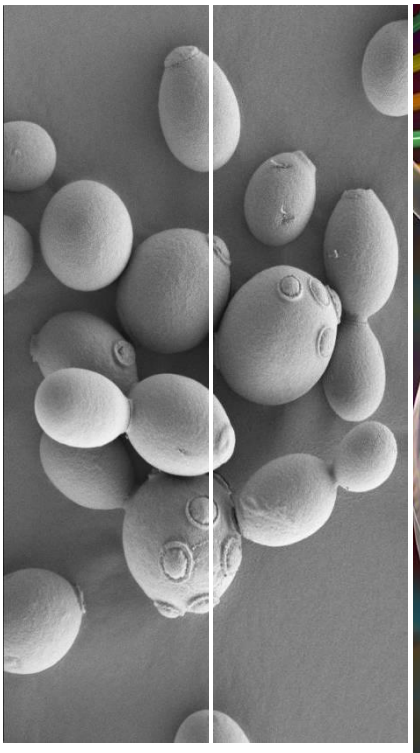


tesis doctoral

Comparative genomics and transcriptomics in *Saccharomyces*

Clara Ibáñez Martínez

CSIC

Supervised by:
Dra. Amparo Querol

Clara Ibáñez Martínez

Comparative genomics and transcriptomics in *Saccharomyces*

Doctoral Thesis

Supervised by:

Dra. Amparo Querol

Dpt. De Bioquímica i Biología Molecular

Doctorat en Biotecnología

Valencia

2015

Instituto de Agroquímica y Tecnología de Alimentos (IATA-CSIC)
Grupo de Biología de Sistemas en Levaduras de Interés Biotecnológico

La Dra. Amparo Querol Simón, Profesora de Investigación del Consejo Superior de Investigaciones Científicas (CSIC) en el Instituto de Agroquímica y Tecnología de Alimentos (IATA),

INFORMA

Que la presente memoria "**Comparative genomics and transcriptomics in *Saccharomyces***" constituye la tesis doctoral de la **Srta. Clara Ibáñez Martínez** para optar al grado de Doctora por la Universidad de Valencia. Asimismo, certifica haber dirigido y supervisado tanto los distintos aspectos del trabajo como su redacción.

Y para que conste a los efectos oportunos, firmamos el presente certificado en Valencia a 11 de Noviembre de 2015

Fdo. Amparo Querol Simón

AGRADECIMIENTOS

Quisiera dedicar la finalización de esta Tesis Doctoral a todas aquellas personas que me han acompañado y facilitado su apoyo, consejo y ánimo a lo largo de este proceso, sin las cuales no hubiera sido posible lograr este objetivo.

En primer lugar, quiero agradecer a la Dra. Amparo Querol, la directora de esta tesis, por haber confiado en mí y haberme dado la oportunidad de trabajar en su grupo de investigación, contando siempre con su consejo y experiencia profesional. Al Dr. Eladio Barrio por el tiempo dedicado al desarrollo del proyecto. También quiero agradecer al Dr. José Manuel Guillamón, por haberme abierto las puertas del mundo de la investigación, por tu apoyo, consejos, y por todos estos años de trabajo compartido y aprendizaje.
Muchas gracias!

A la Dra. Christina Toft por ayudarme con el análisis bioinformático de esta tesis y al Dr. Roberto Pérez-Torrado por su colaboración en algunas de las tareas realizadas.

