CA, USA).

1.3.4 *FSY1* and *ATO3* phylogenies reconstruction

The obtained sequences for each gene were aligned using the Mega *v*6.0 software (Tamura *et al.,* 2008) at their amino acid level. To determine the best evolutionary model that fitted our sequences, we employed jModelTest *v*2.1.7 (Posada, 2008). Firstly, an alignment of the *FSY1* partial sequence from 75 *Saccharomycetaceae* strains was tested for an evolutionary model. According to the AIC, AICc, BIC and DT criteria, three different models were the best obtained ones. After the LRT comparison, the best fitting model was 012212 +I+G+F. A second alignment of 28 complete *FSY1* sequences (1794 bp) was analysed for model selection purposes. In this case, and according to the same criteria, the chosen model chose was 012232+I+G+F. The evolutionary model chosen for the *ATO3* sequences was TIM2+I+G in accordance with all the criteria (AIC, AICc, BIC and DT).

A maximum Likelihood (ML) tree was obtained using the PhyML 3.0 software for each alignment (Guindon S *et al*., 2010). Although the invariant sites (I), gamma shape (G) and base frequencies (F) values were always provided with the model, estimates of them using eight rate categories were recalculated with PhyML. Bootstrap support of 1000 or 100 trees was evaluated.

1.3.5 *FSY1* tree topology comparison

Having obtained tree reconstructions using the *FSY1* complete or partial sequence, different tree topologies were compared because the unexpected phylogenetic positions resulted for the *T. microellipsoides FSY1A* and *FSY1B*/FSY1B2 copies compared to the established family phylogeny. In order to statistically obtain significance for the observed topology, we employed the PAML4.9c software (http://abacus.gene.ucl.ac.uk/software/paml.html) to apply the Shimodaira-Hasewaga test based on the total likelihood of each topology. Four trees were tested by assuming three different sequence comparisons: in the first one, both sequences *FSY1A* and *FSY1B* were included, and only sequences *FSY1A* or *FSY1B* were considered in the other two.

1.3.6 Reconstruction of the *FSY1* chromosome position in pre- and post-WGD species

The synteny arrangement of the pre-WGD and post-WGD species, the synteny from the Yeast Gene Order Browser (YGOB) and from other sources annotated species as *S. eubayanus*, or in our group annotated *T. microellipsoides* strains, were explained to reconstruct the position and the genome context of *FSY1* in the *Saccharomycetaceae* species. The genes adjacent to the fructose transporter sequence were inspected from the bottom linages (*Kluyveromyces* and *Lachancea)* to the *Saccharomyces* genus to show if the gene was maintained at the same or a different chromosome position. We employed the A*rtemis* software to make the annotation work.

1.4. Main conclusions:

- In the Saccharomycetaceae family, the homologous sequence of a gene that codifies for a fructose/H^+ symporter, ***FSY1***, is present in the majority of the pre-WGD species and is absent in nearly all the post-WGD genera, except in the *Saccharomyces* genus.

- In the *T. microellipsoides* genome, two subtelomeric copies of *FSY1* ORF exist, which we named *FSY1A* and *FSY1B*. As the phylogenetic analysis revealed that each copy was similar to one *Saccharomyces* species a horizontal gene transfer from the former species was suggested.

- Both copies' nucleotide divergence from *S. cerevisiae* to their probable donor indicated two independent transfer events of transference from *T. microellipsoides* to *Saccharomyces* species. The introgression in *S. cerevisiae* would be more recent than in *S. eubayanus* and its related hybrids.

- The topology tree obtained for the *FSY1* gene was well supported compared to the topologies expected for the phylogeny of the *Saccharomyces* complex.

- The reconstruction of the *FSY1* chromosomic position from the basal genera of the Saccharomycetaceae family complex to the genera in the upper part of the tree provided much more support to transfer events because this gene was present at a subtelomeric position.

- Regarding a within Region C gene *ATO3*, a homologous copy was found in the majority of the *Saccharomyces* species. The *T. microellipsoides* species was once again suggested as the probable donor.

Chapter 2:

Torulaspora microellipsoides as a species that contributes gene novelties to domesticated *Saccharomyces* strain

2.1. Abstract:

Since *T. microellipsoides* CLIB 830T genome sequencing, the introgression of Region C in some *S. cerevisiae* strains has been attributed to this species. In this chapter it is shown the whole genome sequencing of the four from different sources of existing *T. microellipsoides* isolates, NRRL Y-17058, CBS 6143, CBS 6641 and CBS 6762 (see Strains Table 2). Then, to confirm that the direction of the transfer of this region went from *T. microellipsoides* to *S. cerevisiae* and not the other way round, this region was searched in the new assembled genomes and attempted to reconstruct the chromosomic context of the *FSY1B*/*FSY1B2* genes. An analysis of the *FSY1* and *ATO3* phylogenies was done after we obtained the new genome assemblies.

All the *T. microellipsoides* strains, except CBS 6762, presented a conserved synteny of Region C, which supports the hypothesis of the transfer from this species towards a *S. cerevisiae* wine lineage. The *FSY1B*/*FSY1B2* genes were also found at a subtelomeric location and every strain presented one or both copies, which suggests that they were also transferred from *T. microellipsoides* to *S. eubayanus* and the associated hybrids ancestor. The *FSY1B*/*FSY1B2* region showed a well-conserved synteny with species *Z. rouxii* and *T. delbrueckii* in the core and the subtelomeric region of the pertinent chromosomes.

An analysis of *FSY1* and the *ATO3* phylogeny revealed that strains NRRLY-17058 and CBS 6143 presented more than one duplicated gene. The interpretation of the *ATO3* gene phylogeny was more complicated given its gene redundancy in these non-*Saccharomyces* strains. One of the *ATO3* paralogous genes that we found in strains NRRL Y-17058 and CBS 6143 was inside a “reduced Region C”, and involved *DSF1*, *HXT13*, this *ATO3* gene duplication, *GA10*, *SOR1* and the *FOX2* genes.

When we looked at the *FSY1* phylogeny, we observed that for strains NRRL Y-17058 and CBS 6143, one copy came close to strains CBS 6641/CLIB 830T and the other one to strain CBS 6762. From the resultant tree topology, we suspected that they were probably two hybrid species and their likely parental strains were CBS 6762 and CBS 6641. In order to confirm our hypothesis, we took an alignment of 1,844 concatenated genes retrieved from *de novo* genome annotations and reconstructed a phylogeny for the species. As a result, two well-differentiated groups were drawn by this multi-sequence analysis. We calculated the evolutionary divergence between genomes and the inferred subgenomes. With this measure, CBS 6762 and one subgenome shared by the two hybrids presented a long nucleotide distance compared to other sequences. We concluded that, the CBS 6762 strain would probably not be a *T. microellipsoides* species, but a different species.

New *Torulaspora* species have been discovered in recent years (Saluja et al. 2012) based on their rDNA, and on other sequenced nuclear and mitochondrial genes. To test our hypothesis, which stated that strains CBS 6762 may be a distinct *Torulaspora* species, we decided to perform the 18S/ITS/5.8S and 26S multi-gene phylogeny with the type strain of each newly found species. From the resultant phylogeny, we observed that this strain presented quite a distant position from the other *T. microellipsoides* strains but, at the same time, its close ancestor was shared with them. The greater the isolation of the *Torulaspora* strains, the more sequences we retrieved to refine the phylogenetic analysis about the *Torulaspora* genus. The new analysis gave similar results to those obtained from the type strain sequences: while the CBS 6762 strain was not closely related to any *Torulaspora* species, it still formed a monophyletic group with the CBS 6641 and CLIB 830T *T. microellipsoides* strains.

In spite of the CLIB 830T genome being sequenced in Marsit et al. 2015, no gene annotations existed from it to map our **reads** against them. We used the YGAP approach (Proux-Wéra et al. 2012) for the assembled sequences that we obtained to annotate the CBS 6641 and CBS 6762 genomes. For strains CBS 6143 and NRRL Y-17058, which we suspected to be hybrid species, we created a set of eight *pseudo-chromosomes* (which is the more or less expected number of chromosomes for the species based on karyotype assays) by **mapping** their reads against the scaffolds from the inferred parental strains (CBS 6641 and CBS 6762). From the resultant *pseudo-chromosomes*, we concluded that both hybrid species shared a common origin, but we observed some huge genomic differences along this reconstruction. Previously, a karyotyping analysis revealed that both strains presented a distinct chromosomic band pattern.

Although genome assembly provided an estimation of genomic DNA content, complementary methods were employed to assign a genome size and DNA content to each *T. microellipsoides* strain. We firstly recovered the previous karyotyping analysis to obtain the number and size of chromosomal bands. Secondly, flow cytometry assays were performed to measure DNA content and ploidy.

With all this work, we obtained the genome sequence, assembly and annotation of one *T. microellipsoides* strain (CBS 6641), one novel *Torulaspora* species (CBS 6762) and two hybrid species (CBS 6143 and NRRL Y-17058). Both hybrids shared a common origin, and harbouried one parental genome from *T. microellipsoides* (CBS 6641) and another one from a new *Torulaspora* species before being included in *T. microellipsoides* (CBS 6762).

2.2. Results:

2.2.1 The results Inferred from the *FSY1* and *ATO3* phylogenies

The phylogenies of these two Region C genes shown in Chapter 1 were completed once the genome sequence of all the *T. microellipsoides* strains was obtained. At this point, only the CLIB 830T genome sequence was available, as described in Marsit et al. 2015, but the other isolated strains were not already genome-sequenced. That work revealed a high probability of *T. microellipsoides* being the species responsible for the Region C transfer. In it, the complete region and the conserved distribution of the genes were found, but with changes in *ARB1* ORF with a bigger size and the insertion of a group of genes of approximately 80kb between *ARB1* and the *PUT3* gene.

A similar partial or extended version to *T. microellipsoides* of Region C was found in the newly sequenced *S. cerevisiae* strains in Marsit et al. 2015. Two of these strains, CLIB 382 and CBS 7957 presented four of the additional genes found next to the *ARB1* gene in *T. microellipsoides*: *COG1*, *SDT1*, *KRE1* and *VRG4.* Therefore, these *S. cerevisiae* strains presented a more similar composition of the Region C genes to *T. microelipsoides* than to *S. cerevisiae* EC1118. Therefore, the direction of the introgressed region was investigated here because of the presence of the extended Region C in two S. cerevisiae strains. Moreover, evidence for Region C was confirmed only for one *T. microellipsoides* strain CLIB 830T.

Thus in order to establish the correct Region C transfer direction, the genome sequences of the remaining *T. microellipsoides* strains were obtained. Five strains from different origins and sources of isolation existed in the yeast culture collections (all of which are available in the CBS culture collection), namely: CLIB 830T, NRRL Y-17058, CBS 6143, CBS 6641 and CBS 6762. A detailed description of these strains is shown in Strains Table 2.

First of all, the *FSY1* paralogous genes (*FSY1A* and *FSY1B*) that were already found in the type strain were searched together with the complete Region C (Supplemental Figure 1). With all the retrieved *FSY1B* copies, a reconstruction was performed (shown in the Supplemental Figure 3). Subtelomeric regions of interest were manually annotated using the Artemis software (Carver et al. 2012). The annotation of the genes surrounding the *FSY1B* copy were extracted and the synteny was represented (Supplemental Figure 2).

The phylogenetic incongruences found for the *T. microellipsoides* species in the *FSY1* and *ATO3* phylogenies, but not inversely for the *Saccharomyces* species (Figure 2.1), pointed out that the former species was the probable donor of these genes by horizontal gene transfer (HGT). Not only were the *FSY1* and *ATO3* genes found in the *T. microellipsoides* strains, but Region C was also found in four of the five strains: CLIB 830T, CBS 6641, NRRL Y-17058 and CBS 6143. As suspected, the complete region was absent in the CBS 6762 strain (Supplemental Figure 1). The subtelomeric region harbouring the *FSY1B1* copy was present in CLIB 830T, CBS 6641, NRRL Y-17058 and CBS 6143.

A second *FSY1B* copy was detected, which we called *FSY1B2*, in the CBS 6762, NRRL Y-17058 and CBS 6143 strains (Supplemental Figure 2). In short, the detailed analysis of the *FSY1* phylogeny revealed relevant information about the *T. microellipsoides* strains.

By focusing on this species (Figure 2.1), duplicated sequences in the NRRRL Y-17058 and CBS 6143 strains are observed. The *FSY1A* sequence of these two strains joins the same copy of CLIB 830T and CBS 6641. The *FSY1B1*/*FSY1B2* copies of the same strains are similar to the copies of the CLIB 830T/CBS 6641 strains and the CBS 6762 strains, respectively. From this assignment, it was concluded that it was probably found two hybrid strains, NRRRL Y-17058 and CBS 6143, as well as the parental species involved in the hybridisation event: CBS 6762 and CBS 6641.

The phylogeny of the *ATO3* gene (Figure 1.7) blurred this aspect because of the numerous paralogous sequences found in these species and the absence of Region C in CBS 6762. Perhaps the *ATO3* gene underwent some evolutionary process, and in such a way that copies could recombine among them or fix some important changes through selection. Despite the CBS 6762 strain being absent in this phylogeny, it was observed that the CBS 6641 sequence was the closest to those of the assumed hybrid species. Besides, at least three sequences were obtained for NRRRL Y-17058 and the CBS 6143 strain, which supports the conclusions drawn from the *FSY1* phylogeny.

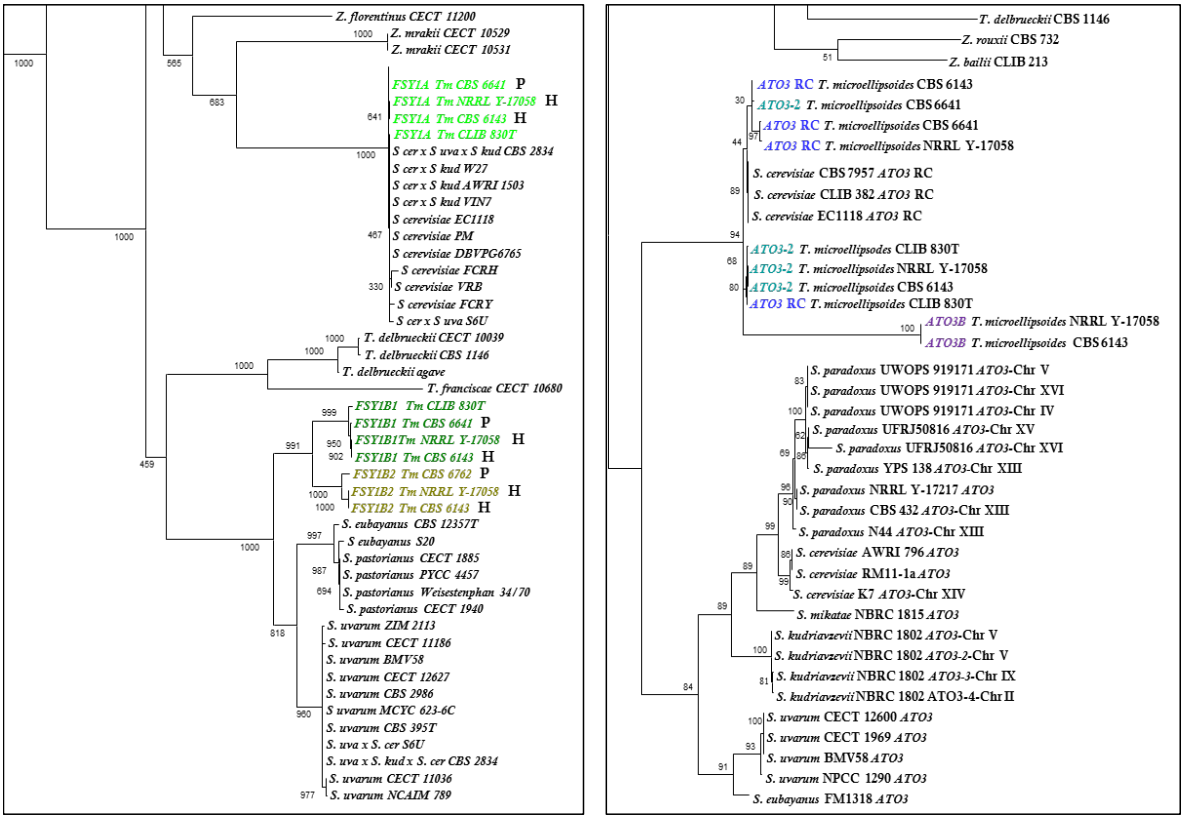


Figure 2.1. Suspected parental and hybrid species from the Region C genes analysis. Zoom on the previously presented ML phylogenies. All the *FSY1* (in green) and *ATO3* (blue or purple) sequences were obtained after genome assemblies. The probable parental (**P**) and hybrid (**H**) genomes deduced from these preliminary analyses are tagged in the image. No assignation was made to the *ATO3* phylogeny due to its complicated interpretation.

2.2.2 Distribution of the *FSY1* and *ATO3* genes in each *T. microellipsoides* strain

The fructose symporter gene distribution of the adjacent genes was first known in the *S. cerevisiae EC1118* strain (Novo, Bigey, Beyne, Galeote, Gavory, Mallet, Cambon, J.-L. Legras, et al. 2009), and later in the *T. microellipsoides* CLIB 830T strain and the new sequenced *S. cerevisiae* strains (Marsit et al. 2015). Now, homologous Region C from strains NRRL Y-17058, CBS 6143,CBS 6641 was obtained. A similar arrangement and composition of genes to that found in the CLIB 830T strain was found. A duplication of the chromosomic core region next to Region C was discovered in strains NRRL Y-17058 and CBS 6143. However, the corresponding subtelomeric region presented a distinct gene composition, and some were Region C paralogous copies. The *PUT3*, *SEO1* and *FOT1* copies came close to the telomere region together with Y'-elements and a *FLO9* homologous gene. As in the CLIB 830T Region C, an additional region of approximately 95 kb maintained a distance between the *PUT3* gene and the *ARB1* gene, and the size of the latter gene was similar to the paralogous one in the former Region C. In addition to search of-interest Region C genes (*FSY1* and *ATO3*), the genome assemblies were screened for the remaining Region C genes. Little *subregions* were found in strains NRRL Y-17058 and CBS 6143 (Figure 2.4), where there was a duplicated *ATO3* gene which presented a substantial number of changes (see *ATO3* tree). This stretch of genes was composed of *DSF1*, *HXT13*, *ATO3*, *GAL10*, *SOR1* and *FOX2*. It was frequently observed that the subtelomeres in these Region C genes were enriched and were quite propagated among the *T. microellipsoides* strains.

Afterwards the *T.microellipsoides* Illumina reads were mapped against the *S. cerevisiae* EC1118 Region C as described in the Methodology (Figure 2.3). A similar coverage distribution in the NRRL Y-17058, CBS 6143 and CBS 6641 strains was

observed on the mapping profiles. For CLIB 830T, one different profile was obtained with a read enrichment in the coverage in the telomere region, which was probably due to the Y'-elements reads. The coverage obtained for strain CBS 6762 was irregular with some large and narrow picks, which suggests that no Region C and core region next to it were present, which was as expected when blast searches were performed.

A second *FSY1* sequence in *T. microellipsoides* CLIB 830T was discovered (similar to *S. eubayanus* and its derived hybrids), but the genomic context of this gene copy was unknown until the genome sequencing of all the available strains was done. All these subtelomeric regions were manually annotated (as mentioned before) and the result is shown in the Supplemental Figure 2. The copy that we named *FSY1B1* was found in *T. microellipsoides* CLIB 830T, NRRL Y-17058, CBS 6143 and CBS 6641. The second paralogous copy was present in the genomes of NRRL Y-17058, CBS 6143 and CBS 6762. In the first two strains, *FSY1B1* and *FSY1B2* lay in homologous regions, a result that corroborates the hypothesis stated here that these two strains arose from ancestral hybridisation. In CBS 6762, there was only one *FSY1* gene similar to the *FSY1B2* copy, while CBS 6641 and CLIB 830T presented two *FSY1* genes: one that was nearly identical to Region C and a second gene that was a copy of *FSY1B1* (see the *FSY1* phylogeny).

The *FSY1* paralogous copies were searched before inside the *T. microellipsoides* genome assemblies, by a *Southern* blot hybridization approach (Figure 2.2) to locate the chromosomic position of each *FSY1* duplicated gene. In this way, two oligonucleotide probes were designed (see Methodology) and their specificity was checked by Sanger sequencing. Both *FSY1A* and *FSY1B1*/*FSY1B2*, were detected in different chromosomes, which supported the idea that they were paralogous copies, but not gene alleles. The origin of these two copies was suspected when the nucleotide divergence between them was calculated. These results were after confirmed by the genome sequencing of all the *T. microellipsoides* strains. However, the chromosomes hybridisation revealed a secondary result: new copies of the *FSY1* gene were interspersed among the genome of some *T. microellipsoides* strains. Thus, the *FSY1A* copy was detected in two different chromosomes of the CLIB 830T strain, but no evidence for this second copy was obtained with the *Illumina* assemblies. Moreover, the absence of the *FSY1A* copy in CBS 6762 was first detected by this approach. Both the *FSY1B* copies found, *FSY1B1*/*FSY1B2*, were first detected by genome sequencing and later by the *Southern* blot hybridisation by lowering stringency conditions. The large number of *FSY1* genes and other Region C genes (also detected by the *Southern* blot hybridization, results not shown) reinforced the notion that this non-*Saccharomyces* species would be responsible for spreading out all these genes to the *Saccharomyces* species.

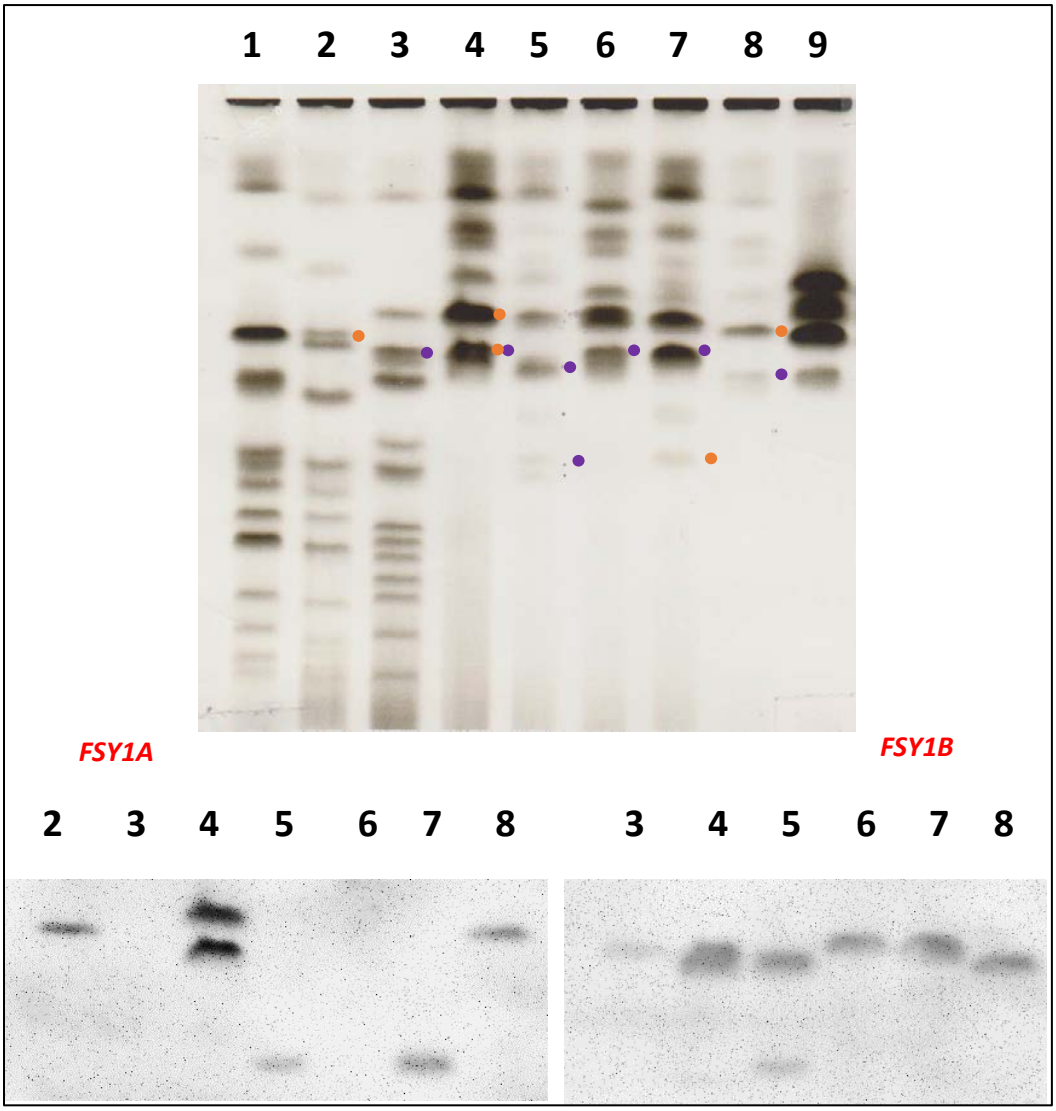


Figure 2.2. **The *FSY1* gene *Southern Blot* hybridization on the *T. microellipsoides* strains**. Upper panel: a PFGE gel containing karyotypes from *Saccharomyces* and non-*Saccharomyces* species. **1**: *S. cerevisiae* S288c, **2**: *S. cerevisiae* EC1118; **3**: *S. uvarum* BMV58, **4:** *T. microellipsoides* CLIB 830T, **5**: *T. microellipsoides* NRRL Y-17058, **6**: *T. microellipsoides* CBS 6762, **7**: *T. microellipsoides* CBS 6641, **8**: *T. microellipsoides* CBS 6143 and **9**: *T. delbrueckii* CECT 10039. The S288c and EC1118 *S. cerevisiae* strains are included as the negative and positive control species carrying the *FSY1* gene. The *T. delbrueckii* species is included to see probe specificity. Bottom panel: from the left to the right, we show the detection of *FSY1A* and *FSY1B*/*FSY1B2* paralogous genes. The orange points indicate the chromosome bands in which the *FSY1A* copy is detected. The purple points denote the bands where *FSY1B* and *FSY1B2* are found.

It was noted that the *T. microellipsoides* regions harboring the *FSY1B1* or *FSY1B2* copies presented a conserved synteny with *T. delbrueckii* and *Z. rouxii* along the core region. Suddenly however, the synteny was lost in the subtelomeric region, and *FSY1* was the last syntenic gene included, but the first included in the subtelomeric part (based on the definition by Yue et al. 2017 of the core and subtelomeric genome zones). From the *FSY1B1*/*FSY1B2* genes towards the telomeres, gene composition was extremely variable between species. The same degree of synteny conservation was observed for a *S. uvarum* strain, which was took as an example. In this species, the *FSY1* gene was found in a different chromosome, which thus broke the observed synteny in the non-*Saccharomyces* species. In view of these results, it was proposed here that the fructose symporter gene could be lost in chromosome VIII of the *S. eubayanus* species cluster and it could be later recovered by two open possibilities. Firstly, if this gene was present in the WGD ancestor and was then lost in all the post-WGD genera except in the *Saccharomyces* genus, then it could jump from chromosome VIII to chromosome IV. Alternatively, it would be acquired by introgression from a *T. microellipsoides* strain in the *S. eubayanus* ancestor.

Inside the *T. microellipsoides* strains, subtelomeric gene diversity started after three contiguous genes to *FSY1B1*/*FSY1B2*: *GTT1*, *SAM4* and *SAM3*. It is note worthy that additional copies of Region C were found with in these gene variable extremes. In the *FSY1B1*-carrying region, an extra copy of *DSF1* was located as the last assembled gene, while we annotated a similar *PUT3* copy from a bacterial origin gene in the *FSY1B2* containing scaffold, specifically a probable amidohydrolase from *Acinetobacter baumanni*.

As previously evidenced in the *ATO3* phylogeny, every *T. microellipsoides* isolate presented at least one duplicated gene of that found in Region C, which was named "*ATO3* RC" in the corresponding phylogeny. The nucleotide similarity between paralogous genes was high, so it was suggested here that the duplication event of these hypothetical ORFs could have recently arosen in time. The degree of similarity can be seen in the tree for the CLIB 830T and CBS 6641 strains.

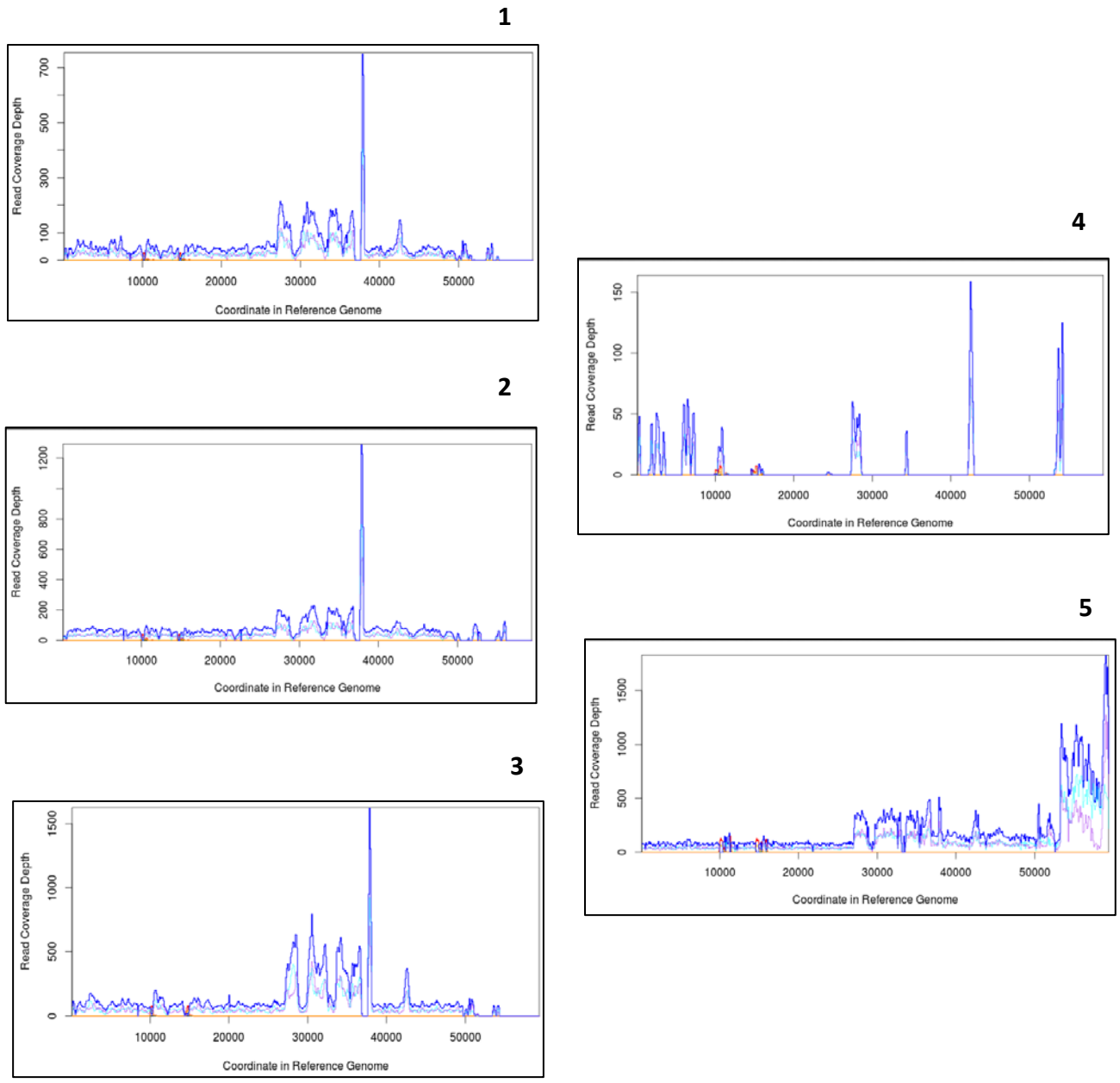


Figure 2.3. The coverage profiles that result from mapping the *T. microellipsoides* reads. Region C from the EC1118 strain (65kb) was employed to map from the sequencing reads. The X-axis presents the nucleotide positions of the sequence from the *ARB1* gene to the telomere. The Y-axis presents the coverage reads that differ in each strain because of sequencing quality. **1**: CBS 6143, **2**: CBS 6641, **3**: NRRL Y-17058, **4**: CBS 6762 and **5**: CLIB 830T.

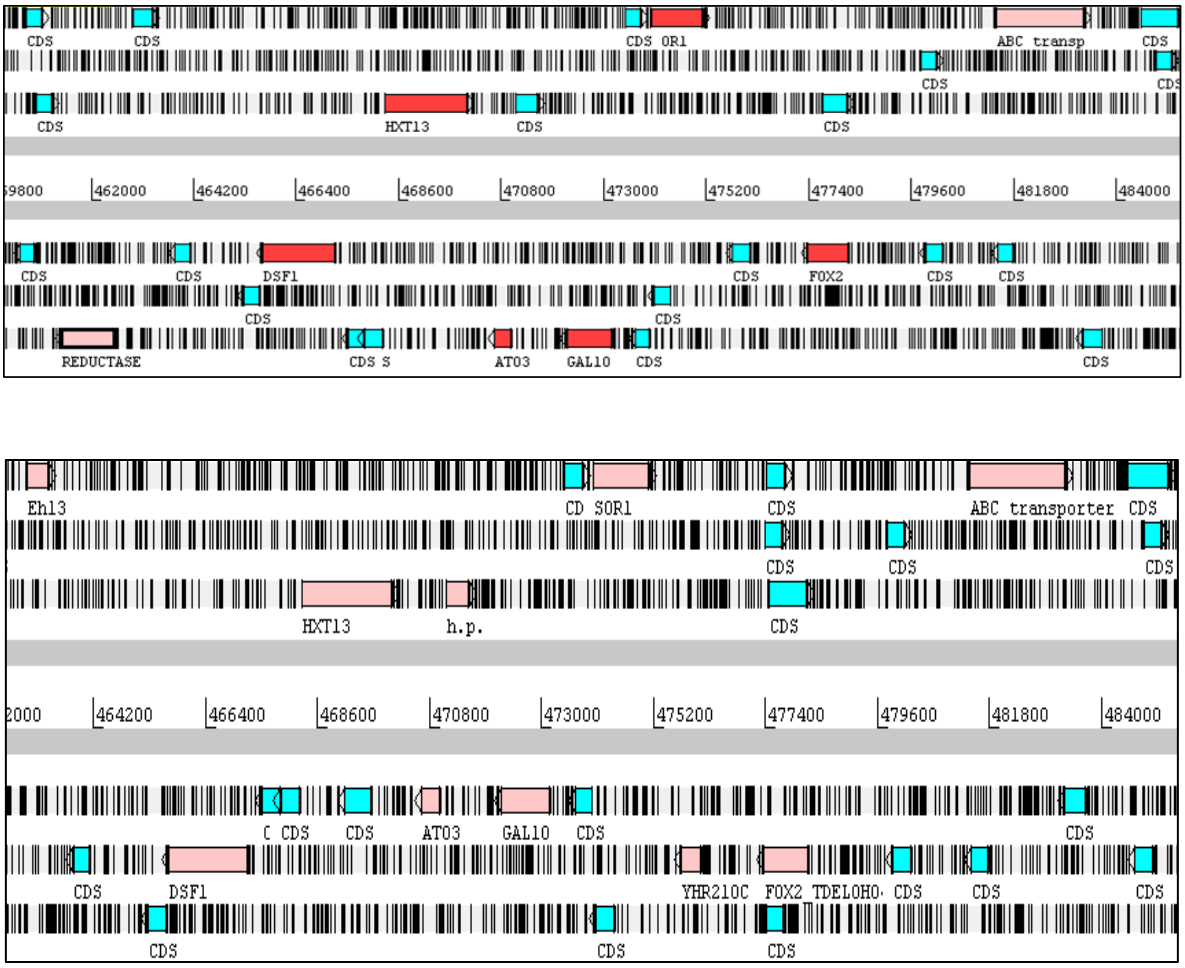


Figure 2.4. The reduced Region C found in some *T. microellipsoides* strains. Strains NRRL Y-17058 (upper panel) and CBS 6143 (bottom panel) each presented a little region composed of genes before being found in a central part of Region C. From left to right, these genes are: *DSF1*, *HXT13*, *ATO3*, *GAL10*, *SOR1* and *FOX2.* Numbers refer to the genomic position of the assembled scaffolds. The ORFs annotated over number positions are in the 5' to 3' sense (Watson strand), while the ORFs below the numbers are in 3' to 5' (Crick strand). Screenshots are captured from the annotations performed in the Artemis software.

2.2.3 Phylogeny of 1,844 *Torulaspora microellipsoides* genes

The results obtained first by the *FSY1* phylogeny reconstruction suggested that two hybrid species probably existed inside the *T. microellipsoides* species and their likely parental species were deduced from the same analysis. To support this hypothesis, all the possible numbers of annotated genes were collected after genome sequencing to perform a new phylogeny for the species. The Yeast Genome Annotation Pipeline for genome annotation (YGAP) was first employed, as described in the Methodology. Once genomes were annotated, it was filtered 1,844 genes that were concatenated to reconstruct the *T. microellipsoides* phylogeny. This number corresponds to the annotated ORFs with no N ambiguities (any base) in the nucleotide sequence and with no sequencing errors. For the predicted hybrid species, subgenomes A and B were obtained by mapping their reads against their parental species inferred from the *FSY1* phylogeny.

The resulting phylogeny (Figure 2.5) shows two clearly distinct groups with a similar branch length from their origin (their common ancestor). In the upper part of the tree a group is located that includes the CBS 6762 and NRRLY-17058/CBS 6143 **B** sequences. A second monophyletic group is composed of the CLIB 830T strain and the NRRLY-17058/CBS 6143 **A** sequences that emerge from the CBS 6641 strain. Bootstrap values are 100 for each node, which indicates the strong support of this reconstruction.

Therefore, with this analysis it was confirmed what it was deduced before for the gene specific trees: two new non-*Saccharomyces* hybrid species (NRRLY-17058 and CBS 6143) and, probably, their corresponding parental species (CBS 6641 and CBS 6762) were definitely found. One of them, CBS 6762, clearly diverged from the other *T. microellipsoides* strains and clustered only with the assumed

hybrid subgenomes. From this unexpectedly obtained relationship, it was hypothesised that CBS 6762 could be a different undescribed species

The conclusions about the discovery of the new and the two hybrid species obtained were based on the phylogenetic species definition. Sporulation, conjugation and spore viability assays would be necessary to confirm the existence of different species based on their biological definition. In this way, *T. microellipsoides* under the study strains were sporulated, but despite our ample lab experience, the small spore size (between 2-5µM) hampered their isolation. Nevertheless, strains were grown in sporulation media (see the Methodology) and then were placed under an optical microscopy to see vegetative cells. Photographs are shown in Figure 2.6. Ascospore formation was observed for all the strains and the asci contained between 1 to 4 spores each. Many images where independent cells conjugated were captured except for the CBS 6143 strain. Therefore, it was detected that at least two different cells can carry out conjugation inside the same strain. So outcrossing, which is the conjugation between haploid cells from different species, is likely to occur.

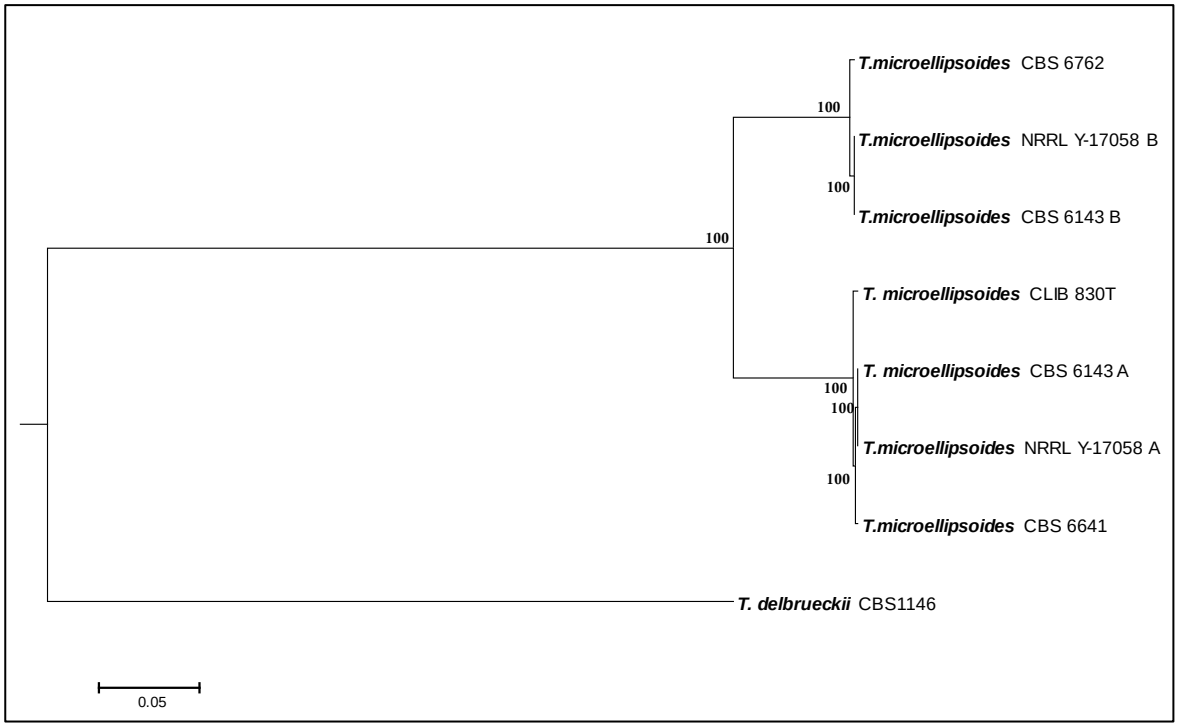


Figure 2.5. **The neighbour-joining phylogeny depicting the relationships between *T. microellipsoides* species**. We included an outgroup species from the same genus, *T. delbrueckii*. Strains NRRL Y-17058 and CBS 6143 are clustered in two monophyletic groups because their genome sequence was split into two subgenomes: A and B. Support values are presented at each node, and correspond to 100 bootstrap replicates. The scale is given in nucleotide substitutions per site.

The evolutionary divergence between the *T. microellipsoides* strains and between each inferred *subgenome* was estimated from the 1,844 concatenated gene sequences (Table 1). Lower nucleotide change values corresponded to the comparisons made between the same *subgenomes* of each hybrid species. The hybrid *subgenomes* named **A** presented a percentage of nucleotide divergence of one change in every 100000 nucleotides between them and less divergence was seen between the hybrid *subgenomes* named **B** (nearly one change in every 1000000 nucleotides). These minimal values suggest a common origin for both hybrids, a result that complements the topology shown in the phylogeny (Table 1). Type strain CLIB 830T differed from the CBS 6641 strain and from hybrids sequence A by 0.5% while fewer nucleotide changes were observed between CBS 6641 and this last hybrid sequences, which was over 0.25%, which meant half one part of the CLIB 830T distance. By taking these values, it was confirmed what was before observed in *FSY1* and the *T. microellipsoides* phylogeny: probably CBS 6641, but not CLIB 830T, could be one of the parental strains, and could contribute with the part of the genome named A. Fewer changes ocurred between the CBS 6762 strain and the hybrid *subgenomes* named B than in CLIB 830T, but more took place than in CBS 6641 (0.40%), when the comparison of the A-*subgenome*was observed. By taking these values together with the phylogenetic position observed in the species tree, it was concluded that the parental origin of the part B subgenome could be attributed to strain CBS 6762 or to a strain close to it.

Table 2.1. The measured divergence between the *T. microellipsoides* genomes conducted in Mega *v*6 (Tamura et al. 2013). The numbers in the box correspond to the number of base differences per sequence from among the shown strains. The analysis included eight sequences and all the codon positions. All the ambiguous positions were removed for each sequence pair. The final dataset had 2785314 positions. Groups of sequences were differentiated by colors to better interpret the results (Blue: subgenome A and CBS 6641, Purple: subgenome B and CBS 6762; Pink: Outgroup species). The *T. delbrueckii* species is included as the outgroup sequence.

	Tm_CLIB_830	Tm_CBS_6641	Tm_NRRL Y-17058_A	Tm_CBS_6143_A	Tm_CBS 6762_B	Tm_NRRL Y-17058_B	Tm_CBS_6143_
Tm_CBS_6641	13075						
Tm_NRRL Y-17058_A	13432	6495					
Tm_CBS_6143_A	13391	6467	18				
Tm_CBS 6762_B	272401	271907	272522	271993			
Tm_NRRL Y-17058_B	272640	272190	272808	272272	11180		
Tm_CBS_6143_B	271777	271332	271959	271445	11169	6	
Td_CBS_1146	852244	850835	852120	850924	849210	849892	847293

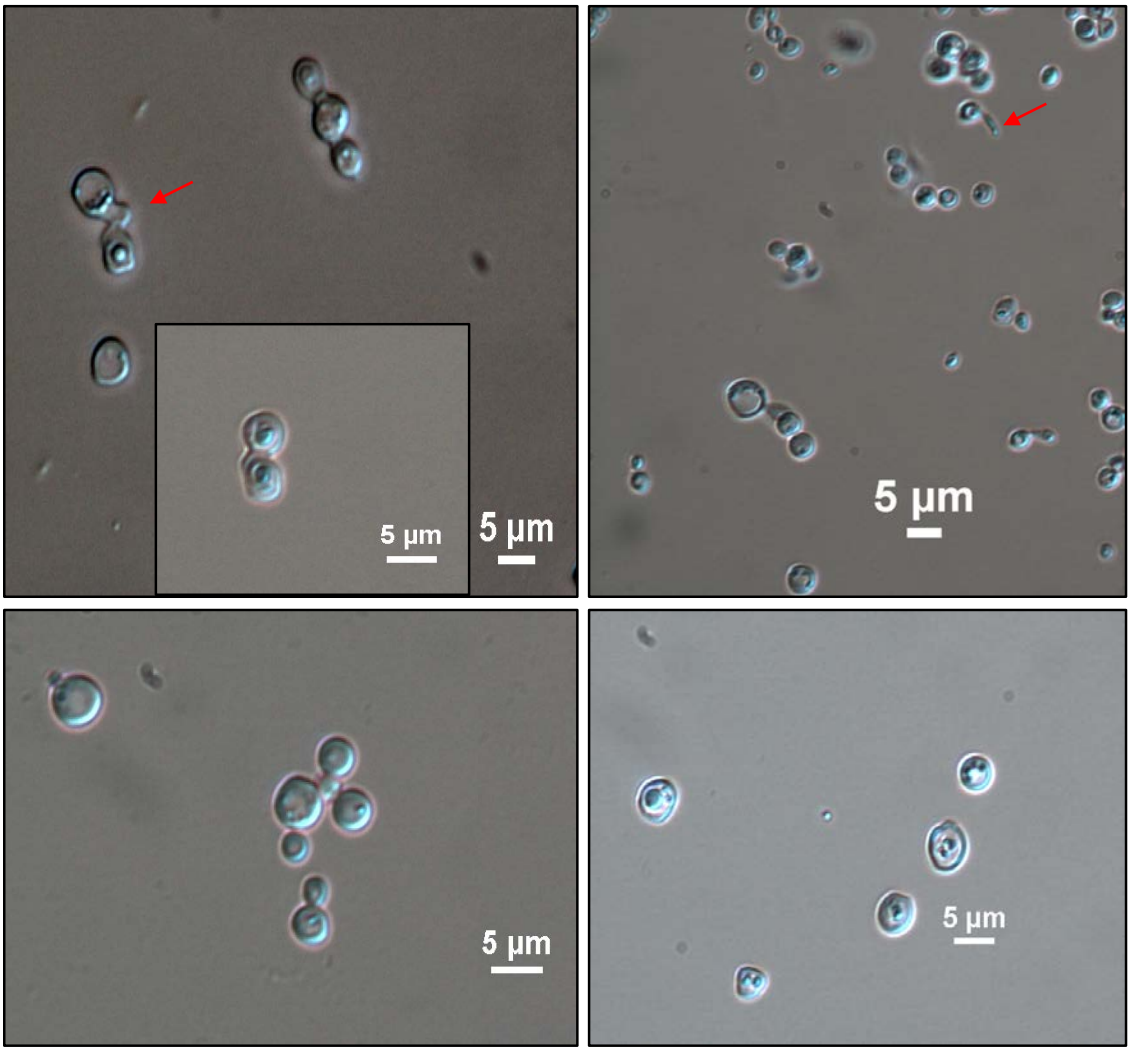


Figure 2.6. Photographs obtained by optical microscopy showing *T. microellipsoides* cells after growing for 15 days on YM restrictive culture media. Panel A (NRRL Y-17058): Two individual cells are conjugating and a "conjugation bridge" between them is observed (red arrow). Cells present single spherical ascospores. Panel B (CLIB 830T): conjugation events between independent cells and probable asci with a tapered protuberance (red arrow) emerging from asci. Panel C (CBS 6641): conjugation between two cells with one spherical ascospore each. Panel D (CBS 6762): different conjugation events. The spores observed in these strains were small in size and only one spore is usually observed. Panel E (CBS 6143): asci showing different spore number (from 1 to 4). No conjugation between independent cells was captured. Bars indicate 5 µM.

2.2.4 Phylogeny of the *Torulaspora* genus

The type strain sequences of the new *Torulaspora* species described in the bibliography were employed to resolve the genus phylogeny, including our *T. microellipsoides* sequences (Figure 2.7). These species were: *T. maleeae*, *T.globosa*, *T. indica*, *T. delbrueckii*, *T. franciscae*, *T. pretoriensis*, *T. quercuum* and the *T. microellipsoides* CLIB 830T strain. In addition, for the analysis those *T. microellipsoides* strains that seemed to be the parental species of the previously identified hybrids were included.

The aim of this analysis was to support the previously obtained results about the phylogenetic relationships among *T. microellipsoides* species. The same phylogenetic reconstruction method as in Saluja et al. 2012 was employed to compare the resultant phylogenies. Hence a **neighbour-joining** tree was reconstructed with the same nucleotide sequences used there. Then a multi gene alignment of the 26S rDNA gene and the 18S/ITS/5.8S rDNA cluster was performed.

From the resulting tree, interesting aspects were observed concerning the *Torulaspora* relationships. For the non-*T. microellipsoides* species, the branch length distance between *T. globose* and *T. indica* was not significatively long and this nucleotide distance was, over 3-7% on average, as assigned in Saluja et al. 2012. This short distance, together with the absence of spore-spore conjugation assays, led to some doubts about these two sequences belonging to two different species. A similar situation probably occurs with *T. franciscae* and *T. pretoriensis*, whose rDNA nucleotide sequences are highly related.

When the topology shown for the *T. microellipsoides* species was inspected, a recent common ancestor for strains CBS 6641 and CLIB 830T could be suggested from the observed phylogeny, exactly as expected for the results of the *FSY1* and *ATO3* phylogenies. Strain CBS 6762 was easily differentiated from these closely related strains by presenting a distant phylogenetic position from them, but similarly to that found between *T. globose* and *T. indica*. However, *T. microellipsoides* and the CBS 6762 strain formed a monophyletic group with a common ancestor species that diverged more quickly compared to the other species.

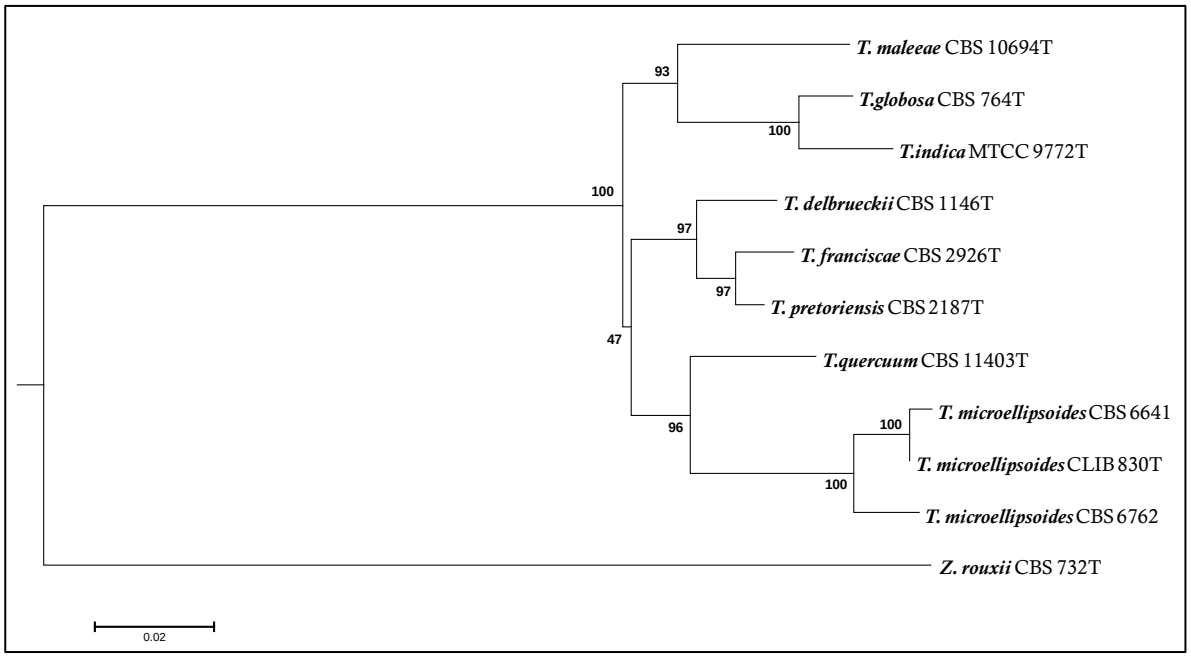


Figure 2.7. The *Torulaspora* genus phylogeny. The neighbour-joining analysis of a complete rDNA region that comprised 26S rDNA and the 18S/ITS/5.8S gene and intergenic conserved sequences. Bootstrap values are represented in each node over 100 replicas. The tree is rooted in the *Z. rouxii* species that belongs to the *Torulaspora* closest genus. The scale is given in nucleotide substitutions per site.

Despite a probable new species (CBS 6762) being inferred from the genus tree, these analyses were repeated by including the nucleotide sequences of all the strains classified as *Torulaspora* sp. and isolated to date to see if we could find any close relationships with the former strain and these new sequences. Strain sequences from the *genbank* database were retrieved, classified either in one of the novel *Torulaspora* species or solely included in the genus (as *Torulaspora* sp.). The 26S rDNA phylogeny (Figure 2.8) was obtained separately from that of 18S/ITS/5.8S rDNA (Figure 2.9) because some strains presented only one of the rDNA sequences.

The 26S sequences for the five strains identified before as *T. microellispoides* strains were obtained by Sanger sequencing because of the difficulty of assembling the rDNA repeated regions. The 26S rDNA genes from CBS 6762 and NRRL Y-17058 were identical, but the same gene in the other assumed hybrid species was identical to CBS 6641. Thus, each hybrid strain inherited one of the rDNA tandem repetitions from one different parental strain. The *Torulaspora* sp. strains included in this reconstruction were 13 from a total of 72, which were quite distanced (looking at their branch length) from the sequences in which the species level was determined. This suggests that some could be addressed as a new species.

Bootstrap values were generally low, perhaps because these long-branched strains were interspersed between the others. Those supporting values were frequently located at the ancestral nodes of these unclassified strains. The relationships of *T. pretoriensis*, *T. franciscae* and *T delbrueckii* were slightly altered compared to the phylogeny obtained for the *Torulaspora* genus (Figure 2.5). The phylogenetic group formed between *T. globosa*/*T. indica* and *T. maleeae* was maintained as with *T. quercuum* and *T. microellipsoides*.

The results obtained for the tree based on 18S/ITS/5.8S rDNA showed that the relationships between species were maintained with respect to the tree obtained for the *Torulaspora* genus. Six *Torulaspora* sp. strains were included in the analysis and of them, three showed a long branch distance compared to their closest species. One of them, *Torulaspora* sp. SG5S08, diverged first from the *T. pretoriensis* species group and the other two, *Torulaspora* sp. EN11S09 and *Torulaspora* sp. EN22S16, presented a long distance from the *T. maleeae* group. All three strains were also included in the 26S rDNA tree, but phylogenetic relationships were shared only by *Torulaspora* sp. EN11S09 and *Torulaspora* sp. EN22S16. Strain *Torulaspora* sp. SG5S08 was related to *T. quercum* species in the 26S rDNA reconstruction instead of being related to *T. pretoriensis.*

The 18S/ITS/5.8S rDNA phylogeny was perfomed only with the CLIB 830T, CBS 6641 and CBS 6762 *T. microellispoides* nucleotide sequences obtained from the CBS strains databases. The reason for this lies in the fact that duplicated genomes of the probable hybrid species hindered the correct assembly of this repeated regions. The topology resolved by the group was the expected one, with related CLIB 830T and CBS 6641 strains, while strain CBS 6762 remained at a distant position.

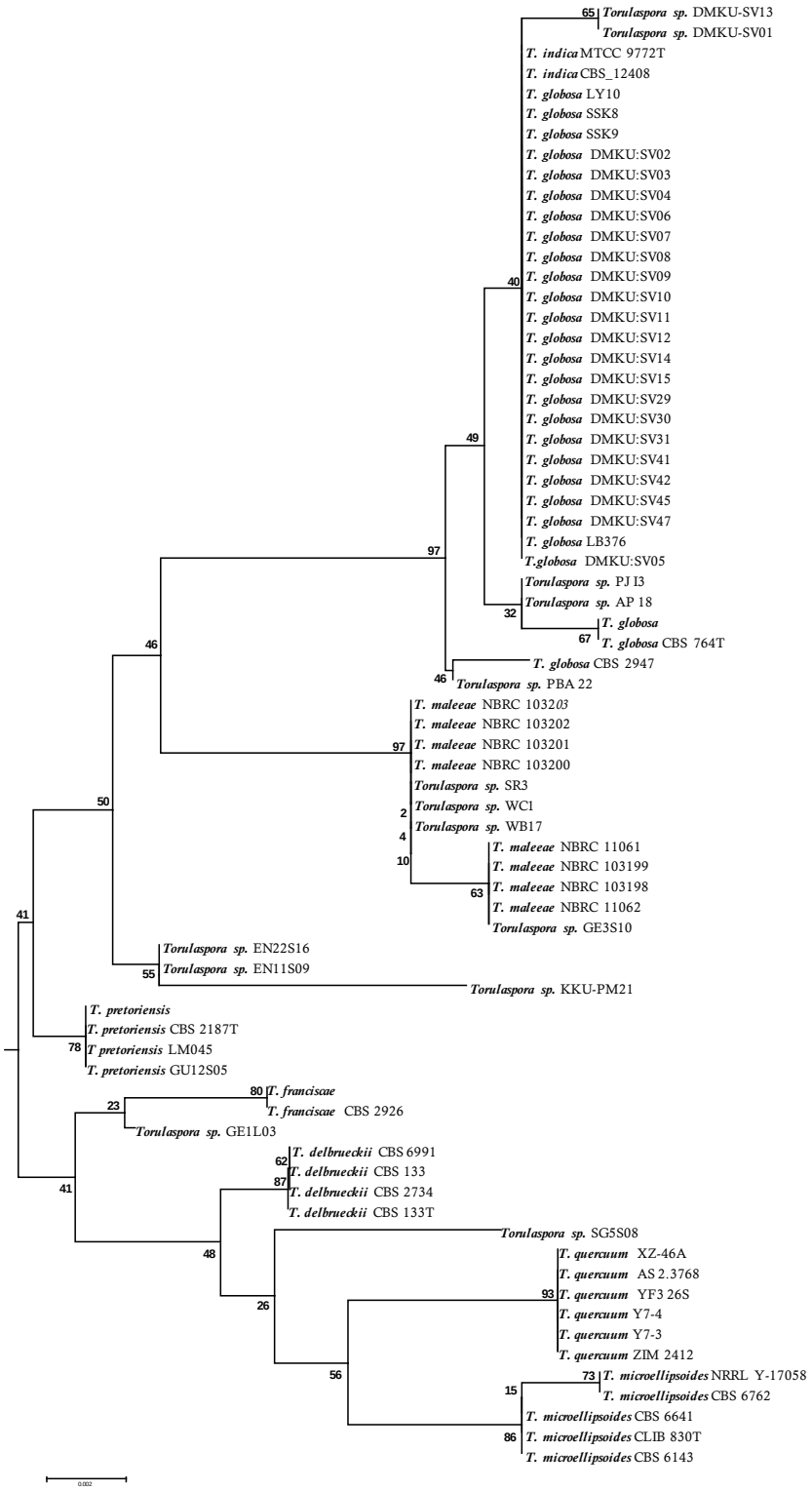


Figure 2.8. Neighbour-joining phylogeny employing the 26S rDNA nucleotide sequences found to belong to the *Torulaspora* genus. Sequences were retrieved from *genbank,* except for the *T. microellipsoides* sequences obtained by Sanger sequencing. Bootstrap values are represented by numbers in each node over 100 replicas. The scale is given in nucleotide substitutions per site.

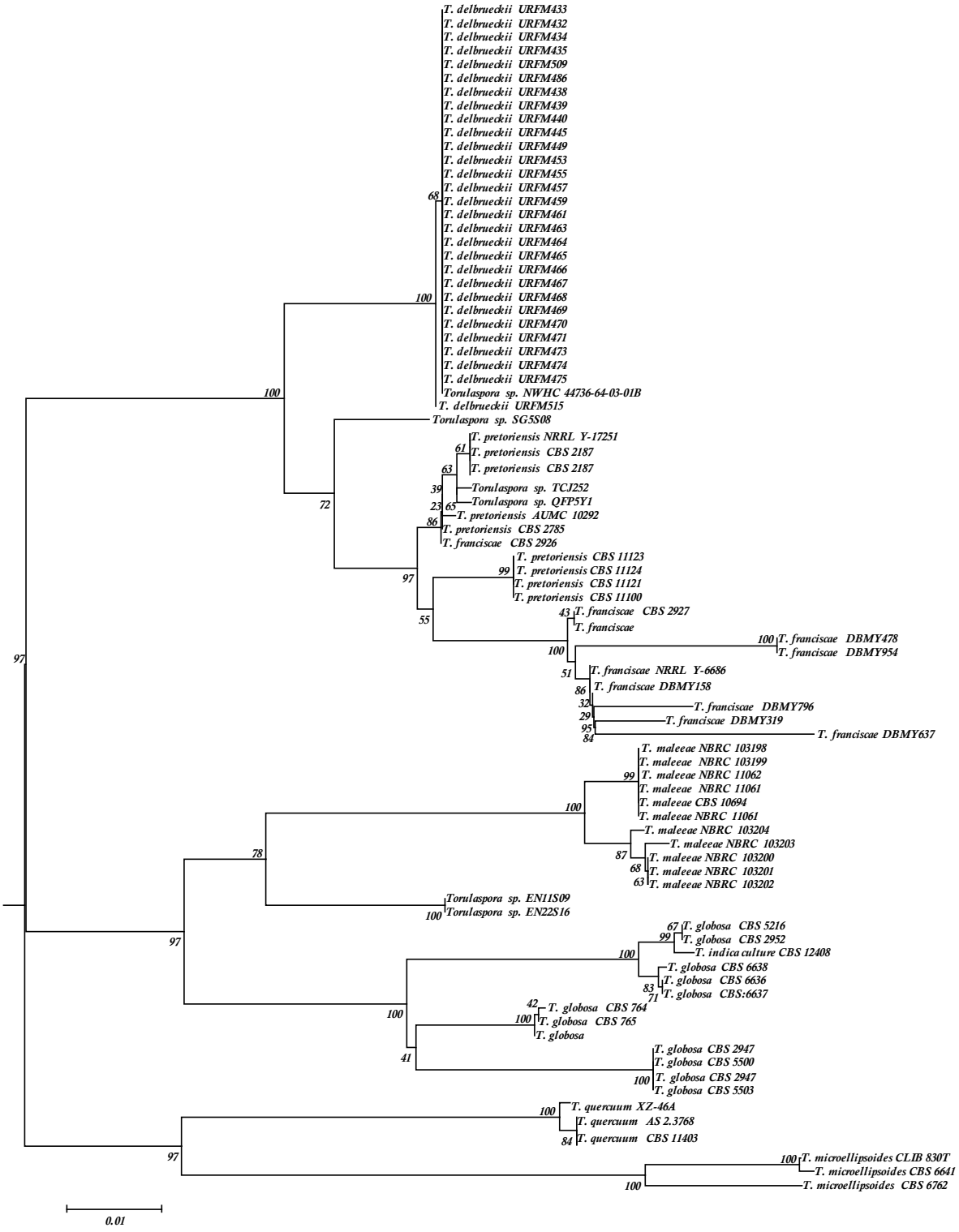


Figure 2.9. Neighbour-joining phylogeny obtained by using the 18S/ITS/5.8S rDNA sequences of them all as *Torulaspora* classified strains. Sequences were extracted from *genbank*, except for the *T. microellipsoides* sequences obtained from the CBS yeast culture collection. Bootstrap values are represented by numbers in each node over 100 replicas. The scale is given in nucleotide substitutions per site.

2.2.5 Pseudo-chromosomes representation of the *T. microellipsoides* hybrid**s**

Even though the first genome of *T. microellipsoides* was formerly sequenced (that from the CLIB 830T strain), no annotation for it existed. So no available reference genome existed against which mappped our reads extracted by the genome sequencing of the remaining isolates. A different chromosome pattern for each *T. microellipsoides* strain by the karyotyping analysis was also observed (see the Methodology). So, it was impossible to assign the scaffolds obtained from the assemblies at any *T. microellipsoides* chromosome. (Figure 2.9).

As previously mentioned, an annotation was carried out by YGAP, and the CLIB 830T, CBS 6641 and CBS 6762 genomes were annotated by this tool. Then, it was took the sequence annotations from the CBS 6641 and CBS 6762 strains to map the NRRLY-17058 and CBS 6143 reads against, as we inferred that they were probably parental strains from the deduced hybrid species. From mapping, we constructed eight *in silico pseudo-chromosomes* (see the Methodology) for the hybrid species to observe the genomes retained from each parental strain and to infer their origin.

The coverage observed after mapping the reads against both inferred parental strains (CBS 6641 and CBS 6762) was frequently constant along the artificial chromosomes, but in some cases it was seen increasing or decreasing coverage (Figure 2.9). Both the NRRL Y-17058 and CBS 6143 strains showed a similar mapping coverage on the 2, 5, 6 and 8 pseudo-chromosomes. This equivalence could be interpreted as both hybrid species having a common origin, and this aspect could also be deduced from the *T. microellipsoides* phylogeny contructed before.

Some chromosomic duplications or delections were observed for the other pseudo-chromosomes, which means that these hybrid strains diverged from each other; e.g. for the reads mapped against ancestral chromosome 1, it was observed an equivalent distribution of the reads between both parental sequences for the CBS 6143 strains. However for strains NRRL Y-17058, a deletion of the reads mapped to CBS 6641 overlapped a duplication of the number of reads mapped against the CBS 6762 strain. Hence, a chromosomic fragment belonging to the CBS 6641 parental strain was probably replaced with its homologous region of CBS 6762 during the hybridisation event. The inverse situation was found for the other altered-in-coverage pseudo-chromosomes, where the reads mapping the CBS 6641 sequences were duplicated and the reads mapping the CBS 6762 strain showed an overlapping deletion. A similar result was obtained for pseudo-chromosome 4 for the CBS 6143 and NRRL Y-17058 strain at the same chromosomic position. Moreover, strain NRRL Y-17058 presented two duplicated regions that came from the CBS 6641 strain in pseudo-chromosomes 3 and 7.

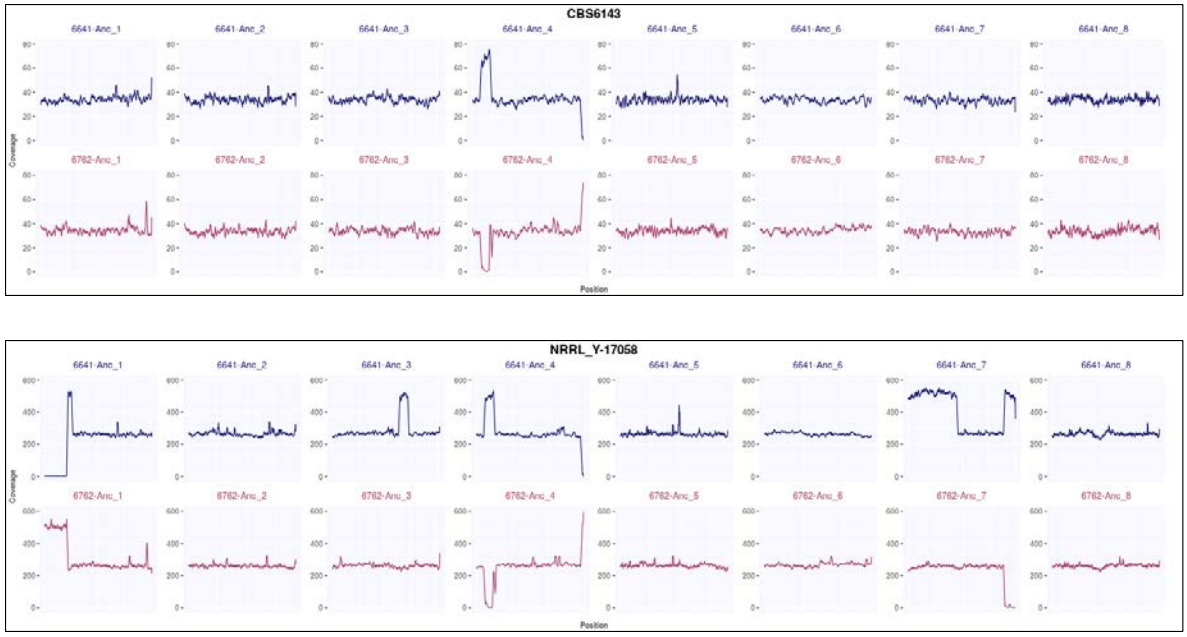


Figure 2.10. The *pseudo-chromosomes* reconstruction of the *T. microellipsoides* hybrid species. From the 26S rDNA *T. microellipsoides* phylogeny, we concluded that the probable parental species that gave way to hybrid strains NRRL Y-17058 and CBS 6143 were CBS 6641 (in blue) and CBS 6762 (depicted in red). Mapping was done against a set of eight artificial chromosomes. A 30% coverage for CBS 6143 and one of 250% in NRRL Y-17058 were established as the base coverage, and any changes over or below these values were taken as duplications or deletions, respectively.

2.2.6 Karyotyping composition of the *T. microellipsoides* strains

Additional analyses to the genome sequencing about, genome size composition were attached to corroborate our hypothesis for hybrid genomes. The chromosome composition of the *T. microellipsoides* strains was obtained from a PFGE (Pulse Field Gel Electrophoresis) analysis and the exactly size of bands was quantified by using *Bionumerics v*7.6 image software (http://www.applied-maths.com/news/bionumerics-version-76-released) including two *S. cerevisiae* reference strains: S288c and EC1118 (see the Methodology for details). The output measures that this software empoys are shown in Supplemental Figure 5. From two sets of samples, including these seven strains each, the sizes of the detected bands were obtained. The results were manually inspected to verify if the programme ignored any band that was observed in other images. As a good resolution of some bands frequently depended on gel electrophoresis, assays were successively repeated until we thought that they had all been solved. An example of a PFGE image was provided by indicating the bands (the green and red points in the Figure 2.10) that were found by looking at different images. As previously mentioned, employing this software to quantify bands, some were not detected (red dot in the bands in Figure 10).

Nine chromosome bands were counted for the CLIB 803T, CBS 6641 and CBS 6762 strains. The phylogenetic analysis suggested a common origin for the first two strains, but the band sizes between them did not share the same sizes at all. CBS 6762 was expected to be a distinct species (as deduced from the phylogenies shown before). However, its chromosome pattern was quite similar to that seen for strain CBS 6641.

In NRRL Y-17058 and CBS 6143, 14 and 13 bands were respectively resolved from the images. For these two strains, the number of chromosomes increased compared to the other *T. microellipsoides* strains. According to the hypothesis stated here it was expected that these strains to probably be two hybrid species between CBS 6641 and CBS 6762. Perhaps some duplicated chromosomes could have similar sizes after hybridisation, but it was not possible to differentiate if some bands corresponded to one chromosome or to a pair of chromosomes. Some other chromosomes likely underwent recombinations or insertions/deletions that modified their initial size, as it was observed in the pseudo-chromosomes representation. The banding pattern of both strains showed differences so, in spite of a common origin being suggested by the phylogenetic analysis for these probable hybrid species, they diverged from each other to accumulate some of these recombinations. This aspect was observed in the karyotype analysis and in the pseudo-chromosome reads mapping.

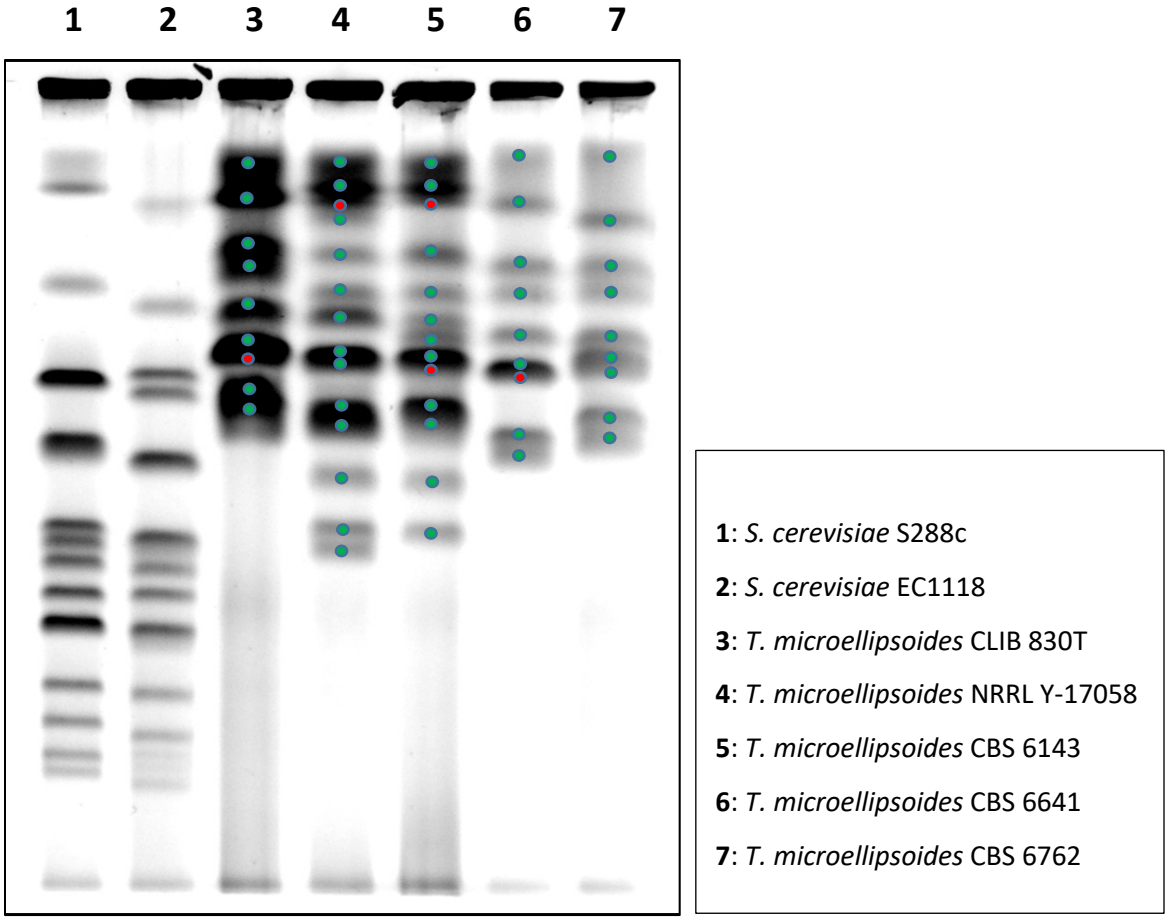
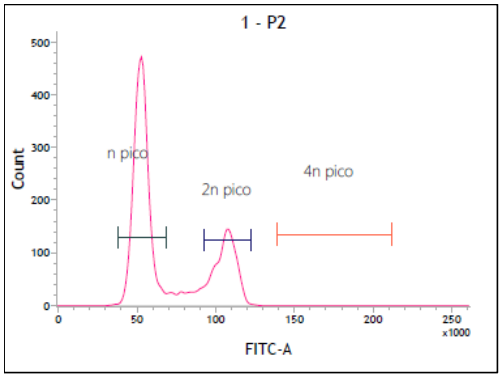
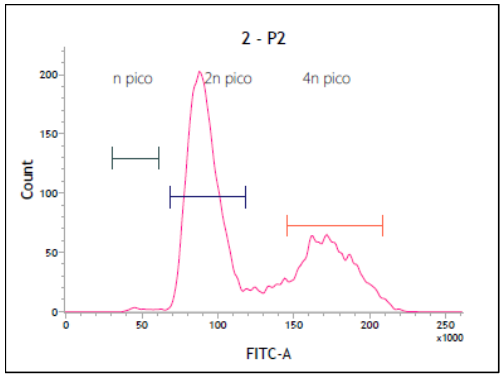
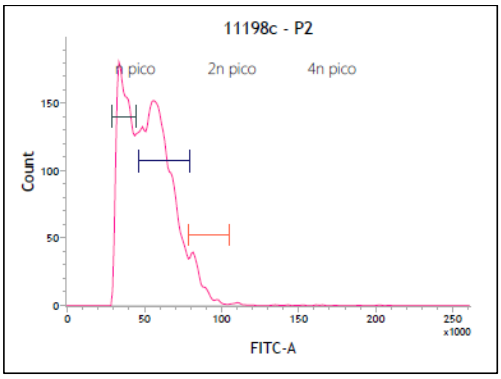


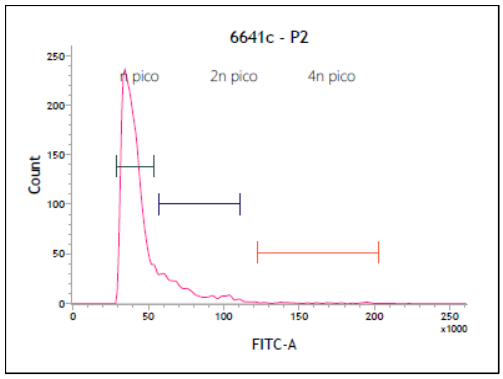
Figure 2.11. The karyotype pattern of the *T. microellipsoides* species. Two *S. cerevisiae* species (1 and 2) were employed to obtain a band size reference. The green circles correspond to the chromosomes detected with the *Bionumerics v*7.6 image software. The red circles show the bands that the programme was unable to detect, but we were able to differentiate after several PFGE replicas.

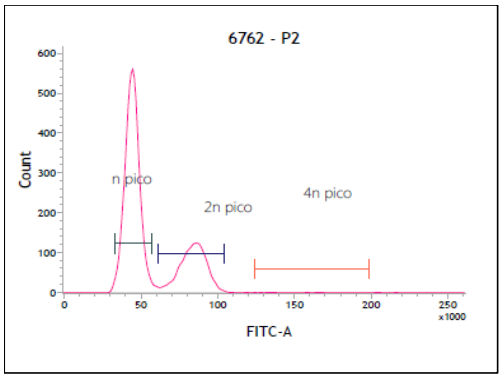
2.2.7 Measured ploidy by flow cytometry

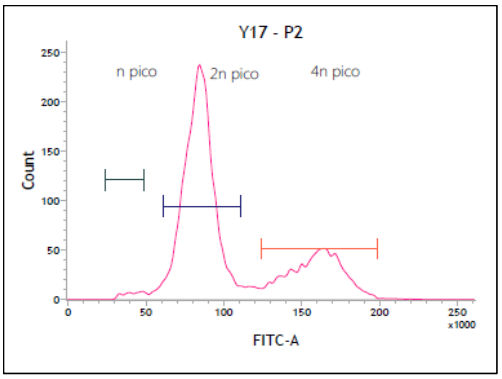












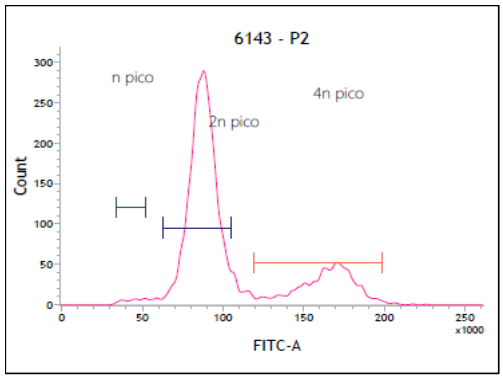


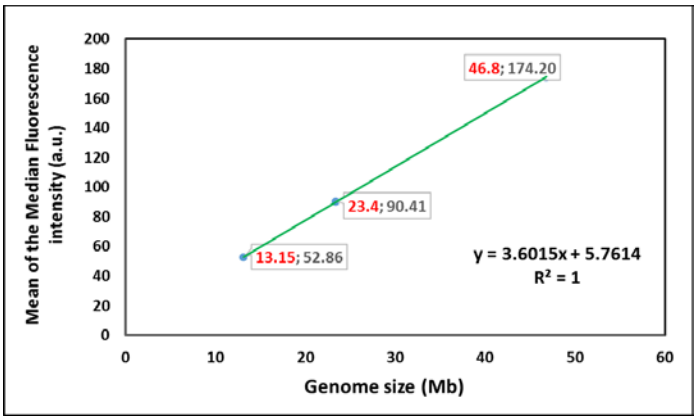
Figure 2.12. The cytometry profiles of the *T. microellipsoides* strains measured by flow cytometry. The fluorescence emitted by an intercalating dye was determined by the picks that appeared on the axis of the count of events (y) and FITC-A (x). We denoted the G0/G1 cell cycle phase as "*n pico"*, and the G2/M cell cycle phase as "*2n pico"* or "*4n pico"*. Box 1 contains the fluorescence spectra of the perfect haploid strain *S. cerevisiae* S288c and Box 2 contains the fluorescence spectra of the perfect diploid strain *S. cerevisiae* FY167. The other boxes contain the fluorescence signal of each *T. microellipsoides* strain. Unsynchronised cell cultures were prepared for the analysis and three technical replicates were included. Picks were sometimes reviewed manually to make a correct association with the cell cycle phase. The fluorescence median values were taken to calculate the genome content from three independent replicas.

The genome size of the *T. microellipsoides* strains was calculated by a flow citometry assay. A total of 1x 10^7 cells were harvested from the unsynchronised YPD cultures grown O/N. Under these conditions, the fluorescence concerning the G0/G1 celular cycle phase was more frequently detected than that from the G2/M phase, which allowed us to determine the real DNA content. Samples were ethanol-fixed and stained with intercalating fluorescent molecule *SITOX Green*. As the of-reference ploidy *T. microellipsoides* strains were unavailable, we compared the fluorescence of our five species with the *S. cerevisiae* S288c and FY1679 reference genomes. A similar approach has been employed by Solieri et al. 2008, who worked with *Z. rouxii* strains.

The profiles obtained for each strain varied from one another (Figure 2.11). So in some cases, it was necessary to manually define the amplitude of the released picks; e.g. while the profiles of the *S. cerevisiae* reference strains and the CBS 6762, NRRL Y-17058 and CBS 6143 *T. microellipsoides* strains showed well-defined picks, CLIB 830T and CBS 6641 presented a more continuous fluorescence profile. Those differences were attributed to the limited growth of these species compared to those with defined profiles. The cultures grown O/N from one isolated colony resulted in a final optical density that was too low for CLIB 830T and CBS 6641 compared to the remaining strains. Although the growth rate was not assayed, the independent cultures of this species group always gave rise to deficient growth for those two strains. From the flow cytometry profiles, it was observed that CLIB 830T showed three overlapped picks, which could be defined by merely looking at the technical replicas. The haploid content was assigned to the less emitting pick and the diploid content to that where fluorescence was duplicated *versus* the former. The pick presented in the middle could involve the cells in the S phase of the cell cycle, the DNA duplication phase. This would suggest that this process could slow down, at least under our growing conditions.

However, strain CBS 6641 presented one clear pick, followed by a discontinuous area of fluorescence. This behavior could be attributed to a probable fast shift to the haploid state of the cell with an non-detectable G2/M phase and a long-time prevalence in the G0/G1 phase (which agrees with the deficient growth seen in the YPD-rich media at 30^0C).

The fluorescence value of the S288c G0/G1 phase, and those of the FY1679 G0/G1 and G2/M phases, were included to represent a correlation *versus* genome size. Maximum fitting was obtained (R=1) and strain DNA content was solved from the resulting standard curve equation. All these data are shown in Figure 2.13. For the *T. microellipsoides* CLIB 830T, CBS 6641 and CBS 6762 strains, the average fluorescence intensity values ranged from 8.98 to 10.90 Mb, while DNA amounts in strains NRRL Y-17058 and CBS 6143 doubled these values with 21.71 and 22.79 Mb respectively. According to the DNA indices explained in Rodrigues et al. 2003, it was calculated two DNA indices corresponding to the relative DNA content of a problem strain compared to a reference. When S288c was used as a reference, it was observed less DNA content for strains CLIB 830T, CBS 6641 and CBS 6762 than for the reference (from 0.72-fold to 0.85-fold), while strains NRRL Y-17058 and CBS 6143 presented an increased amount of DNA (about 1.6-fold). Compared to the CLIB 830T strain, the same content was observed for CBS 6641 and CBS 6762, with the latter strain below them, but doubled for strains NRRL Y-17058 and CBS 6143.



Strains	MFI G0+ G1 ± SD	Mb ± SD	MFI G2 + M ± SD	Mb
S. cerevisiae S288c	52.86 ± 0.57	12.15[a]	105.41 ± 1.95	26.3
S. cerevisiae FY1679	90.41 ± 0.71	23.3[b]	174.20 ± 0.38	46.8
T. microellipsoides CLIB 830	38.09 ± 1.01	8.98 ± 0.24	81.97 ± 1.93	
T. microellipsoides NRRL Y-17	83.96 ± 0.92	21.71 ± 0.24	162.37 ± 2.08	
T.microellipsoides CBS 6143	87.84 ± 0.65	22.79 ± 0.17	166.64 ± 1.01	
T. microellipsoides CBS 6641	39.17 ± 0.60	9.28 ± 0.14	71.35 ± 1.49	
T. microellipsoides CBS 6762	45.03 ± 0.17	10.90 ± 0.04	84.67 ± 0.42	

Strains	DNA ln1	DNA ln2
S. cerevisiae S288c	1	1.4
S. cerevisiae FY1679	1.7	2.4
T. microellipsoides CLIB 830T	0.7	1
T. microellipsoides NRRL Y-17	1.6	2.2
T. microellipsoides CBS 6143	1.7	2.3
T. microellipsoides CBS 6641	0.7	1
T. microellipsoides CBS 6762	0.9	1.2

Figure 2.13. The median fluorescence Intensity correlation with genome size. Upper panel: the correlation between the mean of the median fluorescence intensity (MFI) expressed in arbitrary units (grey numbers) and genome size in Megabases (red numbers) is shown for the values corresponding to the haploid, diploid and tetraploid contents of reference strains S288c and FY1679. Bottom panel: [a, b] the median fluorescence values (MFI) of the *perfect* haploid and diploid strains, *S. cerevisiae* S288c and FY167. The means of the median fluorescence values of three independent replicas are indicated for both the G0/G1 and G2/M cell cycle stages. The values in Mb for each strain with a unknown genome DNA content were obtained using the equation in the upper panel. Standard deviations (SD) are provided for the newly obtained data.

2.3. Methodology:

2.3.1 The yeast strains used in this chapter

The *T. microellipsoides* strains for genome sequencing, flow cytometry, sporulation and the additional analysis were requested from the CBS (*Centraalbureau voor Schimmeltcultures*), the NRRL (*Agricultural Research Service Culture Collection*) or the CLIB (*Collection de Levures d'Intérêt Biotechnologique*) yeast strain collections. Strain name, origin and source of isolation are described in Strains Table 2. The *Torulaspora* strains employed for the phylogenetic analysis of the rDNA genes were taken from *genbank* (https://www.ncbi.nlm.nih.gov/genbank/?). Accession number, origin and source of isolation are indicated in Strains Tables 4 and 5.

The *T.microellipsoides* strain cultures were routinely grown in YPD-rich media (composition as described in Chapter 1, Methodology). The optimal growing temperature for each strain was not exactly determined, but we observed different growth rates at 30^0C. We detected that growing increased at specific temperatures for each strain. So, the number of cells was higher for CLIB 830T at 22^0C, for NRRLY-17058, CBS 6143 and CBS 6762 at 25^0C, and for CBS 6641 at 32^0C. The number of cells/mL grown until the exponential phase (O.D.=0.5) at these tested temperatures were as follows for each strain: **CLIB830T**: 3.5x 10^6 cells/mL , **NRRL-Y-17058**: 4x 10^6 cells/mL ; **CBS 6143**; 4.3x 10^6 cells/mL; **CBS 6641**: 6x 10^6 cells/mL and **CBS 6762**: 3.7x 10^6 cells/mL.

To observe ascii and ascospora formation, we grew *T. microellipsoides* cells in six different restrictive sporulation media: **YM** (yeast extract/malt extract): 1% glucose, 0.3% yeast extract, 0.3% malt extract , 0.5% peptone and 2% agar, **Gorodkowa**: 0.1% glucose, 0.5% sodium Chloride, 1% peptone and 2% agar, **Fowell** (acetate-agar I): 0.5% sodium acetate (tri-hidrate) and 2% agar,pH at 6.5-7, **McClary** (acetate-agar II) 0.1% glucose, 0.18% potassium chloride, 0.82% sodium acetate, 0.25% yeast extract and 2% agar, **Corn Meal Agar**: 4.2% corn flour and 2% agar, **Malt extract-agar**: 5% malt extract and 2% agar and **acetate-agar IV**: 1% potassium acetate, 0.125% yeast extract, 0.1% glucose and 2% agar. Ascospores and conjugation were observed after 2 weeks on the Corn Meal Agar media and the YM media, which was the preferred to sporulate cells.

2.3.2 Genome sequencing of the *T. microellipsoides* species

The assembly of the *Torulaspora microellipsoides* type strain CLIB830T genome was already available in Marsit S. *et al*., 2013. The genomic DNA from the *T. microellipsoides* NRRLY-17058 strain was prepared with these *Illumina NextSeq* sequencing libreries: *Mate Pair Library Prep Kit v2* and *True Seq DNA PCR-free sample preparation kit*. The sequencing kit that we employed was the *NextSeq 500/550 High Output Kit v2* (500 cycles). The remaining *T. microellipsoides* strains, CBS 6143, CBS 6641 and CBS 6762, were prepared from the same commercial libraries and sequenced by the *MiSeq Illumina* approach with the *MiSeq Reagent kit v2* (500 cycles) and the *MiSeq Reagent kit v3* (600 cycles).

De novo assemblies using the *Illumina* reads were built with different tools. To determine the proper *k-mer* parameter, the Velvet *v*1.1.06 software (Zerbino and Birney, 2008) was employed. A 95-nucleotide *k-mer* was the best obtained for the reads quality. Next, the definitive assembly and scaffolding were obtained with Sopra *v*1.4.6 (Dayarian, Michael and Sengupta 2010). *SSPACE v2.0* was used (Boetzer et al., 2011) for scaffold enlarging. Finally, the gaps between scaffolds were filled by *GapFiller* v1.11 (Boetzer and Pirovano, 2012).

The assembly output file describing the total number of scaffolds, N50 and the estimated genome size parameter is shown in Supplemental Table 1. The annotation of genomes was carried out using the *Yeast Genome Annotation Pipeline*, *YGAP*, on the website http://wolfe.ucd.ie/annotation/ (Proux-Wéra *et al*., 2012). To annotate the hybrid genomes, the assemblies were processed as from the post-WGD strains to allow the programme to assign two sequences for an ancestral gene instead of one sequence (two *pillars*). The number of sequences annotated in *YGOB* (http://ygob.ucd.ie/), with a reference in the ancestral genome, were: 9167 for CBS 6143, 9046 for NRRL Y-17058, 4613 for CBS 6641 and 4620 for CBS 6762.

The subtelomeric regions of interest were manually annotated because the genome references were unavailable in *YGOB* or in any other resource. We employed the *Artemis annotation tool* (Carver T. *et al*., 2012) to find all the unannotated ORFs, and then trimmed these frames to Methionine as the start amino acid and to finally perform blast searches with the *blastx* searching tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastx&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome). The coverage profiles were obtained from the mapping of the *T. microellipsoides* reads by employing the *breseq* pipeline (Deatherage and Barrick 2014).

2.3.3 Representation of the hybrids' genome structure

As a set of the reference *T. microellipsoides* chromosomes was not available, a *pseudo-chromosome* reconstruction for the inferred hybrid species was obtained by taking the annotation of the predicted parental strains which was subdivided into eight ancestral chromosomes (CBS 6641 and CBS 6762). The order of the codifying sequences in each *pseudo-chromosome* followed that of the ancestral genome in YGOB (http://ygob.ucd.ie/). Afterwards these sequences were concatenated to simulate eight chromosomes whose intergenic regions were omitted. The eight *pseudo-chromosomes* of each predicted parental strain were concatenated and the reads from the potential hybrids (NRRLY-17058 and CBS 6143) were mapped to those references using *Bowtie2* (Langmead and Salzberg, 2012) with the parameter defined in *-local.*

2.3.4 The *T. microellipsoides* and *Torulaspora* sp. phylogenies

Four *Torulaspora* phylogenies are presented in this Chapter. For the *T. microellipsoides* species, 1,844 sequences obtained from annotations were employed in a multi gene alignment. Hybrid subgenomes A and B, were classified according to the nucleotide distance obtained from each probable parental strain (CBS 6641 and CBS 6762). Then the phylogenetic analysis was performed.

Reconstruction with the type strains of the *Torulaspora* species was performed with the 26S rDNA sequences retrieved from *genbank.* The corresponding accession numbers are listed in Strains Table 4. An increased number of *Torulaspora* sp. sequences appeared in the databases, so we took advantage of this and, we performed two additional reconstructions with the 26S and 18S/ITS/5.8S rDNA sequences retrieved from *genbank* and from the CBS strains catalogue. Each phylogeny was obtained by the neighbour-joining method of reconstruction to compare our results with that obtained in the publication of Saluja et al. 2012. Analyses were conducted with *Mega v*6 (Tamura et al. 2013).

2.3.5 Karyotype analysis and Southern blot hybridisation

For the karyotyping analysis, yeasts were grown for 24 h to reach the stationary phase in 50 mL flasks with 10 mL of YPD media at different temperatures and 190 rpm. The collected cells were washed with 50 Mm EDTA, pH 8, and suspended in CPES buffer (40 mM citric acid, 120 mM Na_2HPO_4, 20 mM EDTA, 1.2 M de Sorbitol and 5 mM DTT) with 200µg/mL lyticase. Then 1% agarose dissolved in CPE buffer (the same as the CPES without Sorbitol and DTT) was added to perform blocks that contained genomic DNA with *SeaPlaque Agarose* (Lonza). These blocks were constructed for *S. cerevisiae* S288c, *S. cerevisiae* EC1118 and all the *T. microellipsoides* strains. The agarose blocks were first incubated with CPE buffer for 2 h and then with 1mg/mL of Proteinase K in the Solution 3 buffer (10 mM Tris, 0.45M EDTA and 1% lauryl sulphate, final pH=8) between 2-4 h. A 1% gel for the electrophoresis was prepared using *SeaKem Gold Agarose* (Lonza) and the electrophoresis buffer was 0.5% TBE. In the CHEF apparatus, gels were run under these conditions: 60 seconds for 12 h and 120 seconds for 14 h with an angle of 120^0 and a speed of 6V/cm. Chromosomic DNA was stained with EtBr for 1 h. The banding patterns of the karyotype images were analyzed by *Bionumerics v*7.6 (http://www.applied-maths.com/news/bionumerics-version-76-released). The detected band sizes of all the *T. microellipsoides* isolates were obtained by taking *S. cerevisiae* S2888c and EC1118 as the pattern references.

The *Southern* blot hybridisation was performed on the former karyotype gels. Probes consisted of labelled oligonucleotides that were synthetised using the *PCR DIG Probe Synthesis Kit* (Sigma-Aldrich). The oligonucleotides sequences were as follows: ***FSY1A*F**: 5'-GTAGCTGAACAAGAAAAGGAG-3' and ***FSY1B*F:** 5'-GCCGCTGAGCAAGATCAACAG-3'. A common reverse primer sequence was employed for PCR amplification: ***FSY1allR***: 5'-GCRATACCACCGTARAAGCC-3'. The programme's probe labelling conditions were as follows: initial denaturing at 95^0C for 2 minutes; 30 cycles comprising these steps: 1) 95^0C for 30s; 2) 60^0C for 30s; 3) 72^0C for 40s, and a final extension at 72^0C for 7 minutes. Nylon membranes were treated under high-medium stringency conditions to ensure probe specificity. The hybridisation buffer composition was: 50% Formamide, 5X SSC, 2% commercial blocking solution, 0.1% N-laurylsarcosine and 0.02% SDS. One low-astringency wash in 2X SSC /0.1% SDS at RT and a second high-astringency wash in 0.5X SSC/0.1% SDS at 68^0C were carried out. For the supplementary washes, we used the *DIG Wash and Block Buffer Set* (Roche, Mannheim, Germany). A signal was revealed using *Anti-Digoxigenin-AP, Fab fragments* (Sigma-Aldrich) and CDP-Star (Sigma-Aldrich).

2.3.6 Flow cytometry analysis and genome size content estimation

We complemented the karyotyping analysis with flow cytometry assays to estimate the DNA content of each *T. microellipsoides* strain. To this end, yeast cells were inoculated in 15 mL of YPD media and grown O/N until the stationary phase. The number of cells collected per assay was $1x10^7$ cells/mL. These cells were washed and fixed O/N at 4^0C with 70% EtOH. After fixing, they were treated O/N with a 2mg/mL RNase A solution and the next day they were incubated in a 4.5µL and 5mg/mL HCl-Pepsin solution. For fluorescence detection purposes, 50 µL of cells were mixed with 1mL of SYTOX Green solution (Thermo-Fisher).

Once we obtained the fluorescence values, we referred to the 2003 publication of Rodrigues et al., where the genome content estimation of *Z. bailii* strains was deduced by taking two *S. cerevisiae* strains as the DNA reference. According to these authors, a direct correlation between fluorescence and DNA content can be established. Two DNA indices were calculated from the median fluorescence intensity (MFI): the first index normalized the fluorescence values against the *S. cerevisiae* haploid MFI. The second index employed a similar ratio, but against the MFI of one tested genome. A correlation was found from our fluorescence data and we calculated both DNA indices for the *T. microellipsoides* strains.

2.4. Main conclusions

- The genome sequencing of the *T. microellipsoides* strains revealed that all the available isolates presented a conserved Region C among their genomes, except strain CBS 6762, which harboured only *ARB1*, a 95-kb gene region, and the *SEO1* and *FOT1* genes.
- The genomic fragments harbouring the *FSY1B1*/*FSY1B2* genes were found by genome sequencing in all the *T. microellipsoides* strains, which were located in a subtelomeric region. We found a conserved synteny of these genomic regions with that found in *Z. rouxii* and *T. delbrueckii*
- The former conclusions suggest that the *FSY1B1*/*FSY1B2* genes were ancestral genes that gave rise to the *FSY1A* paralogous gene, included in Region C.
- The phylogenies obtained for genes *FSY1* and *ATO3*, showed that strains NRRL Y-17058 and CBS 6143 presented an increased number of copies of both genes. From these reconstructions, we hypothesised that these two species could in fact be hybrid species.
- In these last gene reconstructions, we found out the probable parental species of these two inferred hybrid strains. One *T. microellipsoides* strain sequence, fromCBS 6641, was similar to two homologous sequences from the two hybrid strains. Another *T. microellipsoides* strain, CBS 6762, was close to the other sequences found on the hybrid genomes.
- A 1,844 multi gene phylogeny was reconstructed by taking the genome annotations of the *T. microellipsoides* genome sequences. Hybrid genomes were annotated based on their deduced parental strains and the resulting

- phylogeny supported our conclusions for the hybrid origin of strains NRRL Y-17058 and CBS 6143 seen in phylogenies *FSY1* and *ATO3*.
- From the *T. microellipsoides* phylogeny, we obtained a new result in addition to the confirmation of two hybrid species. One of the inferred parental strains, CBS 6762, presented a distant phylogenetic position from the *T. microellipsoides* strain, and only was closely related to one subgenome of the hybrid strains. So, we have probably found a new putative non-*Saccharomyces* species related to *T. microellipsoides*. Therefore, an interspecies hybribridisation event could be proposed for hybrid formation.
- A complementary measure of the nucleotide divergence between the strains classified as T*. microellipsoides* showed a long distance for strain CBS 6762 compared with the other T*. microellipsoides,* CLIB 830T and CBS 6641. However, a shorter distance was observed compared to one subgenome of both hybrid strains, which confirmed its parental origin.
- The parental origin of the hybrid strains could not be confirmed by the spore-to-spore conjugation assays run between inferred parental species because we failed to isolate the arising spores. However, we observed ascospores and a conjugation between the cells from the same strain. Thus the conjugation between species would probably be possible to lead to interspecific hybridisation.
- A posterior phylogeny with the *Torulaspora* genus type strains showed a similar position from strain CBS 6762. Hence, we concluded that it could be a new closely related species to *T. microellipsoides*.
- From the isolation of the new *Torulaspora* isolates not classified in any species, we performed two phylogenies for the 26S rDNA and the 18S/ITS/5.8S rDNA regions. The resulting reconstructions gave similar results

to the former phylogenies, so CBS 6762 was no more related to other *Torulaspora* strains.

- A pseudo-chromosomes representation was created for the predicted hybrid species by mapping their sequencing reads against the annotated scaffolds of the inferred parental strain. Reads were frequently shared out between the two parental sequences in both hybrid species for half the chromosome set, which suggests a common origin of both NRRL Y-17058 and CBS 6143. However, some differential duplications and deletions were detected in each representation, which suggests that these species diverged after the hybridisation event.
- The karyotiping analysis of the *T. microellipsoides* strains was firstly performed by Southern blot hybridisation, and revealed a different chromosome location of the *FSY1* copies and other duplications not detected by genome sequencing. This analysis also showed a wide variability in chromosome composition between the *T. microellipsoides* strains, with differences in number and sizes. By taking the chromosome banding of the *S. cerevisiae* references, we measured the bands presented in each strain. These data complemented other assays to obtain the genome size for each one. Divergency between hybrid genomes was reflected in the different karyotype patterns shown for them and also complemented the *pseudo-chromosomes* studies.
- Flow cytomnetry assays helped us to determine the fold change in genome size between the *T. microellipsoides* strains compared to the *S. cerevisiae* genome references, which confirmed the 2-fold content for the hybrid species. The profiles for strains CLIB 830T and CBS6641 showed undefined picks, which we attributed to deficient growth under our culture conditions.

General discussion section

In this thesis work we paid attention to elucidate the evolutionary mechanism concerning the genes that could perform important functions in wine fermentations carried out by yeast overall hexose transport. While we undertook our research, novelty and unexpected findings appeared, such as the discovery of introgression, not previously seen, thanks to deep genome sequencing with approaches like *PacBio*, or the description of hybrid species from a non-*Saccharomyces* genus, once again from genome sequencing by the *Ilumina* system in our own group. Finally, these results are becoming more important in our research and are the main goals of this thesis.

Here we describe, together with the obtained results, the connection made with others published by researchers, our contributions and our limitations.

1. Inferred yeast species donors of novel gene regions to *S. cerevisiae* strains

Our first result about this first point of the discussion came from collaborating with the group directed by Dr. Sylvie Dequin. The genome sequencing of the ***T. microellipsoides*** **CLIB 830T** strain confirmed what we already detected in our group by first the *FSY1* gene reconstruction: **a novel Region C** (described in Novo *et al*. 2009) **was found in that species, arranged with the same distribution, at an ending chromosome position and, finally, with as much as 99.5% similarity** (Marsit et al. 2015). Some particularities in the CLIB 830T Region C were found: we observed that the first gene, *ARB1*, presented a bigger nucleotide sequence than that found in the *S. cerevisiae* strains and this gene was located at 80kb from the second gene of Region C, *PUT3.*

Some years earlier, the EC1118 novel Region B, the smallest found in this strain (17kb), which was located at a central position of the chromosome and a 3-fold repeat inside the genome, was attributed to a different pre-WGD species, *Zygosaccharomyces bailli* (Galeote et al. 2011). In this case, the percentage of similarity between homologous genes was lower than for Region C, at about 95%. For the last region found in the EC1118 strain, Region A, no species has been assigned. Thus it will probably arise quickly with future genome sequencing projects.

While the model followed to incorporate Region B was done by forming a circular molecule (Galeote et al. 2011), the formation of a first hybrid species to acquire Region C was proposed, which involved carrying out subtelomeric recombinations (Novo, Bigey, Beyne, Galeote, Gavory, Mallet, Cambon, J.-L. Legras, et al. 2009). This last hypothesis agrees with newer theories about an ancestral hybridisation by allopolyploidisation, which could give rise to all the species from **WGD.** Thus, some species from the ZT clade (*Zygosachharomyces-Torulaspora*) could be one of the subgenomes of this ancestral hybrid (Marcet-Houben and Gabaldón 2015a).

Moreover, the fact that yeast species cohabit in natural must environments with *S. cerevisiae* is well know, and are commonly called contaminant species because their excess could cause wine spoilage. Fleet and Heard 1993 previosly mentioned which species grow in the different fermentation stages: *Candida*, *Debaryomyces*, *Hanseniaspora*, *Hansenula*, *Kloeckera*, *Metschnikowia*, *Pichia*, *Schizosaccharomyces*, *Torulaspora* and *Zygosaccharomyces*. It is common knownledge that non-*Saccharomyces* species grow on the first days of fermentation to then be displaced by *S. cerevisiae*. However, some *Z. bailii* and *Z. fermentati* strains are high ethanol concentration-tolerant, so they dominate on the last days of fermentation (Fleet 2003). The *FSY1* expression of EC1118 *S. cerevisiae* 3-fold increases in 2% ethanol compared to 0.5% fructose, as seen in Galeote et al. 2010. Therefore, the species that harbour this gene, such as *T.microellipsoides*, could grow in the last stages of fermentation, where the amounts of fructose and glucose are small and ethanol concentrations are high. Regarding these pre-WGD species, two isolates of strains *T. microellipsoides* and *Z. rouxii* have been found after 8 fermentation days of White Riesling grapes from Arkansas (USA), which suggests that these strains could also exist in final fermentation stages (Moore et al. 1988). In the thesis work by Esteve-Zarzoso, *T.*

microellipsoides strains were isolated from another white grape variety of natural must, Palomino must. This species was isolated from the fermentations of some grapes harvested later than others. Therefore perhaps these grapes had a large amount of sugar when fermentation took place.

This makes sense given the source of isolation known for the *T. microellipsoides* strains, as they are isolated from sugar-enriched environments: apple juice, lemonade, tea-beer, berries and a sandalwood tree exudate. With this information, we wondered if non-*Saccharomyces* species could play a more or less important role in fermentation. From the biotechnological application point of view, they have been found to contribute to the final aromatic composition. A recopilation of enzymatic reactions produced by non-*Saccharomyces* wine yeast is reviewed in Esteve-Zarzoso et al. 1998. In addition, mixed cultures with *Saccharomyces*/non-*Saccharomyces* have resulted in an unexpected final chemical composition of wines, which implies the possibility of a new wine flavour composition (Ciani et al. 2010).

Over the last few years, the employment of non-*Saccharomyces* species in mixed and sequential fermentations (first innoculating the non-*Saccharomyces* species) has emerged and improves wine quality. This seems relevant for the production of new volatile compounds and others of secondary metabolism products, and also for lowering alcohol concentration in wines. To this end, *T. delbrueckii* is the mostly widely selected species as it minimises undesired effects in final productions, such as acidity. Sequential fermentation using this species has been analysed in lychee wines, where the conclusion was that it clearly contributed with more alcohols and esters, and also with the retention of terpenoids (Chen and Liu 2016). For red wine, the use of *T. delbrueckii* and also *S.pombe* sequential and mixed cultures with *S. cerevisiae* has resulted in great aromatic complexity, and

also in wine colour stability (Loira et al. 2015). The specific transcriptional responses of *T. delbrueckii* and *S. cerevisiae* in early co-cultivation stages consisted in recognising one another to start outcompeting (Tronchoni et al. 2017). Moreover, *Dekkera* and *Brettanomyces* species have already been employed in the fermentations of sour beers and strains of *Wickerhamomyces anomalus* and *Torulaspora delbrueckii*, and are now being used for brewing fermentations given their fruity esters production and their contribution to lower alcohol concentrations (Basso et al. 2016).

Hybridisation between cohabiting species in fementative environments is largely known. Brewing species *S. pastorianus* and wine strain *S. cerevisiae* S6U are well-known examples in the yeast scientific community. A very interesting work about hybridisation between yeasts is that by Marinoni et al. 1999, who generate hybrids of *Saccharomyces* and non-*Saccharomyces* (ancient *sensu lato*) and observed the karyotype composition of the resulting hybrids. Phylogenetically close parental species produced a hybrid that conserved both parental genomes, while the more distant ones conserved only one of the subgenomes, but presented some traces of the other parental genome. These experiments demostrate that despite being rare, hybrids of distant yeast species might happen, and some genetic material may be transferred to each other.

The introgression of gene regions, or horizontal gene transfer, is a well-reported event in the bibliography. Here we report the Region C introgression from two phylogenetically distant yeasts, but numerous examples of horizontal gene transfer to *S. cerevisiae* and to other yeasts can be found in the bibliography. The bacteria or yeasts used as the donor, and the acquisition and maintenance of these genes, are because they contribute with an advantage for cells. Tree topologies have revealed that the origin of the *URA1* gene of the Saccharomycetaceae family

(some revised species) lies in *Lactobacillales* and that the *BDS1* gene (which codifies for sulphatase) groups with α- proteobacteria. Moreover, from the pathway for biotin production identified in *S. cerevisiae* and composed of six genes, two have a bacterial origin (*BIO3* and *BIO4*). It is also known that *S. cerevisiae* clinical strains present patogenic genes from bacteria; e.g. during the genome sequencing of clinical strain *S. cerevisiae* YJM789 (Wei et al. 2007), the genes from pathogenic bacteria, such as *E. faecalis* and *S. enterica,* are found. Not only in *S. cerevisiae*, but also in the genome sequencing of other *Saccharomycetaceae* yeast, like *Z. rouxii*, *S. kluyveri*, *K. waltii*, *K. thermotolerans*, *K. lactis* and *E. gossypii*, the genes of HGT have shown a bacterial origin (Rolland et al. 2009).

2. From *T. microellipsoides* to *Saccharomyces* genus gene introgressions.

Although many *S. cerevisiae* yeast strains were checked for the presence of genes from the three novel regions, **no other *Saccharomyces* species were take for this search**. The prevalence of these genes among *S. cerevisiae* wine strains was quite clear, and can be explained by their possible advantage to grow under fermentation conditions. This hypothesis has already been tested in Marsit et al. 2015, who reported improved of the oligopeptide transport for the EC1118 strain as it harbours the *FOT* genes located in Region C.

Complete Regions B and C are found only in *S. cerevisiae* strains (not only wine strains) and in non-*Saccharomyces* species *Z. bailli* and *T. microellipsoides*, respectively. When we looked at the percentage of similarity in these blocks of genes between donors and acceptors, we can speculate that both transfers were recent events, perhaps in a common ancestor of one wine lineage and other *S. cerevisiae* lineages. Another possibility is that the acceptor species is the common ancestor of all the *S. cerevisiae*, and then some lineages lose all or part of the regions because of their environment specialisation.

Since genome sequencing of strain CLIB 830T, Region C was detected in *T. microellispoides* with a few minor modifications. An 80kb insertion of genes was between belonging to Region C genes *ARB1* and *PUT3*. When performing the genome searching of Region C, in two of the new *S. cerevisiae* sequences, this region was more similar to that of the T. *microellipsoides* region than to that of the two *S. cerevisiae* strains. These findings made us wonder if the direction of the region transfer went from *T. microellipsoides* to *S. cerevisiae* or the other way round. Then we decided to perform the genome sequencing of all the existing *T. microellipsoides* strains to search for Region C. We found that all the strains from different origins except CBS 6762, presented the region, which suggested that the direction of gene transfer went from the non-*Saccharomyces* species to *S. cerevisiae*.

As mentioned before, one of our first analyses was performed with the *FSY1* gene presented in Region C. This gene codifies for a fructose/H^+ transport system, which is active at low fructose concentrations, and thus shows high-affinity for this sugar. It was interesting for us to know more about the existence of this gene in yeast species given its importance in fructose consumption during fermentation. We took all the available *Saccharomycetaceae* strains from our lab collection to check the presence of this gene.

Following this approach, we were able to know which species was the donor of the *FSY1* gene (or complete Region C). Unexpectedly, **we discovered an additional diverging sequence in the CLIB 830T *T. microellipsoides* strain** that clustered together with those of *S. eubayanus*, *S. pastorianus* and *S. uvarum*. We noticed that in nucleotide composition, this sequence (which we called ***FSY1B***) differed from that of *S. cerevisiae* (we called it ***FSY1A***) by 25.5% on average (74.5% of homology). This large difference suggested that we had found a paralogous copy

of the gene rather than a gene allele. By the *Southern* blot hybridisation approach, we checked that each one was in a different chromosome (although a second band of ***FSY1A*** appeared in the same chromosome of ***FSY1B***), which confirmed the origin of the **paralogous** duplication. The *FSY1B* sequence similarity between these brewing species and *T. microellipsoides* CLIB 830T was less than the *FSY1A* gene compared to its homologous sequence in *S. cerevisiae*, which was about 85% on average. We concluded from the obtained data that the differences in the homology percentage of both sequences with these *T. microellipsoides* was because **both genes were acquired in the different events of the gene transfer from this species to these *Saccharomyces* species**. Copy B of this gene was present in every tested *S. eubayanus* strain and derived hybrids, which suggests that the transfer could happen in the ancestor of the species. One probable advantage in the growth of this species could be assignated to the fructose trasporter, maybe in brewing environments. The only presence of copy A of this gene in some *S. cerevisiae* strains, could be also attributed to some advantage in wine fermentation or to any other fermentative process because the other two *S. cerevisiae* strains presenting this gene came from food and beer sources.

From the genome sequencing of the *T. microellipsoides* isolates, we obtained a duplicated copy of ***FSY1B*** in some strains. The synteny reconstruction of the regions harbouring this pair *FSY1B1*/*FSY1B2* showed a conserved order in relation to the synteny in *T. delbrueckii* and *Z.rouxii*, which suggests that these were the *FSY1* ancestral copies, while *FSY1* from Region C (*FSY1A*) probably appeared in a posterior gene duplication event and gave rise to this paralogous copy. We were also able to explain the absence of these transferred genes in the other post-WGD genera as a widespread loss of them in the WGD ancestor due to the colonisation of non-wine fermenting environments. This would be the most parsimonic explanation.

With the *ATO3* gene, we detected 90.8% homology on average between the *Saccharomyces* new *ATO3* group and those sequences of Region C (Supplemental Figure 4). The most remarkable result related to this new copy was that the majority of the *Saccharomyces* species presented it with more than one representative strain for each one and, once again, the most probable species responsible for the transfer would be *Torulaspora microellipsoides*. We previously concluded that that pre-WGD species was the donor of Region C to *S. cerevisiae* and, despite them sharing the position in the *ATO3* phylogeny, we thought that *T. microellipsoides* was probably the donor rather than *S. cerevisiae*. As a 90.8% homology was similar to that of *FSY1B*/*FSY1B2* (85.2%), we hypothesised that the former could also be addressed as an ancient introgression, which would at least imply the *Saccharomyces* ancestor. We cannot assert whether the acquisition of both genes happened during the same evolutionary event. Perhaps the *FSY1B*/*FSY1B2* copy was already present in the common ancestor of the *Saccharomyces* genus and was later lost in nearly all the species, except for the ancestor of *S. eubayanus*.

Therefore, *T. microellipsoides* contributed with a greater introgression of 65kb, and also with the transfer of both *FSY1* to *S. eubayanus* and of *ATO3* to the *Saccharomyces* genus, as well as some post-WGD genera. Then this *Torulaspora* genus species would appear to be a great contributor of genetic material.

Both fructose transporter and ammonia exporter genes were detected in more than one copy in a subtelomeric region of the strains under study. Subtelomers are zones that come close to chromosome ends rich-in-repetitions because they contain sequences of transposable elements (Y' elements, X elements and TG repetitions) that are actually well characterised (Yue et al. 2017). These repeats favour the recombination between chromosomes, a molecular mechanism

that probably led to not only the increase in the paralogous copies in each strain, but also to the introgression of all these genes after the hybrid intermediate had formed. As we conclude, the transfer of *FSY1B*/*FSY1B2* and *FSY1A* occurred during different evolutionary events due to their contrasting accumulations of nucleotide changes. So it was required for hybridisation events at least. If *ATO3* is not acquired simultaneously to *FSY1B*/*FSY1B2*, then we have to add another hybridisation event.

As we mentioned in the Introduction, the work of Marcet-Houben and Gabaldón 2015 speculates that the hybridisation between two different species would lead to the ancestor species of the WGD event. After a profound analysis of what they called *phylomes*, they concluded that the implicated species probably belonged to the ZT clade (*Zygosaccharomyces* and *Torulaspora*). The results presented herein for *FSY1B*/*FSY1B2* and *ATO3* could suggest that what we found out is a little trace of that ancient WGD, but no conclusions can be reached with this hypothesis because having only two genes to study is very risky.

3. Genome diversity of the *T. microellipsoides* strains and the discovery of a new non-*Saccharomyces* species

From the genome sequencing of existing *T. microellipsodies* isolates, we found two hybrid species (NRRL Y-17058 and CBS 6143) and we predicted the hybrid origins of these strains by inferring the two probable parental species. We deduced that strain CBS 6641 could be one of the parents and that a CBS 6762-like species would probably be the other parent. We proposed that this last strain could be a new putative non-*Saccharomyces* species related to *T. microellipsoides* by looking at its phylogenetic position and its nucleotide divergence in relation to the remaining *T. microellipsoides* strains. Then we hypothesised that the found hybrid species would be the product of interspecific hybridisation between different species.

In the last 10 years, numerous publications have emerged about the discovery of new non-*Saccharomyces* species and hybrid species from the already known pre-WGD genera. They were all found close to the *Zygosaccharomyces* genus and were specifically more related to species *Z. rouxii* and *Z. bailli*. James et al. 2005, described the existence of three natural hybrids inside the *Zygosaccharomyces* genus by a phylogenetic reconstruction of nuclear and mitochondrial genes (*ADE2*, *HIS3* and *SOD2*). These authors concluded that these hybrid species resulted from an interspecific hybridisation, which we observed for strains NRRL Y-17058 and CBS 6143. The nuclear DNA from those hybrids appeared to originate from *Z. rouxii* and a novel species related to *Z. mellis* and *Z. rouxii*. Furthermore, the above-cited identified the NCYC 3042 strain as this novel species, whose 26S rDNA sequence was strictly related to *Z. rouxii* and *Z. mellis*, and they referred to it as *Z. pseudorouxii*.

Afterwards, in the work of Solieri et al., 2007, from the two strains included in a phylogenetic analysis with other *Zygosaccharomyces* strains, the authors concluded that they had probably found two new putative species, strains ABT 301 and ABT 601, that were closely related to *Z. rouxii*.

Some years later, Suh et al. 2013 classified some strains before being included in the *Z. bailli* species in two different groups, the *Z. parabailli* group which included 10 strains and was the closely related group to *Z. bailli*; the *Z. pseudobailli* group, which first diverged from *Z. parabailli* and included five strains.

Therefore, although hybridisation studies have traditionally focused in the *Saccharomcyces* genus, in recent years more information has emerged about the existence of non-*Saccharomcyes* hybrid species of an interspecific origin. Here we report new findings on this kind of hybrid species that are isolated from natural environments and which we know come from a different species compared to previously reported hybrids: *Torulaspora microellipsoides*.

General conclusions

1) *T. microellipsoides* has been revealed as an important gene contributor to the *Sacccharomyces* genus, being the responsible of diverse gene introgressions.

2) Genome analysis of strains classified as *T. microellipsoides* has shown a hugh diversity inside this species. Hybrid and parental species have been inferred from phylogenetic analysis and a new species close to *T. microellipsoides* have been proposed.

3) All the *T. microellipsoides* strains (including CBS 6762) presented genes that were classified as genes coming from an introgression, first in EC1118, which reforced the idea of the direction of the transference, from *T. microellipsoides* to *Saccharomyces* species.

List of supplemental figures and strains tables

Supplemental Table 1 (ST1). **Statistic values from the *de novo* assembling with the Velvet and Sopra software.** The assembling quality could be measured by looking at the output files from the assemblers. The parameters that determine the degree of quality are: the total number of obtained scaffolds, the N50 value: this means that half the bases are covered by scaffolds that are longer than the N50 value and the estimated genome size. Tm= *T. microellipsoides*.

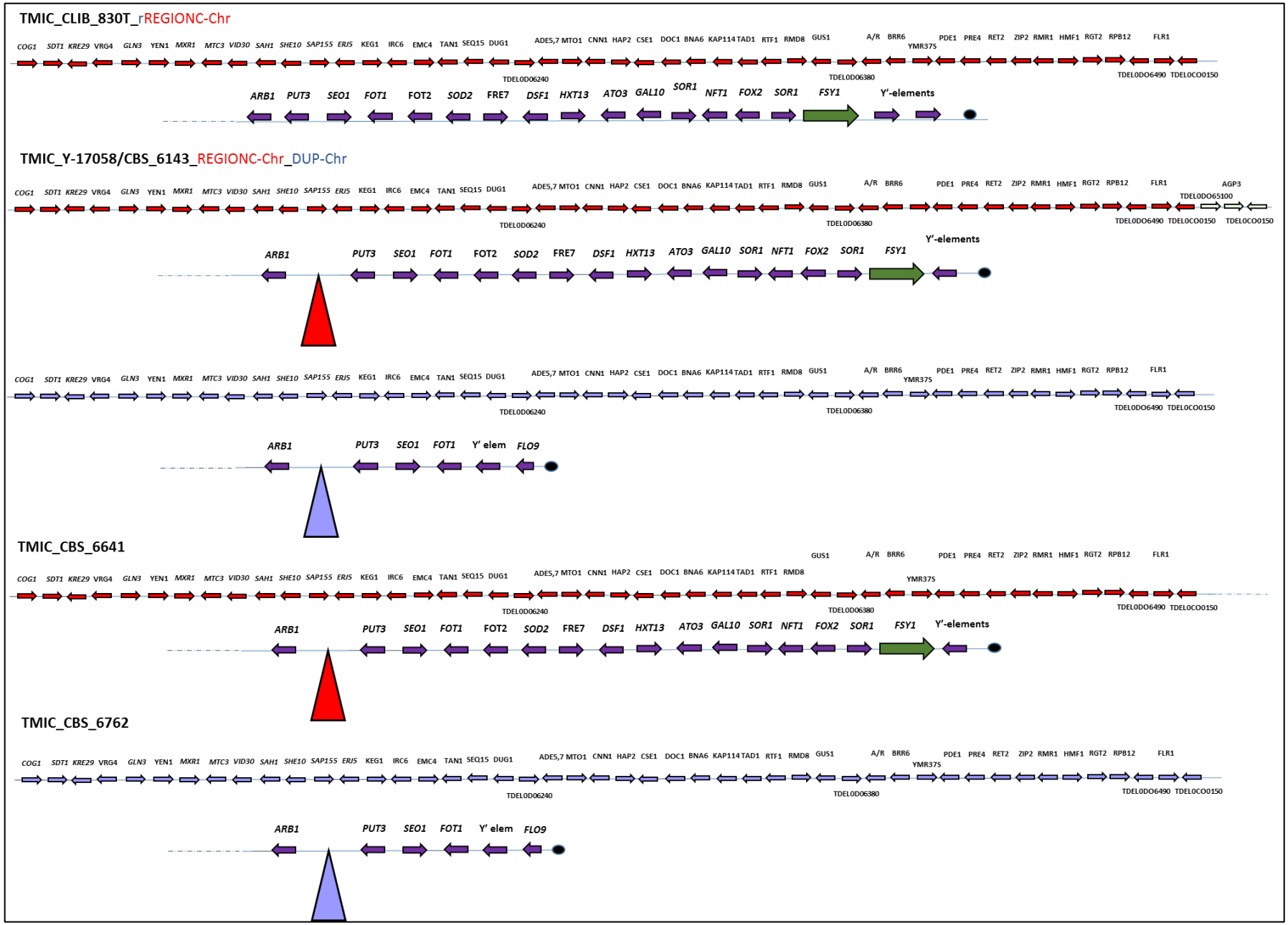
Tm strain	Total scaffolds	N50	Genome size
NRRLY-17058	408	619942	20.186.889 Mb
CBS 6143	615	88607	20.487.845 Mb
CBS 6641	159	258883	10.238.615 Mb
CBS 6762	215	231345	10.294.020 Mb

Supplemental Figure 1 (SF1). **The reconstruction of the chromosome harbouring the Region C genes in *T.microellipsoides* species**. The *T. microellipsoides* synteny of these genomic regions was obtained from the *Artemis* manual annotation. The first genes represented from left to right are *COG1*, *SDT1*, *KRE29* and *VRG4*. These genes were discovered in *S. cerevisiae* strains CBS 7957 and the CLIB 382, and here we show that all the *T. microellipsoides* strains also present them. An identical annotation was performed for strains CLIB 830T and CBS 6641 but the orthologous genomic region found in strain CBS 6762 presented a different subtelomeric gene composition. In this species, Region C was reduced to the *FOT1* gene and the following annotated ORFs were an Y' element and a *FLO9* homologous gene. With the *T. microellipsoides* hybrid strains, we detected two duplicated genomic regions whose gene content differed from a *TDEL0C00150* homologous gene to *ARB1* gene. One of the regions was similar to that found in the CBS 6762 strain, and the other one which harboured Region C was similar to that presented in CBS 6641, except for the insertion of three genes (*TDEL0D065100*, *AGP3* and a second copy of *TDEL0C00150*). Genes are represented by arrows that do not show their real size. The Region C genes are drawn in purple and the *FSY1* gene is highlighted in green. The red and blue triangles (colors despite their similar origin) denote the 95kb gene stretch found in the *T. microellipsoides* strains between *ARB1* and the *PUT3* gene.

Supplemental Figure 2 (SF2). The s**yntenic context of the *FSY1B1*/*FSY1B2* genes**. The gene composition of the genomic region where the ***FSY1B1*/*FSY1B2*** gene copies are located is shown for the *T. microellipsoides* strains, *T. delbrueckii*, *Z. rouxii* and *S. uvarum*. The core chomosomic and subtelomeric regions are drawn with a green and a yellow background, respectively as in Yue et al. 2017. The gene annotation retrieved from YGOB is presented for the *T. delbrueckii*, *Z. rouxii* and *S. uvarum* genes, while the *T. microellipsoides* genes were *de novo* annotated). Genes are represented by arrows that do not show their real size. An orange triangle in the *T. delbrueckii* subtelomeric region is drawn to indicate an rDNA region. Black circles denote the end of chromosomes which, for the *T. microellipsoides* strains, correspond to

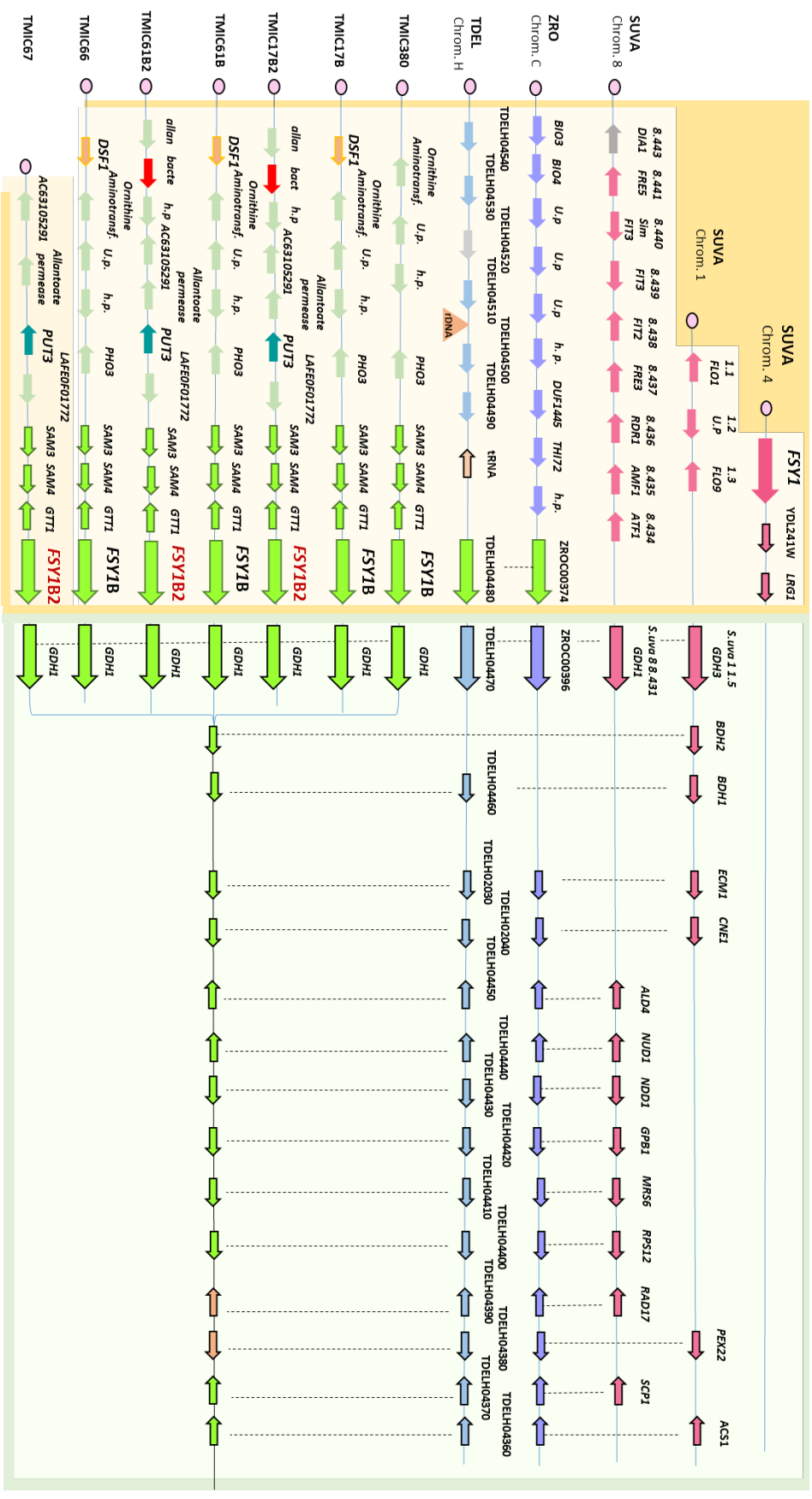
Supplemental Figure 3 (SF3). The n**ucleotide divergence matrix between the *ATO3* genes**. A distance matrix for the aligned 543 bp was obtained with the Mega *v*6.0 software (Tamura et al. 2013). The nucleotide change of the *ATO3* genes found in *T. microellipsoides* homologous Region C and the additional copies distributed along the genome (from 23 to 35 rows and columns) were compared to the new *ATO3* genes found in the *Saccharomyces* species (from 1 to 22 rows and columns). The most important data to be compared are highlighted with a black square.

Supplemental Figure 4 (SF4). **Band size quantification in the *T. microellipsoides* strains.** The karyotype image analysis containing the S288c and the EC1118 *S. cerevisiae* strains and all the *T. microellipsoides* isolates was carried out by *Bionumerics v*7.6 to estimate the genome size of these non-*Saccharomyces* strains. The resulting bands and their sizes calculated from the *S. cerevisiae* S288c genome reference are observed in the first part of the figure. The reference band sizes and the data obtained with the analysis are shown in the second part of the figure. Green boxes represent the bands detected by the programme and the sizes estimated, while red boxes correspond to those bands that the programme failed to recognize and that we after calculated taking the by us accepted band sizes.


TMIC_CLIB_830T_rREGIONC-Chr
COG1 SDT1 KRE29 VRG4 GLN3 YEN1 MXR1 MTC3 VID30 SAH1 SHE10 SAP155 ERJ5 KEG1 IRC6 EMC4 TAN1 SEQ15 DUG1
ADE5,7 MTO1 CNN1 HAP2 CSE1 DOC1 BNA6 KAP114 TAD1 RTF1 RMD8 GUS1
A/R BRR6 YMR37S PDE1 PRE4 RET2 ZIP2 RMR1 HMF1 RGT2 RPB12 FLR1
TDEL0D06240 TDEL0D06380 TDEL0DO6490 TDEL0CO0150
ARB1 PUT3 SEO1 FOT1 FOT2 SOD2 FRE7 DSF1 HXT13 ATO3 GAL10 SOR1 NFT1 FOX2 SOR1 FSY1 Y'-elements
TMIC_Y-17058/CBS_6143_REGIONC-Chr_DUP-Chr
AGP3 TDEL0DO65100 TDEL0CO0150
TMIC_CBS_6641
TMIC_CBS_6762
FLO9 Y' elem

SF1

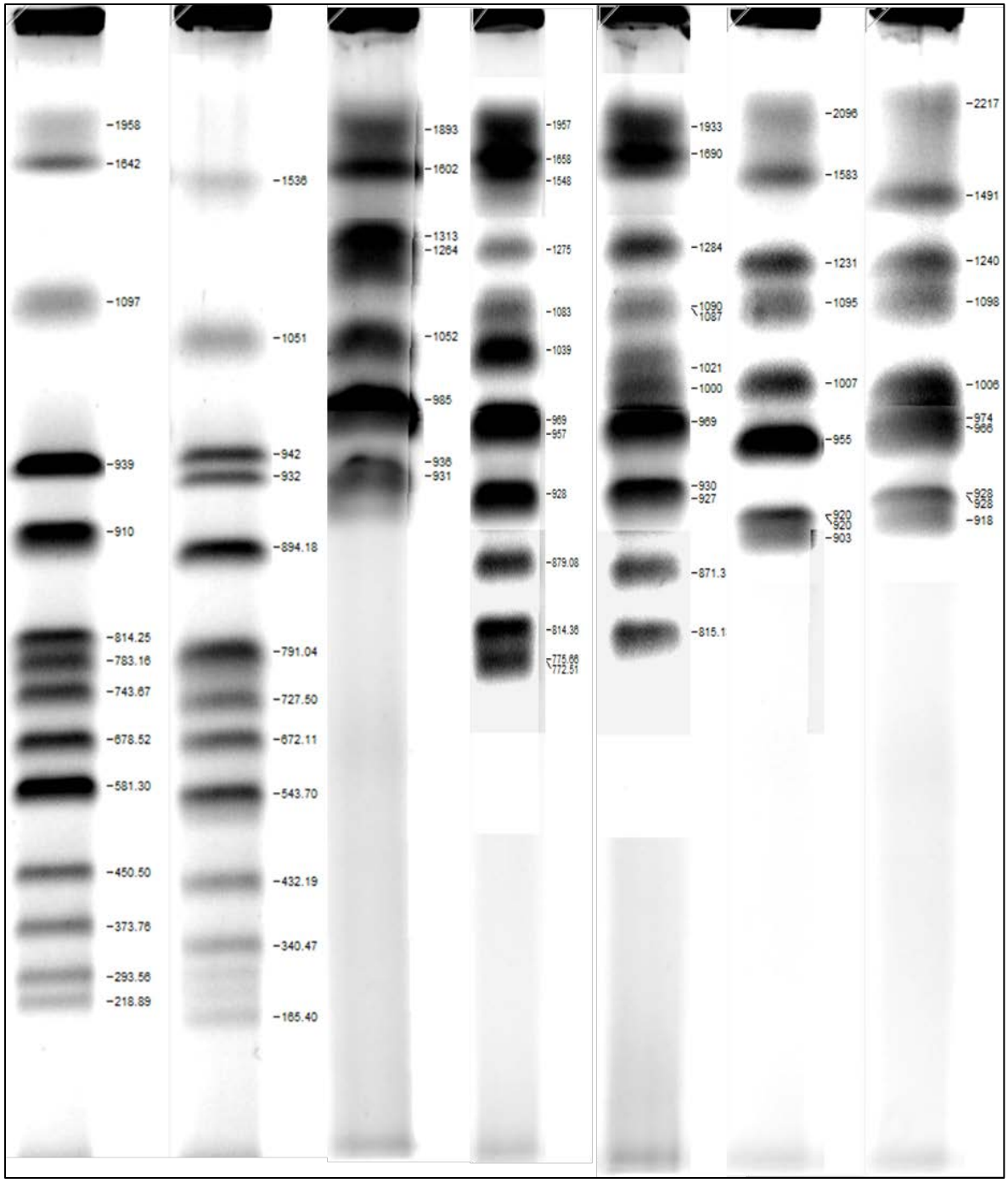
SF2


SUVA
Chrom. 4
FSY1
YDL241W
LRG1
SUVA
Chrom. 1
1.1
FLO1
1.2
U.p
1.3
FLO9
SUVA
Chrom. 8
8.443
DIA1
8.441
FRE5
8.440
Sim
FIT3
8.439
FIT3
8.438
FIT2
8.437
FRE3
8.436
RDR1
8.435
AMF1
8.434
ATF1
ZRO
Chrom. C
BIO3
BIO4
U.p
U.p
U.p
h. p.
DUF1445
THI72
h.p.
ZROC00374
TDEL
Chrom. H
TDELH04540
TDELH04530
TDELH04520
TDELH04510
TDELH04500
TDELH04490
tRNA
TDELH04480
rDNA
TMIC380
Ornithine
Aminotransf.
U.p.
h.p.
PHO3
SAM3
SAM4
GTT1
FSY1B
TMIC17B
DSF1
Ornithine
Aminotransf.
U.p.
h.p.
PHO3
SAM3
SAM4
GTT1
FSY1B
TMIC17B2
allan
bact
h.p
AC63105291
Ornithine
Aminotransf.
U.p.
h.p.
Allantoate
permease
LAFE0F01772
PUT3
SAM3
SAM4
GTT1
FSY1B2
TMIC61B
DSF1
Ornithine
Aminotransf.
U.p.
h.p.
PHO3
Allantoate
permease
LAFE0F01772
SAM3
SAM4
GTT1
FSY1B
TMIC61B2
allan
bacte
h.p
AC63105291
Ornithine
Aminotransf.
U.p.
h.p.
Allantoate
permease
LAFE0F01772
PUT3
SAM3
SAM4
GTT1
FSY1B2
TMIC66
DSF1
Ornithine
Aminotransf.
U.p.
h.p.
PHO3
SAM3
SAM4
GTT1
FSY1B
TMIC67
AC63105291
Allantoate
permease
LAFE0F01772
PUT3
SAM3
SAM4
GTT1
FSY1B2
S.uva 1 1.5
GDH3
S.uva 8 8.431
GDH1
ZROC00396
TDELH04470
TDELH04460
TDELH02030
TDELH02040
TDELH04450
TDELH04440
TDELH04430
TDELH04420
TDELH04410
TDELH04400
TDELH04390
TDELH04380
TDELH04370
TDELH04360
GDH1
BDH2
BDH1
ECM1
CNE1
ALD4
NUD1
NDD1
GPB1
MRS6
RPS12
RAD17
PEX22
SCP1
ACS1

SF3

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
1	*ATO3*_SpUWOPS91171-chrV																																			
2	*ATO3*_SpUWOPS91171-chrXVI	0																																		
3	*ATO3*_SpUWOPS91171-chrIV	0	0																																	
4	*ATO3*_SpUFRJ50816-chrXV	3	3	3																																
5	*ATO3*_SpUFRJ50816-chrXVI	3	3	3	0																															
6	*ATO3*_SpYPS138-chrXIII	4	4	4	1	1																														
7	*ATO3*_SpNRRLY-17217	7	7	7	6	6	7																													
8	*ATO3*_SpCBS432-chrXIII	7	7	7	6	6	7	0																												
9	*ATO3*_SpN44-chrXIII	9	9	9	8	8	9	4	4																											
10	*ATO3*_SceAWRI796	21	21	21	20	20	21	16	16	15																										
11	*ATO3*_SceRM11-1a	21	21	21	20	20	21	16	16	15	0																									
12	*ATO3*_SceK7-chrXIV	21	21	21	20	20	21	16	16	13	2	2																								
13	*ATO3*_SmikNBRC1815	31	31	31	29	29	27	27	27	28	28	28	28																							
14	*ATO3*_SkNBRC1802-ChrV	36	36	36	37	37	36	31	31	33	31	31	31	31																						
15	*ATO3-2*_SkNBRC1802-ChrV	36	36	36	37	37	36	31	31	33	31	31	31	31	0																					
16	*ATO3*_SkNBRC1802-ChrX	37	37	37	38	38	37	32	32	34	32	32	32	32	1	1																				
17	*ATO3*_SkNBRC1802-ChrI	36	36	36	37	37	36	31	31	33	31	31	31	31	0	0	1																			
18	*ATO3*_SuvaCECT12600	44	44	44	43	43	44	42	42	42	44	44	44	46	35	35	36	35																		
19	*ATO3*_SuvaCECT1969	44	44	44	43	43	44	42	42	42	44	44	44	46	35	35	36	35	0																	
20	*ATO3*_SuvaBMV58	44	44	44	43	44	42	42	42	44	44	44	44	46	35	35	36	35	0	0																
21	*ATO3*_SuvaNPCC1290	43	43	43	42	43	41	41	41	43	43	43	43	47	36	36	37	36	1	1	1															
22	*ATO3*_SeubFM1318	38	38	38	37	37	37	37	36	43	43	43	43	39	39	39	40	39	16	16	16	15														
23	*RC*_CBS6143	53	53	53	53	53	53	53	50	52	52	51	52	50	50	50	51	50	45	45	45	46	46	3												
24	*ATO3 (2)*_CBS6641	53	53	53	53	52	53	53	50	52	52	51	55	49	49	50	49	49	45	45	45	46	46	1	4											
25	*RC*_CBS6641	54	54	54	54	54	54	54	51	53	53	52	53	51	51	51	52	51	46	46	46	47	45	1	4	2										
26	*RC*_NRRL Y-17058	52	52	52	52	52	52	52	49	51	51	50	51	49	49	50	49	49	46	46	46	48	44	2	5	3	1									
27	*RC*_Sc7957	51	51	51	51	51	51	51	48	50	50	49	50	50	50	51	50	47	47	47	47	48	44	2	5	3	1	0								
28	*RC*_6765	51	51	51	51	51	51	51	48	50	50	49	50	50	50	51	50	47	47	47	48	44	2	5	3	1	0	0								
29	*RC*_EC1118	51	51	51	51	51	51	51	48	50	50	49	50	50	50	51	50	47	47	47	48	44	2	5	3	1	0	0								
30	*ATO3 (2)*_CLIB830T	50	50	50	50	50	49	48	48	47	51	51	50	46	46	46	47	46	48	48	48	49	46	9	6	10	8	7	7	7						
31	*ATO3 (2)*_NRRL Y-17058	50	50	50	50	50	49	48	48	47	51	51	50	46	46	46	47	46	48	48	48	49	46	9	6	10	8	7	7	7	0					
32	*ATO3 (2)*_CBS6143	50	50	50	50	49	48	48	47	51	51	50	50	46	46	46	47	46	48	48	48	49	46	9	6	10	8	7	7	7	0	0				
33	*RC*_CLIB830T	51	50	50	50	49	48	48	47	51	51	51	50	46	46	46	47	46	48	48	48	49	46	10	7	11	9	8	8	8	1	1	1			
34	*ATO3B*_NRRL Y-17058	79	79	79	78	78	79	79	79	78	79	79	78	80	78	78	79	78	72	72	72	73	71	57	57	57	56	56	56	56	55	55	55	56		
35	*ATO3B*_CBS6143	79	79	79	78	78	79	79	79	78	79	79	78	80	78	78	79	78	72	72	72	73	71	57	57	57	56	56	56	56	55	55	55	56	0	

SF4


1958
1642
1097
939
910
814.25
783.16
743.67
678.52
581.30
450.50
373.76
293.56
218.89
1536
1051
942
932
894.18
791.04
727.50
672.11
543.70
432.19
340.47
185.40
1893
1602
1313
1264
1052
985
936
931
1957
1658
1548
1275
1083
1039
969
957
928
879.08
814.36
775.66
772.51
1933
1690
1284
1090
1087
1021
1000
969
930
927
871.3
815.1
2096
1583
1231
1095
1007
955
920
920
903
2217
1491
1240
1098
1006
974
966
928
928
918

SF4-*cont.*

Chr num.	REFERENCE	S288c	S288c	EC1118	EC1118	CLIB 830T	CLIB 830T	NRRL Y-17	NRRL Y-17	CBS 6143	CBS 6143	CBS 6641	CBS 6641	CBS 6762	CBS 6762
1	1900	1884	1958	1518	1536	1746	1893	1738	1957	1800	1933	2010	2096	2018	2217
2	1640	1639	1642	1049	1051	1596	1602	1628	1658	1629	1690	1544	1583	1490	1491
3	1120	1120	1120	941	942	1333	1313	1618	1550	1531.3	1507.75	1238	1231	1267	1240
4	1100	1095	1097	932	932	1253	1264	1299	1275	1308	1284	1097	1095	1157	1098
5	945	939	939	894.8	894.18	1054	1052	1094	1083	1095	1090	1007	1007	1013	1006
6	915	915	910	790.66	791.04	989	985	1048	1039	1035	1021	956	955	972	974
7	815	813.36	814.25	729.1	727.5	963	965	978	969	1007	1000	950	949	964	966
8	785	780.75	783.16	677	672.11	937	936	952.55	957	975	969	923	920	931	928
9	745	745.35	743.67	553.06	543.7	932	931	932	928	955.8	949.3	916	903	924	918
10	680	680.57	678.52	431.77	432.19			928	912.44	934	930				
11	610	610.3	581.3	344.05	340.47			885.94	879.08	928	927				
12	555	555	555	x	x			821.89	814.36	888.42	871.31				
13	450	448.29	450.5	x	x			793.27	775.66	826.43	815.13				
14	375	372.88	373.76	178.7	165.4										
15	295	296.04	293.56												
16	225	221.89	218.89												

***Yeast culture collection in Strains Table 1, 2 ,3 ,4 and 5**: **ATCC**= American Type Culture Collection (Manassas, EEUU), **AWRI**= Australian Wine Research Insitute (Glen Osmond, Australia), **CBS=** Centraalbureau voor Schimmeltcultures (Utrecht, the Netherlands), **NBRC**= Nite Biological Resource Center (Tokyo, Japan), **NCYC=** National Collection of Yeast Cultures (Norwich**,** United Kingdom), **NRRL**= Agricultural Research Service Culture Collection (Peoria, EEUU), **MCYC=** Microbiology Collection of Yeast Cultures (Polytechnic University of Madrid, Spain), **UWOPS**= University of Western (Ontario, Canada), **PYCC**= Portuguese Yeast Culture Collection (Caparica, Portugal), **DBVPG** (University of Perugia, Italy), **NCAIM=** National Collection of Agricultural and Industrial Microorganism **(**Budapest, Hungary), **ZIM**= Culture Collection of Industrial Microorganism (University of Ljubljana, Slovenia), **UCD**= University of California (Davis, EEUU), **UFRJ**= Universidade Federal do Rio de Janeiro (Rio de Janeiro, Brazil), **KCTC**= Korean Collection for Type Cultures (JeonGeup, South Korea), **CECT=** Spanish Type Culture Collection (University of Valencia, Spain), **CLIB**= Collection de Levures d'Intérêt Biotechnologique (Jouy-en-Josas, France). The other strain references correspond to the commercial name or a personal yeast collection.

Strains Table 1. Yeast strains for which the *FSY1* gene was searched in the databases using the BLAST algorithm: (CODE: presence: + /absence: -).

Yeast strains	**Strain Reference***	**Source/Country**	***FSY1***
Botrytis cinerea	SAS 56	Grapevine	**+**
Candida albicans	SC 5314	Clinic	**+**
Candida dubliniensis	CD 36	Clinic	**+**
Candida lusitaniae	ATCC 42720	Clinic	**-**
Candida parapsilosis	CDC 317	Clinic	**+**
Candida tropicalis	MYA-3404	Clinic	**+**
Debaryomyces hansenii	CBS 767	Carlsberg Laboratories	**+**
Eremothecium gossypii	ATCC 10895	Cotton	**-**
Kluyveromyces aestuarii	ATCC 18862	Estuarine mud	**+**
Kluyveromyces lactis	NRRL Y-1140	Cream (USA)	**+**
	JA6	Lab strain	**+**
Kluyveromyces marxianus	KCTC 17555	Fermented maize dough (Mexico)	**+**
Kluyveromyces wickerhamii	UCD 54-210	Insect (*Drosophila montana*)	**+**
Lachancea kluyveri	NRRL Y-12651	Insect (*Drosophila pinicola*)	**+**
Lachancea thermotolerans	CBS 6340	Preserved plumbs	**+**
Lodderomyces elongisporus	NRRL YB-4239	Orange juice	**+**

Meyerozyma guilliermondii	ATCC 6260	Clinic	**+**
Millerozyma farinosa	CBS 7064	Sugar (Germany)	**+**
Nakaseomyces glabrata	CBS 138	Clinic	**-**
Pichia pastoris	GS115	Unknown	**-**
Saccharomyces castellii	NRRL Y-12630	Soil	**-**
Saccharomyces cerevisiae	AWRI 1631	Wine (Australia)	**-**
	EC1118	Wine (France)	**+**
	JAY291	Bio ethanol production	**-**
	M22	Clinic	**-**
	RM11-1a	Grapevine	**-**
	S288c	Lab strain	**-**
	YJM789	Unknown	**-**
	YPS163	Clinic	**-**
	DBVPG 6765	Unknown	**+**
	SK1	Lab strain	**-**
	Y55	Lab strain	**-**
	YPS128	Environment	**-**
	DBVPG 6044	Africa	**-**
	DBVPG 1788	Finland	**-**
	DBVPG 1373	The Netherlands	**-**
	DBVPG 1853	Ethiopia	**-**
	BC187	Napa Valley	**-**
	YPS606	Pennsylvania	**-**
	L-1374	Chile	**-**
	L-1528	Chile	**-**
	Y12	Africa	**-**
	DBVPG 1106	Australia	**-**
	UWOPS 83-787.3	Bahamas (USA)	**-**
	UWOPS 87-2421	Hawaii	**-**
	K11	Japan	**-**
	YS4	The Netherlands	**-**
	YS9	Singapore	**-**
	322134S	Royal Victoria Infirmary	**-**

	378604X	Royal Victoria Infirmary	-
	273614N	Royal Victoria Infirmary	-
	YJM978	Italia	-
	Y9	Japan	-
	UWOPS 03-461.4	grapevine	-
	UWOPS 05-217.3	Malaysia	-
	W303	Lab strain	-
	UWOPS 05-227.2	Malaysia	-
	DBVPG 6040	The Netherlans	-
	YIIc17_E5	France	-
	YJM981	Italia	-
	YJM975	Italia	-
	NCYC 110	West Africa	-
Saccharomyces eubayanus	CBS 12357T	*Cyttaria hariotii* (Argentina)	+
Saccharomyces kudriavzevii	NBRC 1802	Decaying leaves (Japan)	-
Saccharomyces mikatae	NBRC 1815	Tree bark	-
Saccharomyces paradoxus	NRRL Y-17217	Exudates of trees	-
	N-45	Eastern Russia	-
	UFRJ50816	Brazilian forests	-
	N44	Eastern Russia	-
	N17	Russia	-
	T21.4	United Kingdom	-
	Q59.1	United Kingdom	-
	YPS138	USA	-
	S36.7	United Kingdom	-
	Y7	United Kingdom	-
	Q32.3	United Kingdom	-
	Z1.1	United Kingdom	-
	Q95.3	United Kingdom	-
	DBVPG 4650	Italy	-
	N43	Eastern Russia	-
	CBS 5829	Floor wasteland	-

	DBVPG 6304	USA	-
	A4	Canada	-
	A12	Canada	-
	Y6.5	United Kingdom	-
	Q62.5	United Kingdom	-
	Q89.8	United Kingdom	-
	KPN3828	Russia	-
	CBS 432	Russia	-
	KPN3829	Russia	-
	UFRJ50791	Brazil	-
	NBRC 1804	Japan	-
	CBS 1146	Unknown	-
	UWOPS 91-917.1	Hawaii	-
Saccharomyces pastorianus	Weihenstephan 34/70	Beer	+
	PYCC 4457	Beer	+
Saccharomyces uvarum	623-6C	Insect	+
	CBS 395T	Juice of *Ribes nigrum* (Netherlands)	+
Scheffersomyces stipitis	CBS 6054	Bio-ethanol production	+
Torulaspora delbrueckii	CBS 1146	Unknown	+
Vanderwaltozyma polyspora	DSM 70294	Soil (South Africa)	-
Yarrowia lipolytica	CLIB122	Unknown	-
Zygosaccharomyces bailli	CLIB 213	Beer (Japan)	+
Zygosaccharomyces rouxii	CBS 732	Grape must	+

Strains Table 2. Yeast strains for which the *FSY1* gene was searched by PCR amplification (CODE: presence: **+** /absence: **-)**. **¥** ***Sc*****=** *S. cerevisiae*, ***Su***= *S. uvarum*, ***Sk***= *S. kudriavzevii*.

Yeast strains	**Strain Reference***	**Source/Country**	***FSY1***
Hanseniaspora uvarum	CECT 11105	Most (Italy)	-
Kazachstania africanus	CECT 1963	Soil (South Africa)	-
Kazachstania barnetii	AQ 1449	Unknown	-
Kazachstania exiguus	CECT 1206	Unknown	-
Kazachstania humilis	AQ 883	Masato (South America)	-
Kazachstania kunashirensis	CECT 11346	Soil (Russia)	-
Kazachstania lodderae	CECT 1126	Soil (South Africa)	-
Kazachstania martiniae	CECT 12692	Grape juice (Spain)	-
Kazachstania rosinii	CECT 11357	Soil	-
Kazachstania servazzii	CECT 11353	Soil (Finland)	-
Kazachstania sinensis	CECT 11332	Death bird (China)	-
Kazachstania spencerorum	CECT 11347	Soil (South Africa)	-
Kazachstania transvaalensis	CECT 11354	Soil (South Africa)	-
Kazachstania unisporus	CECT 10682	Unknown	-
Kluyveromyces aestuarii	CECT 1949	Unknown	+
Kluyveromyces dobzhanskii	CECT 2284	Oaks (Hungary)	+
Kluyveromyces marxianus	CECT 1442	Unknown	+
Kluyveromyces wickerhamii	CECT 1946	*Drosophila montana* (U.S.A.)	+
Lachancea fermentati	CECT 10382	Alpechin (Spain)	-
Lachancea thermotolerans	AQ 2301	Oaks (Hungary)	**+**
Nakaseomyces bacillisporus	CECT 1979	*Quercus emoryi* exudate	-
Nakaseomyces delphensis	CECT 1954	Dried figs (South Africa)	-
Nakaseomyces glabrata	CECT 1448	Clinic	-
Naumovozyma castellii	CECT 11356	Soil (Finland)	-
Naumovozyma dairenensis	CECT 11345	Dry fruit of *Diospyros sp.*	-
Saccharomyces arboricolus	CBS 10644	Tree bark (China)	-
Saccharomyces cariocanus	NCYC 2890	Insects (Brazil)	-
Saccharomyces cerevisiae	IGAL01	Olives (Spain)	-
	FRCH	Wine (France)	+
	FCRY	Wine (France)	+

	VRB	Wine	+
	T73	Wine (Spain)	-
	CBS 2087	Lychee flower (China)	-
	CBS 8857	Sorghum beer (Western Africa)	-
	CBS 435	Sake (Japan)	-
	ZA29	Wine (South Africa)	-
	ARS	Tequila (Mexico)	-
	GUY	Tequila (Mexico)	-
	ARG7	Wine (Argentina)	+
	CECT 1885	Wine (Spain)	+
	PM	Wine	+
	K1H	Wine	-
	FCha	Wine	-
	PE7	Wine (Spain)	-
	CBS 7957	Food (Brazil)	+
	CLIB 382	Super-attenuated beer (Japan)	+
Saccharomyces eubayanus	PE35M	Ferment (America)	-
	TEMOHAYA	Agave (Mexico)	-
Saccharomyces kudriavzevii	S20	Cider	+
	CECT 11186	Beer (Denmark)	+
	CR90	Oak bark (Spain)	-
	CR85	Oak bark (Spain)	-
Saccharomyces paradoxus	CR89	Oak bark (Spain)	-
	CA111	Oak bark (Spain)	-
	120M	Pulque (Mexico)	-
	CECT 1939	Trees	-
	K54	Vineyards (Croatia)	-
	CBS 5313	Pulque (Mexico)	-
Saccharomyces pastorianus	CBS 406	Trees (Holland)	-
Saccharomyces uvarum	115M	Pulque (Mexico)	-
	CECT 1940	Beer (Denmark)	+
	CECT 12627	Wine (Spain)	+
	BMV58	Wine (Spain)	+
	NCAIM 789	Tree exudates (Hungary)	+
	ZIM 2113	Most (Slovenia)	+
¥ Híbrido Sc x Su	CBS 2986	Wine (Swiss)	+
¥ Híbrido Sc x Sk	CECT 11036	Beer altered	+

¥ Híbrido Sc x Sk	S6U	Wine	+
¥ Híbrido Sc x Sk	AWRI 1503	Wine (Australia)	+
¥ Híbrido Sc x Sk	AMH	Wine (Swiss)	-
¥ Híbrido Sc x Sk	W27	Wine (Swiss)	+
¥ Híbrido Sc x Sk x Su	SOY3	Wine (Croatia)	-
Tetrapisispora blattae	HA1837	Wine (Swiss)	-
Tetrapisispora phaffii	CBS 2834	Wine (Swiss)	+
Torulaspora delbrueckii	CECT 1964	Insects (Germany)	-
Torulaspora delbrueckii	CECT 10646	Soil (South Africa)	-
Torulaspora franciscae	AGAVE	Agave fermentation (Mexico)	+
Torulaspora globosa	CECT 10039	Fraxinus angustNBRClia (Spain)	+
Torulaspora microellipsoides	CECT 10680	Soil (Spain)	+
	CECT 10655	Soil (Caribbean)	-
	CLIB830T	Apple juice (Germany)	+
	NRRLY-17058	Tea-beer (Finland)	+
Zygosaccharomyces bailii	CBS 6762	Lemonade (Switzerland)	+
	CBS 6641	Exudate of *Myoporum sp.* (U.S.A.)	+
Zygosaccharomyces bisporus	CBS 6143	Tea-beer (Finland)	+
Zygosaccharomyces fermentati	CECT 11042	Grape must (Italy)	+
Zygosaccharomyces mellis	CECT 11055	Tea-beer	+
Zygotorulaspora mrakii	CECT 10382	Alpechin (Spain)	+
	CECT 11057	Honey (USA)	+
Zygotorulaspora florentinus	CECT 10531	Alpechin (Spain)	+
Vanderwaltozyma yarrowii	CECT 10529	Alpechín (Spain)	+
	CECT 11200	Sulphated grape most	+
	CECT 1958	Unknown	-

Strains Table 3. Yeast strains for which the *T. microellipsoides ATO3* gene (present in the described Region C) was searched by using sequences obtained either in our lab or databases.

Yeast strains	Strain Reference*	Source/Country
Candida glabrata	CBS 138	Faeces of Man
Eremothecium cymbalariae	DBVPG 7215	*Brachynema germari* (Iran)
Eremothecium gossypii	ATCC 10895	Cotton
Kazachstania africanus	CBS 2517	Soil (South Africa)
Kazachstania naganishii	CBS 8797	Decaying leaves (Japan)
Kluyveromyces lactis	CLIB 210	Lab strain
Lachancea thermotolerans	CBS 6340	Mirabelle-plum conserve (Russia)
Lachancea waltii	NCYC2644	Exudate of *Ilex integra* (Japan)
Lachancea kluyveri	CBS 3082	*Drosophila pinicola*
Naumovozyma castellii	CBS 4309	Soil (Finland)
Naumovozyma dairenensis	CBS 421	Dry fruit of *Diospyrus* sp.
Saccharomyces cerevisiae	EC1118	Wine (France)
	S288c	Lab strain
	CBS 7957	Food (Brazil)
	CLIB 382	Super-attenuated beer (Japan)
	AWRI 796	Wine (Australia)
	RM11-1a	Vineyard (USA)
	K7	Sake(Japan)
Saccharomyces eubayanus	CBS 12357T	*Cyttaria hariotii* (Argentina)
Saccharomyces kudriavzevii	NBRC 1802	Decaying leaves (Japan)
Saccharomyces mikatae	NBRC 1815	Tree bark
Saccharomyces paradoxus	NRRL Y-17217	Exudates of trees
	UFRJ50816	*Drosophila* spp. (Brazil)
	YPS138	Soil beneath *Quercus velutina* (U.S.A.)
	UWOPS91-917.1	Flux of *Myoporum sandwichense* (Hawaii)
	N-44	Exudate of *Quercus mongolica* (Russia)
	CBS432	Bark of *Quercus* spp. (Russia)
Saccharomyces uvarum	BMV58	Wine (Spain)
	CECT 12600	Mistela (Spain)
	CECT 1969	Juice of Ribes *nigrum* (Netherlands)

	NPCC 1290	Traditional fermentation (Argentina)
Tetrapisispora blattae	CBS 6284	Gut of *Blatta orientalis*
Tetrapisispora phaffii	CBS 4417	Uncultivated urface soil (South Africa)
Torulaspora delbrueckii	CBS 1146	Unknown
Torulaspora microellipsoides	CLIB830T	Apple juice (Germany)
	NRRLY-17058	Tea-beer (Finland)
	CBS 6762	Lemonade (Switzerland)
	CBS 6641	Exudate of *Myoporum* sp. (U.S.A.)
	CBS 6143	Tea-beer (Finland)
Vanderwaltozyma polyspora	DSMZ 70294	Soil (South Africa)
Zygosaccharomyces bailli	CLIB 213	Beer (Japan)
Zygosaccharomyces rouxii	CBS 732	Grape must

Strains Table 4. *Torulaspora* strains for which the 26S rDNA gene was obtained from *genebank*.

Accession num.	*Torulaspora* strains	Source/Origin
AB836696.1	*Torulaspora sp*. **DMKU-SV13**	soil in vetiver rhizosphere
AB836684.1	*Torulaspora sp.* **DMKU-SV01**	soil in vetiver rhizosphere
HE574639.1	*Torulaspora sp.* **APSS 805**	coal mine soil
KY109871.1	*Torulaspora indica* **CBS 12408**	Coal mine soil, Singareni Collries (India)
AB499015.1	*Torulaspora globosa* **LY10**	soil from forest (Thailand)
AB499987.1	*Torulaspora globosa* **SSK8**	soil from forest (Thailand)
AB499988.1	*Torulaspora globosa* **SSK9**	soil from forest (Thailand)
AB836685.1	*Torulaspora globosa* **DMKU:SV02**	soil in vetiver rhizosphere
AB836686.1	*Torulaspora globosa* **DMKU:SV03**	soil in vetiver rhizosphere
AB836687.1	*Torulaspora globosa* **DMKU:SV04**	soil in vetiver rhizosphere
AB836689.1	*Torulaspora globosa* **DMKU:SV06**	soil in vetiver rhizosphere
AB836690.1	*Torulaspora globosa* **DMKU:SV07**	soil in vetiver rhizosphere
AB836691.1	*Torulaspora globosa* **DMKU:SV08**	soil in vetiver rhizosphere
AB836692.1	*Torulaspora globosa* **DMKU:SV09**	soil in vetiver rhizosphere
AB836693.1	*Torulaspora globosa* **DMKU:SV10**	soil in vetiver rhizosphere
AB836694.1	*Torulaspora globosa* **DMKU:SV11**	soil in vetiver rhizosphere
AB836695.1	*Torulaspora globosa* **DMKU:SV12**	soil in vetiver rhizosphere
AB836697.1	*Torulaspora globosa* **DMKU:SV14**	soil in vetiver rhizosphere
AB836698.1	*Torulaspora globosa* **DMKU:SV15**	soil in vetiver rhizosphere
AB836712.1	*Torulaspora globosa* **DMKU:SV29**	soil in vetiver rhizosphere
AB836713.1	*Torulaspora globosa* **DMKU:SV30**	soil in vetiver rhizosphere
AB836714.1	*Torulaspora globosa* **DMKU:SV31**	soil in vetiver rhizosphere
AB836724.1	*Torulaspora globosa* **DMKU:SV41**	soil in vetiver rhizosphere
AB836725.1	*Torulaspora globosa* **DMKU:SV42**	soil in vetiver rhizosphere
AB836727.1	*Torulaspora globosa* **DMKU:SV45**	soil in vetiver rhizosphere
AB836729.1	*Torulaspora globosa* **DMKU:SV47**	soil in vetiver rhizosphere
KJ159059.1	*Torulaspora globosa* **LB376**	soil sample (Brazil)
AB836688.1	*Torulaspora globosa* **DMKU:SV05**	soil in vetiver rhizosphere
JF327428.1	*Torulaspora sp.* **PJ I3**	Thailand
HE575669.1	*Torulaspora sp*. **AP 18**	soil
U72166.1	*Torulaspora globosa* **NRRL Y-12650**	soil (West Indies)
KF300889.1	*Torulaspora globosa* **CBS 764**	soil (West Indies)
KF300888.1	*Torulaspora globosa* **CBS 2947**	soil (Papua New Guinea)
HE575668.1	*Torulaspora sp.* **PBA 22**	soil
AB303872.1	*Torulaspora maleeae* **NBRC 103203**	mangrove (Thailand)
AB303871.1	*Torulaspora maleeae* **NBRC 103202**	moss (Thailand)
AB303870.1	*Torulaspora maleeae* **NBRC 103201**	moss (Thailand)
AB303869.1	*Torulaspora maleeae* **NBRC 103200**	moss (Thailand)

	Torulaspora sp. **SR3**	soil from forest (Thailand)
AB500184.1		
AB456554.1	*Torulaspora sp.* **WB17**	water in mangrove forest (Thailand)
AB087395.1	*Torulaspora maleeae* **NBRC 11061**	leaf, Rhizophora *stylosa* (Japan)
AB303868.1	*Torulaspora maleeae* **NBRC 103199**	moss (Thailand)
AB303867.1	*Torulaspora maleeae* **NBRC 103198**	moss (Thailand)
AB303866.1	*Torulaspora maleeae* **NBRC 11062**	rhizosphere of mangrove, Bruguiera *gymnorrhiza* (Japan)
FJ527215.1	*Torulaspora sp*. **GE3S10**	soil (Taiwan)
FJ527214.1	*Torulaspora sp.* **EN22S16**	soil (Taiwan)
FJ527213.1	*Torulaspora sp.* **EN11S09**	soil (Taiwan)
AB714269.1	*Torulaspora sp*. **KKU-PM21**	Thailand
U72157.1	*Torulaspora pretoriensis* **NRRL Y-17251**	soil (South Africa)
KF300887.1	*Torulaspora pretoriensis* **CBS 2187**	soil (South Africa)
AB617947.1	*Torulaspora pretoriensis* **LM045**	leaves (Thailand)
FJ873560.1	*Torulaspora pretoriensis* **GU12S05**	soil (Taiwan)
U73604.1	*Torulaspora franciscae* **NRRL Y-17532**	mangrove sediments from east and west coast (India)
KF300890.1	*Torulaspora franciscae* **CBS 2926**	soil (Spain)
FJ527110.1	*Torulaspora sp*. **GE1L03**	plant (Taiwan)
KF300892.1	*Torulaspora delbrueckii* **CBS 6991**	drink
KF300893.1	*Torulaspora delbrueckii* **CBS 133**	food (Indonesia)
KF300891.1	*Torulaspora delbrueckii* **CBS 2734**	missing
AJ508558.1	*Torulaspora delbrueckii* **CBS 133T**	food (Indonesia)
EU289351.1	*Torulaspora sp*. **SG5S08**	Taiwan
FJ888525.1	*Torulaspora quercuum* **XZ-46A**	oral cavities of healthy Tibetan volunteers
FJ888524.1	*Torulaspora quercuum* **AS 2.3768**	Leaf of Quercus sp. (China)
KJ183053.1	*Torulaspora quercuum* **YF3**	yak milk dreg (Tibet)
KJ183052.1	*Torulaspora quercuum* **Y7-4**	yak milk dreg (Tibet)
KJ183051.1	*Torulaspora quercuum* **Y7-3**	yak milk dreg (Tibet)
HE660063.1	*Torulaspora quercuum* **ZIM 2412**	cheese (Serbia)

Strains Table 5. *Torulaspora* strains for which the 18S/ITS/5.8S rDNA genes were obtained from *genebank* and the CBS yeast culture collection.

Accession num.	*Torulaspora* strains	Source and Origin
KX859660.1	*Torulaspora delbrueckii* **URFM433**	cheese (France)
KX859657.1	*Torulaspora delbrueckii* **URFM432**	cheese (France)
KX859663.1	*Torulaspora delbrueckii* **URFM434**	cheese (France)
KX859666.1	*Torulaspora delbrueckii* **URFM435**	cheese (France)
KX859667.1	*Torulaspora delbrueckii* **URFM509**	cheese (France)
KX859672.1	*Torulaspora delbrueckii* **URFM486**	cheese (France)
KX859677.1	*Torulaspora delbrueckii* **URFM438**	cheese (France)
KX859680.1	*Torulaspora delbrueckii* **URFM439**	cheese (France)
KX859683.1	*Torulaspora delbrueckii* **URFM440**	cheese (France)
KX859697.1	*Torulaspora delbrueckii* **URFM445**	cheese (France)
KX859713.1	*Torulaspora delbrueckii* **URFM449**	cheese (France)
KX859725.1	*Torulaspora delbrueckii* **URFM453**	cheese (France)
KX859730.1	*Torulaspora delbrueckii* **URFM455**	cheese (France)
KX859733.1	*Torulaspora delbrueckii* **URFM457**	cheese (France)
KX859739.1	*Torulaspora delbrueckii* **URFM459**	cheese (France)
KX859741.1	*Torulaspora delbrueckii* **URFM461**	cheese (France)
KX859743.1	*Torulaspora delbrueckii* **URFM463**	cheese (France)
KX859744.1	*Torulaspora delbrueckii* **URFM464**	cheese (France)
KX859745.1	*Torulaspora delbrueckii* **URFM465**	cheese (France)
KX859746.1	*Torulaspora delbrueckii* **URFM466**	cheese (France)
KX859747.1	*Torulaspora delbrueckii* **URFM467**	cheese (France)
KX859750.1	*Torulaspora delbrueckii* **URFM468**	cheese (France)
KX859751.1	*Torulaspora delbrueckii* **URFM469**	cheese (France)
KX859752.1	*Torulaspora delbrueckii* **URFM470**	cheese (France)
KX859753.1	*Torulaspora delbrueckii* **URFM471**	cheese (France)
KX859755.1	*Torulaspora delbrueckii* **URFM473**	cheese (France)
KX859756.1	*Torulaspora delbrueckii* **URFM474**	cheese (France)
KX859757.1	*Torulaspora delbrueckii* **URFM475**	cheese (France)
KX148748.1	*Torulaspora sp.* **NWHC 44736-64-03-01B**	snake
KX859687.1	*Torulaspora delbrueckii* **URFM515**	cheese (France)
FJ153205.1	*Torulaspora sp.* **SG5S08**	Taiwan
KY105669.1	*Torulaspora pretoriensis* **CBS 2187**	soil (South Africa)
KF300901.1	*Torulaspora pretoriensis* **CBS 2187**	soil (South Africa)
AY046188.1	*Torulaspora pretoriensis* **NRRL Y-17251**	soil (South Africa)
HM044873.1	*Torulaspora sp.* **TCJ252**	Asia
KY977702.1	*Torulaspora sp.* **QFP5Y1**	larval gut, Bactrocera tryoni (Australia)
KX218261.1	*Torulaspora pretoriensis* **AUMC 10292**	mud sample from fresh water canal (Egypt)
KY105667.1	*Torulaspora pretoriensis* **CBS 2785**	soil (South Africa)
KY105650.1	*Torulaspora franciscae* **CBS 2926**	soil (Spain)

KY105672.1	*Torulaspora pretoriensis* **CBS 11123**	Unknown
KY105668.1	*Torulaspora pretoriensis* **CBS 11124**	Unknown
KY105670.1	*Torulaspora pretoriensis* **CBS 11121**	Unknown
KY105671.1	*Torulaspora pretoriensis* **CBS 11100**	Unknown
KY105651.1	*Torulaspora franciscae* **CBS 2927**	soil (Spain)
AY046186.1	*Torulaspora franciscae* **NRRL Y-17532**	mangrove sediments from east and west coast (India)
KJ706695.1	*Torulaspora franciscae* **DBMY478**	mangrove sediments from east and west coast (India)
KJ707171.1	*Torulaspora franciscae* **DBMY954**	mangrove sediments from east and west coast (India)
KJ706377.1	*Torulaspora franciscae* **DBMY158**	mangrove sediments from east and west coast (India)
FJ153222.1	*Torulaspora franciscae* **NRRL Y-6686**	soil (Spain)
KJ707013.1	*Torulaspora franciscae* **DBMY796**	mangrove sediments from east and west coast (India)
KJ706536.1	*Torulaspora franciscae* **DBMY319**	mangrove sediments from east and west coast (India)
KJ706854.1	*Torulaspora franciscae* **DBMY637**	mangrove sediments from east and west coast (India)
AB304154.1	*Torulaspora maleeae* **NBRC 103198**	moss (Thailand)
AB304155.1	*Torulaspora maleeae* **NBRC 103199**	moss (Thailand)
AB304153.1	*Torulaspora maleeae* **NBRC 11062**	soil, rhizosphere of Bruguiera gymnorrhiza (Japan)
AB304152.1	*Torulaspora maleeae* **NBRC 11061**	leaf (Japan)
KY105663.1	*Torulaspora maleeae* **CBS 10694**	plant (Japan)
NR 136949.1	*Torulaspora maleeae* **NBRC 11061**	leaf, Rhizophora *stylosa* (Japan)
AB304160.1	*Torulaspora maleeae* **NBRC 103204**	moss (Thailand)
AB304159.1	*Torulaspora maleeae* **NBRC 103203**	mangrove (Thailand)
AB304156.1	*Torulaspora maleeae* **NBRC 103200**	moss (Thailand)
AB304157.1	*Torulaspora maleeae* **NBRC 103201**	moss (Thailand)
AB304158.1	*Torulaspora maleeae* **NBRC 103202**	moss (Thailand)
FJ873436.1	*Torulaspora sp.* **EN11S09**	Taiwan
HM461650.1	*Torulaspora sp.* **EN22S16 SB100**	Taiwan
KY105658.1	*Torulaspora globosa* **CBS 5216**	soil (Italy)
KY105653.1	*Torulaspora globosa* **CBS 2952**	soil (South Africa)
KY105662.1	*Torulaspora indica* **CBS 12408**	Coal mine soil, Singareni Collries (India)
KY105659.1	*Torulaspora globosa* **CBS 6638**	Unknown
KY105655.1	*Torulaspora globosa* **CBS 6636**	Unknown
KY105657.1	*Torulaspora globosa* **CBS 6637**	Unknown
KY105656.1	*Torulaspora globosa* **CBS 764**	soil (West Indies)
KY105661.1	*Torulaspora globosa* **CBS 765**	Unknown
AY046184.1	*Torulaspora globosa* **NRRL Y-12650**	soil (West Indies)
KF300895.1	*Torulaspora globosa* **CBS 2947**	soil (Papua New Guinea)
KY105652.1	*Torulaspora globosa* **CBS 5500**	mating
KY105654.1	*Torulaspora globosa* **CBS 2947**	soil (Papua New Guinea)

KY105660.1	*Torulaspora globosa* **CBS 5503**	mating
FJ888525.1	*Torulaspora quercuum* **XZ-46A**	oral cavities of healthy Tibetan volunteers
FJ888524.1	*Torulaspora quercuum* **AS 2.3768**	Leaf of Quercus sp. (China)
KY105673.1	*Torulaspora quercuum* ***CBS 11403***	plant (China)

Bibliography

Almeida P et al. 2015. A population genomics insight into the Mediterranean origins of wine yeast domestication. Mol. Ecol. 24:5412–5427. Available from:http://dx.doi.org/10.1111/mec.13341

Basso RF et al. 2016. Could non-*Saccharomyces* yeasts contribute on innovative brewing fermentations? Food Res. Int. 86:112–120. Available from:http://dx.doi.org/10.1016/j.foodres.2016.06.002

Berthels NJ et al. 2004. Discrepancy in glucose and fructose utilisation during fermentation by *Saccharomyces cerevisiae* wine yeast strains. FEMS Yeast Res. 4:683–689. Available from:http://dx.doi.org/10.1016/j.femsyr.2004.02.005

Boles E et al. 1997. The molecular genetics of hexose transport in yeasts. FEMS Microbiol. Rev. 21:85–111. Available from:http://dx.doi.org/10.1111/j.1574-6976.1997.tb00346.x

Borneman AR 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. Available from: http://dx.doi.org/10.1111/j.1567-1364.2011.00773.x

Borneman AR et al. 2015. Genomic insights into the *Saccharomyces* sensu stricto complex. Genetics 2 281-291. Available from: http://dx.doi.org/10.1534/genetics.114.173633

Brown CA et al. 2010. Rapid expansion and functional divergence of subtelomeric gene families in yeasts. Curr. Biol. 20:895–903. Available from: https://doi.org/10.1016/j.cub.2010.04.027

Butler G et al. 2004. Evolution of the *MAT* locus and its Ho endonuclease in yeast species. Proc. Natl. Acad. Sci. 101:1632–1637. Available from: http://doi.org/10.1073/pnas.0304170101

Carver T et al. 2012. Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics. 28:464-469. Available from: http://dx.doi.org/10.1093/bioinformatics/btr703

Chen D et al. 2016. Impact of simultaneous and sequential fermentation with *Torulaspora delbrueckii* and *Saccharomyces cerevisiae* on non-volatiles and volatiles of lychee wines. LWT - Food Sci. Technol. 65:53–61. Available from: http://dx.doi.org/10.1016/j.lwt.2015.07.050

Ciani M et al. 2010. Controlled mixed culture fermentation: a new perspective on the use of non-*Saccharomyces* yeasts in winemaking. FEMS Yeast Res.10:123-133. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19807789

Coelho MA et al. 2013. Extensive Intra-Kingdom Horizontal Gene Transfer Converging on a Fungal Fructose Transporter Gene. PLoS Genet. e1003587. Available from: https://doi.org/10.1016/j.mib.2003.10.005

Conant GC et al. 2007. Increased glycolytic flux as an outcome of whole-genome duplication in yeast. Mol. Syst. Biol. 3:129. Available from: http://doi.org/10.1038/msb4100170

Deatherage DE and Barrick JE 2014. Identification of mutations in laboratory evolved microbes from next-generation sequencing data using breseq. Methods Mol. Biol. 1151:165–188. Available from: https://doi.org/10.1007/978-1-4939-0554-6_12

Diezemann A and Boles E 2003. Functional characterization of the Frt1 sugar transporter and of fructose uptake in *Kluyveromyces lactis*. Curr. Genet. 43:281–288. Available from: http://doi.org/10.1007/s00294-003-0392-5

Doehlemann G 2005. Molecular and functional characterization of a fructose specific transporter from the gray mold fungus *Botrytis cinerea*. Fungal Genet. Biol.7:601-610. Available from: http://doi.org/10.1016/j.fgb.2005.03.001

Dujon BA and Louis EJ. 2017. Genome Diversity and Evolution in the Budding Yeasts (Saccharomycotina). 206:717–750. Available from: http://doi.org/10.1534/genetics.116.199216

Dujon BG et al. 2004. Genome evolution in yeasts. Nature 430:35–44. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15229592

Dunn B and Sherlock G. 2008. Reconstruction of the genome origins and evolution of the hybrid lager yeast *Saccharomyces pastorianus*. Genome Res. 18:1610–1623. Available from: http://doi.org/10.1101/gr.076075.108

Esteve-Zarzoso B et al. 1998. The role of non-*Saccharomyces* yeasts in industrial winemaking. Int. Microbiol. 1:143–148. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10943352

Fay JC and Benavides JA 2005[a]. Evidence for Domesticated and Wild Populations of *Saccharomyces cerevisiae*. PLOS Genet. 1:e5. Available from: https://doi.org/10.1371/journal.pgen.0010005

Fay JC and Benavides JA 2005[b]. Hypervariable noncoding sequences in *Saccharomyces cerevisiae*. Genetics 170:1575–1587. Available from: http://doi.org/10.1534/genetics.105.042283

Fleet GH. 2003. Yeast interactions and wine flavour. Int. J. Food Microbiol. 86:11–22. Available from: http://doi.org/10.1016/S0168-1605(03)00245-9

Fleet GH. 2007. Yeasts in foods and beverages: impact on product quality and safety. Curr. Opin. Biotechnol.18:170–175. Available from: https://doi.org/10.1016/j.copbio.2007.01.010

Galeote V et al 2011. Amplification of a *Zygosaccharomyces bailii* DNA segment in wine yeast genomes by extrachromosomal circular DNA formation. PLoS One. 6: e17872. Available from: http://doi.org/10.1371/journal.pone.0017872

Galeote V et al 2010. *FSY1*, a horizontally transferred gene in the *Saccharomyces cerevisiae* EC1118 wine yeast strain, encodes a high-affinity fructose/H+ symporter. Microbiology. Available from: http://dx.doi.org/10.1099/mic.0.041673-0

Goffeau A. et al 1996. Life with 6000 genes. Science 274:563–567. Available from: http://dx.doi.org/10.1126/SCIENCE.274.5287.546

Gonçalves P et al 2000. *FSY1*, a novel gene encoding a specific fructose/H(+) symporter in the type strain of *Saccharomyces carlsbergensis*. J. Bacteriol. 182:5628–5630. Available from: http://doi.org/10.1128/JB.182.19.5628-5630.2000

Gordon JL et al 2009. Additions, Losses, and Rearrangements on the Evolutionary Route from a Reconstructed Ancestor to the Modern *Saccharomyces cerevisiae* Genome. PLOS Genet.5:e1000485. Available from: https://doi.org/10.1371/journal.pgen.1000485

Hagman A et al. 2013. Yeast "Make-Accumulate-Consume" Life Strategy Evolved as a Multi-Step Process That Predates the Whole Genome Duplication. PLoS One 8:e68734. Available from: https://doi.org/10.1371/journal.pone.0068734

Hittinger CT. 2013. *Saccharomyces* diversity and evolution: A budding model genus. Trends Genet.29:5 Available from: http://dx.doi.org/10.1016/j.tig.2013.01.002

James SA et al. 2005. Molecular evidence for the existence of natural hybrids in the genus *Zygosaccharomyces*. FEMS Yeast Res. 5:747–755. Available from: https://doi.org/10.1016/j.femsyr.2005.02.004

Jones T et al. 2004. The diploid genome sequence of *Candida albicans*. Proc. Natl. Acad. Sci. U. S. A. 101:7329 LP-7334.
Available from: http://doi.org/ 10.1073/pnas.0401648101

Jordan P et al. 2016. Hxt13, Hxt15, Hxt16 and Hxt17 from *Saccharomyces cerevisiae* represent a novel type of polyol transporters. Sci. Rep. 6:23502.
Available from: http://dx.doi.org/10.1038/srep23502

Kellis M et al. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature 428:617.
Available from: http://dx.doi.org/10.1038/nature02424

Knop M. 2006. Evolution of the hemiascomycete yeasts: on life styles and the importance of inbreeding. Bioessays 28:696–708.
Available from: http://www.ncbi.nlm.nih.gov/pubmed/16929561

Kowallik V and Greig D. 2016. A systematic forest survey showing an association of *Saccharomyces paradoxus* with oak leaf litter. Environ. Microbiol. Rep. 8:833–841. Available from: http://dx.doi.org/10.1111/1758-2229.12446

Kurtzman C. 2003. Phylogenetic circumscription of *Saccharomyces*, *Kluyveromyces* and other members of the Saccharomycetaceae, and the proposal of the new genera *Lachancea*, *Nakaseomyces*, *Naumovia*, *Vanderwaltozyma* and *Zygotorulaspora*. FEMS Yeast Res. 3:233-245
Available from: https://dx.doi.org/10.1016/S1567-1356(03)00175-2

Kurtzman C. 2003. Phylogenetic circumscription of , and other members of the Saccharomycetaceae, and the proposal of the new genera *Lachancea*, *Nakaseomyces*, *Naumovia*, *Vanderwaltozyma and Zygotorulaspora* and FEMS Yeast Res. 4:233–245.
Available from: http://dx.doi.org/10.1016/S1567-1356(03)00175-2

Kurtzman CP, Fell JW, Boekhout T. 2011. The Yeasts: A Taxonomic Study 5th edn. Amsterdam: Elsevier

Kurtzman CP et al. 2015. Advances in yeast systematics and phylogeny and their use as predictors of biotechnologically important metabolic pathways. FEMS Yeast Res. 15:1–17. Available from: http://dx.doi.org/10.1093/femsyr/fov050

Kurtzman CP and Robnett CJ. 2003. Phylogenetic relationships among yeasts of the "*Saccharomyces* complex" determined from multigene sequence analyses. FEMS Yeast Res. 3:417–432.
Available from: http://doi.org/10.1016/S1567-1356(03)00012-6

Leandro MJ et al. 2011. The osmotolerant fructophilic yeast *Zygosaccharomyces rouxii* employs two plasma-membrane fructose uptake systems belonging to a new family of yeast sugar transporters. Microbiology 157:601–608. Available from: https://doi.org/10.1099/mic.0.044446-0

Legras J et al. 2007. Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. Mol. Ecol. 16:2091–2102. Available from: https://doi.org/10.1111/j.1365-294X.2007.03266.x

Libkind D et al. 2011. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. Proc. Natl. Acad. Sci. 108:14539–14544. Available from: http://doi.org/10.1073/pnas.1105430108

Liti G et al. 2009. Population genomics of domestic and wild yeasts. Nature 458:337–341. Available from: http://doi.org/10.1038/nature07743

Loira I et al. 2015. Use of *Schizosaccharomyces pombe* and *Torulaspora delbrueckii* strains in mixed and sequential fermentations to improve red wine sensory quality. Food Res. Int. 76:325–333. Available from: http://dx.doi.org/10.1016/j.foodres.2015.06.030

Lopes CA et al. 2010. Natural hybrids of *S. cerevisiae* × *S. kudriavzevii* share alleles with European wild populations of *Saccharomyces kudriavzevii*. FEMS Yeast Res. 10:412–421. Available from: http://dx.doi.org/10.1111/j.1567-1364.2010.00614.x

Marcet-Houben M and Gabaldón T. 2015. Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker's Yeast Lineage.Hurst LD, editor. PLOS Biol. 13:e1002220. Available from: http://doi.org/10.1371/journal.pbio.1002220

Marinoni G et al. 1999. Horizontal transfer of genetic material among *Saccharomyces* yeasts. J. Bacteriol. 181:6488–6496 Available from:https://jb.asm.org/content/181/20/6488.long

Marsit S and Dequin S. 2015. Diversity and adaptive evolution of *Saccharomyces* wine yeast: a review. FEMS Yeast Res. 15:fov067-fov067. Available from: http://dx.doi.org/10.1093/femsyr/fov067

Marsit S et al. 2015. Evolutionary advantage conferred by an eukaryote-to-eukaryote gene transfer event in wine yeasts. Mol. Biol. Evol. 32:1695–1707. Available from: http://dx.doi.org/10.1093/molbev/msv057

Moore KJ et al. 1988. Indigenous Yeast Microflora on Arkansas White Riesling (Vitis vinifera) Grapes and in Model Must Systems. J. Food Sci. 53:1725–1728. Available from:http://dx.doi.org/10.1111/j.1365-2621.1988.tb07827.x

Nakao Y et al. 2009. Genome sequence of the lager brewing yeast, an interspecies hybrid. DNA Res. 16:115–129. Available from: http://dx.doi.org/10.1093/dnares/dsp003

Naseeb S et al. 2017. *Saccharomyces jurei* sp. nov., isolation and genetic identification of a novel yeast species from *Quercus robur*. 67:2046–2052. Available from: http://dx.doi.org/10.1099/ijsem.0.002013

Naumov GI. 2000. *Saccharomyces bayanus* var. *Uvarum* comb.nov., a new variety established by genetic analysis. Mikrobiologiia 69:410–414. Available from: https://link.springer.com/article/10.1007/BF02756744

Naumov GI et al. 2000. Three new species in the *Saccharomyces sensu stricto* complex: *Saccharomyces cariocanus*, *Saccharomyces kudriavzevii* and *Saccharomyces mikatae*. Int. J. Syst. Evol. Microbiol. 50 Pt 5:1931–1942. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11034507

Nguyen H-V and Gaillardin C. 1997. Two Subgroups within the *Saccharomyces bayanus* Species Evidenced by PCR Amplification and Restriction Polymorphism of the Non-Transcribed Spacer 2 in the Ribosomal DNA Unit. Syst. Appl. Microbiol. 20:286–294. Available from: http://dx.doi.org/10.1016/S0723-2020(97)80075-6

Novo M et al. 2009. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. Proc. Natl. Acad. Sci. 106:16333–16338. Available from: https://doi.org/10.1073/pnas.0904673106

Okuno M et al. 2016. Next-generation sequencing analysis of lager brewing yeast strains reveals the evolutionary history of interspecies hybridization. DNA Res. 23:67–80. Available from: http://dx.doi.org/10.1093/dnares/dsv037

Pérez-Ortín JE et al. 2002. Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains. Genome Res.12:1533–1539. Available from: http://doi.org/10.1101/gr.436602

Peris D et al. 2012. Comparative genomics among *Saccharomyces cerevisiae* × *Saccharomyces kudriavzevii* natural hybrid strains isolated from wine and beer reveals different origins. BMC Genomics 13:407. Available from: https://doi.org/ 10.1186/1471-2164-13-407

Peris D et al. 2014. Population structure and reticulate evolution of *Saccharomyces eubayanus* and its lager-brewing hybrids. Mol. Ecol. 23:2031–2045. Available from: https://doi.org/10.1111/mec.12702

Peter J et al. 2018. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. Nature 556:339–344. Available from: https://doi.org/10.1038/s41586-018-0030-5

Pina C et al. 2004. Ffz1, a new transporter specific for fructose from *Zygosaccharomyces bailii*. Microbiology 150:2429–2433. Available from: https://doi.org/10.1099/mic.0.26979-0

Piškur J et al. 2006. How did *Saccharomyces* evolve to become a good brewer? Trends Genet. 22:183–186. Available from: https://doi.org/ 10.1016/j.tig.2006.02.002

Proux-Wéra E et al. 2012. A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. BMC Bioinformatics 13:237. Available from: https://doi.org/10.1186/1471-2105-13-237

Pulvirenti A et al. 2006. *Saccharomyces uvarum*, a proper species within *Saccharomyces sensu stricto*. FEMS Microbiol. Lett. 192:191–196. Available from: https://doi.org/10.1111/j.1574-6968.2000.tb09381.x

Querol A et al. 2003. Adaptive evolution of wine yeast. Int. J. Food Microbiol. 86:3–10. Available from: https://doi.org/10.1016/S0168-1605(03)00244-7

Rodrigues F et al. 2003. The spoilage yeast *Zygosaccharomyces bailii* forms mitotic spores: A screening method for haploidization. Appl. Environ. Microbiol. 69:649–653. Available from: http://doi.org/10.1128/AEM.69.1.649-653.2003

Rodrigues de Sousa H et al. 2004. Differential regulation by glucose and fructose of a gene encoding a specific fructose/H+ symporter in *Saccharomyces sensu stricto* yeasts. Yeast 21:519–530. Available from: http://dx.doi.org/10.1002/yea.1118

Rodrigues de Sousa H et al. 1990. Active fructose transport in *Saccharomyces sensu stricto*: taxonomic implications. In: Acta Varia. Acta Varia. p. 27–134.

Rolland T et al. 2009. Insertion of horizontally transferred genes within conserved syntenic regions of yeast genomes. PLoS One 4:e6515. Available from: http://doi.org/10.1371/journal.pone.0006515

Saluja P et al. 2012. *Torulaspora indica* a novel yeast species isolated from coal mine soils. Antonie Van Leeuwenhoek 101:733–742. Available from: https://doi.org/10.1007/s10482-011-9687-6

Sampaio JP and 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. Microbiol. 74:2144–2152. Available from: http://doi.org/ 10.1128/AEM.02396-07

Shull GH. 1948. What is "heterosis"? Genetics 33:439 LP-446. Available from: http://www.genetics.org/content/33/5/439.abstract

Sicard D and Legras JL. 2011. Bread, beer and wine: Yeast domestication in the *Saccharomyces sensu stricto* complex. Comptes Rendus - Biol.334:229–236. Available from: http://dx.doi.org/10.1016/j.crvi.2010.12.016

Smith MM. 1987. Molecular evolution of the *Saccharomyces cerevisiae* histone gene loci. J. Mol. Evol. 24:252–259. Available from: https://doi.org/10.1007/BF02111238

Solieri L et al. 2008. Genome size and ploidy level: New insights for elucidating relationships in *Zygosaccharomyces* species. Fungal Genet. Biol. 45:1582–1590. Available from: https://doi.org/10.1016/j.fgb.2008.10.001

Solieri L et al. 2007. A new putative *Zygosaccharomyces* yeast species isolated from traditional balsamic vinegar. Yeast [Internet] 24:403–417. Available from: http://doi.wiley.com/10.1002/yea.1471

Suh S-O et al. 2013. Proposal of *Zygosaccharomyces parabailii* sp. nov. and *Zygosaccharomyces pseudobailii* sp. nov., novel species closely related to *Zygosaccharomyces bailii*. Int. J. Syst. Evol. Microbiol. 63:1922–1929. Available from: http://doi.org/10.1099/ijs.0.048058-0

Tamura K et al. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30:2725–2729. Available from: http://dx.doi.org/10.1093/molbev/mst197

Thomson JM et al. 2005. Resurrecting ancestral alcohol dehydrogenases from yeast. Nat. Genet. 37:630. Available from: http://dx.doi.org/10.1038/ng1553

Tronchoni J et al. 2017. Early transcriptional response to biotic stress in mixed starter fermentations involving *Saccharomyces cerevisiae* and *Torulaspora delbrueckii*. Int. J. Food Microbiol.241:60–68. Available from: http://dx.doi.org/10.1016/j.ijfoodmicro.2016.10.017

Tsai IJ et al. 2008. Population genomics of the wild yeast *Saccharomyces paradoxus*: Quantifying the life cycle. Proc. Natl. Acad. Sci. 105:4957 LP-4962. Available from: https://doi.org/10.1073/pnas.0707314105

Walther A et al. 2014. Genome Sequence of *Saccharomyces carlsbergensis*, the World's First Pure Culture Lager Yeast. G3 Genes|Genomes|Genetics 4. Available from: http://doi.org/10.1534/g3.113.010090

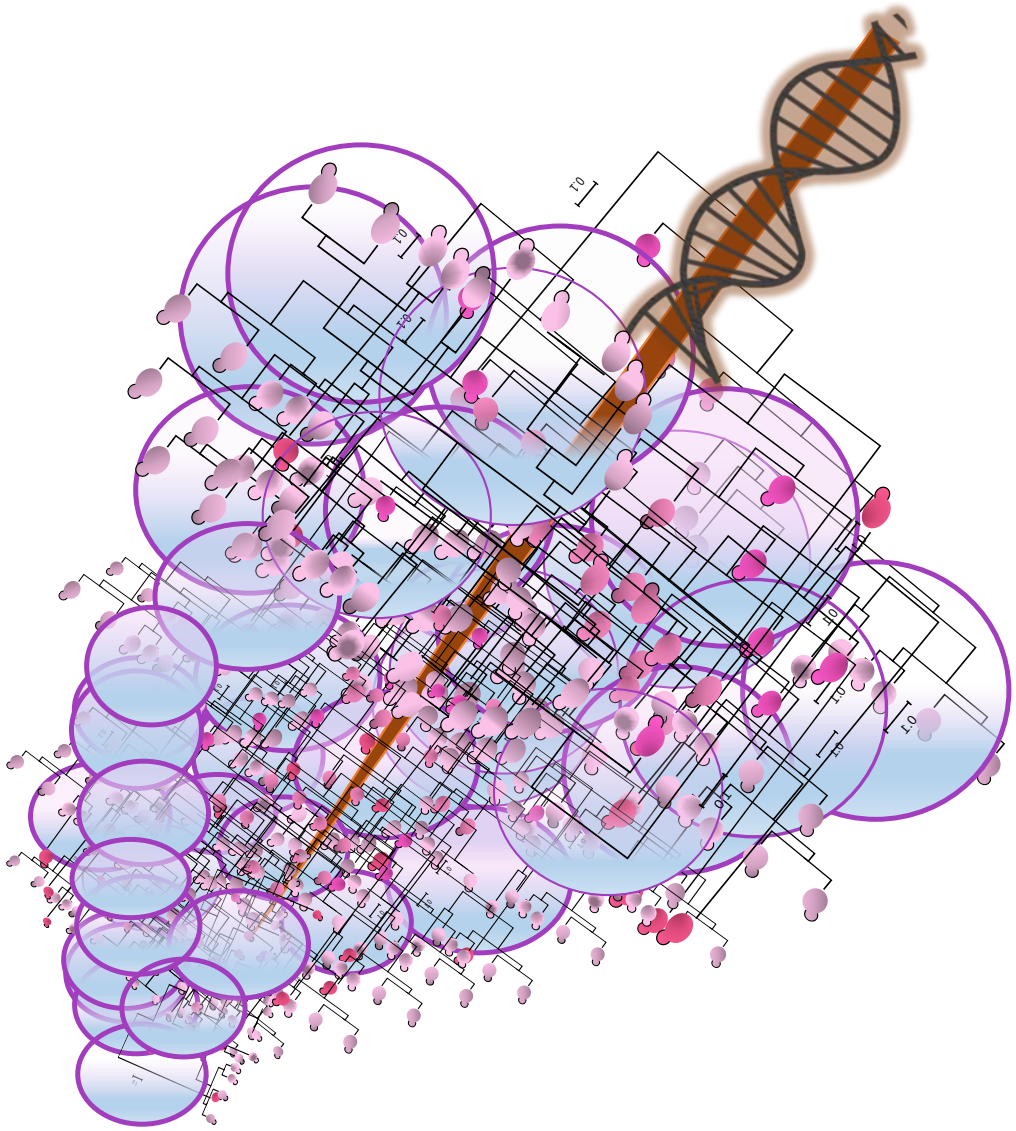
Wang S-A and Bai F-Y. 2008. *Saccharomyces arboricolus* sp. nov., a yeast species from tree bark. Int. J. Syst. Evol. Microbiol. 58:510–514. Available from: http://doi.org/10.1099/ijs.0.65331-0

Wei W et al. 2007. Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. Proc. Natl. Acad. Sci. U. S. A. 104:12825–12830. Available from: http://doi.org/10.1073/pnas.0701291104

Wolfe KH. 2015. Origin of the yeast whole-genome duplication. PLoS Biol. 13: e1002221. Available from: https://doi.org/10.1371/journal.pbio.1002221

Wolfe KH and Shields DC. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. 387. Available from: https://www.nature.com/articles/42711

Yue J-X et al. 2017. Contrasting evolutionary genome dynamics between domesticated and wild yeasts. Nat. Genet. 49:913–924. Available from: https://www.nature.com/doifinder/10.1038/ng.3847



CSIC
Consejo Superior de Investigaciones Científicas

iata
Instituto de Agroquímica
y Tecnología de Alimentos