Gracias al 307 y a todos los demás compañeros de la tercera planta del IATA que en algún momento habéis compartido conmigo quejas y lamentaciones, alegrías y satisfacciones, comidas, cenas… y miles de momentos más que no caben aquí. Habéis sido un gran apoyo día tras día y, aunque algunos estéis lejos, tengo la suerte de poder sentirme siempre acompañada. Ha sido genial haber podido compartir con vosotros todo este tiempo, o gran parte de él. Agradecer a David Lázaro y Laura por brindarme su amistad, apoyo, ánimo y colaboración en todo momento y sobre todo cuando más necesitaba de ellos, sin poner nunca peros o darme negativas, sino todo lo contrario. A mis distinguidos compañeros, David Peris y Jordi Tronchoni, los cuales ya han emprendido otros caminos y con quienes he compartido las mismas experiencias. Debo agradecerles que a pesar de lo difícil del camino, siempre hayan confiado en mi capacidad para finalizar este trabajo animándome a seguir hasta el final. A mi compañera Amparo Gamero, mil gracias por haberme hecho la vida mucho más fácil durante la fase final de este trabajo, un beso para ti "Amparito". A vosotras, Ali, María y Fani, las chicas del 303, gracias por vuestro entusiasmo y cariño a lo largo de todos estos años. Por tantos momentos compartidos dentro y fuera. Esto no lo hubiera conseguido sin vosotras. Nunca. Habéis sido mi punto de apoyo, mi fuerza, con vuestro entusiasmo y vuestra comprensión he podido superar dificultades que parecían imposibles. Ali, por haber sido siempre mi parte racional y mi gran amiga. Gracias por todo lo que hemos vivido juntas antes y durante este tiempo y por todo lo que nos depara el futuro. Unidas ya para siempre.
En general quisiera agradecer a todas y cada una de las personas que, aunque no aparecen aquí con nombres y apellidos, han estado presentes de alguna forma durante el

desarrollo de este trabajo y han hecho posible que hoy vea la luz. Tanto ellas como yo sabemos que desde lo más profundo de mi corazón les agradezco el haberme brindado todo el apoyo, colaboración, ánimo y sobre todo cariño y amistad.

Asimismo, agradezco a los profesionales del Servicio de Genómica del SCSI de la Universitat de València que han colaborado en este estudio, por ayudarme amablemente con el análisis, especialmente a Vicente Sentandreu por su inestimable ayuda durante todo este período.

A Paula, gracias por dedicarme parte de tu tiempo y ayudarme con los formatos de esta Tesis. A Raúl, por convertirte en alguien especial en este último tiempo, gracias por tu cariño y el apoyo que me has demostrado. Y a todos los amigos que en algún momento han sufrido los efectos de escribir una tesis, por su apoyo y escucha en los momentos de estrés.

Por último, mi familia, a los que va dedicada esta tesis. Quiero agradecer de forma especial el apoyo incondicional de mis padres y mi hermano, que me han dado fuerza a lo largo de estos años para continuar hacia delante. Por apoyarme en todas las decisiones que he tomado a lo largo de la vida, y por enseñarme a luchar por lo que quiero. Sin vosotros no sería quien soy y no estaría donde estoy. Gracias de todo corazón. Os quiero!

A todos, mi agradecimiento

"O valor das coisas nao está no tempo em que elas duram, mas na intensidade com que acontecem. Por isso existem momentos inesquecíveis, coisas inexplicáveis e pessoas incomparáveis"

Fernando Pessoa

A mis padres y a mi hermano

1. INDEX

2. INTRODUCTION

2.1. General characteristics of yeast

2.1.1. Yeasts and ecology

Ascomycete yeasts (phylum Ascomycota: subphylum Saccharomycotina: class Saccharomycetes: order Saccharomycetales) comprise a monophyletic lineage with a single order of about 1500 known species (Kurtzman *et al*., 2011). Pasteur was the first to demonstrate experimentally that fermented beverages result from the action of living yeast transforming sugar into ethanol, while Hansen provided the first insights on the distribution of yeast in their natural habitats, being recognized as the founder of yeasts systematic (Phaff *et al*., 1978).

DNA sequencing has revolutionized yeast taxonomy. About 40 different yeast species have been sequenced so far (Figure I.1) and genomic-level aspects of yeast evolution are gradually being unveiled. The most attention has been focused on the Saccharomycotina (or Hemiascomycetes) (Casaregola *et al*., 2011).

Yeasts are widely dispersed in nature with a wide variety of habitats, and are often isolated from sugar-rich materials. They are commonly found on plant leaves, flowers, and fruits, as well as in soil. Yeasts are also found on the surface of the skin and in the intestinal tracts of warm-blooded animals, where they may live symbiotically or as parasites (Kurtzman & Fell, 1998).

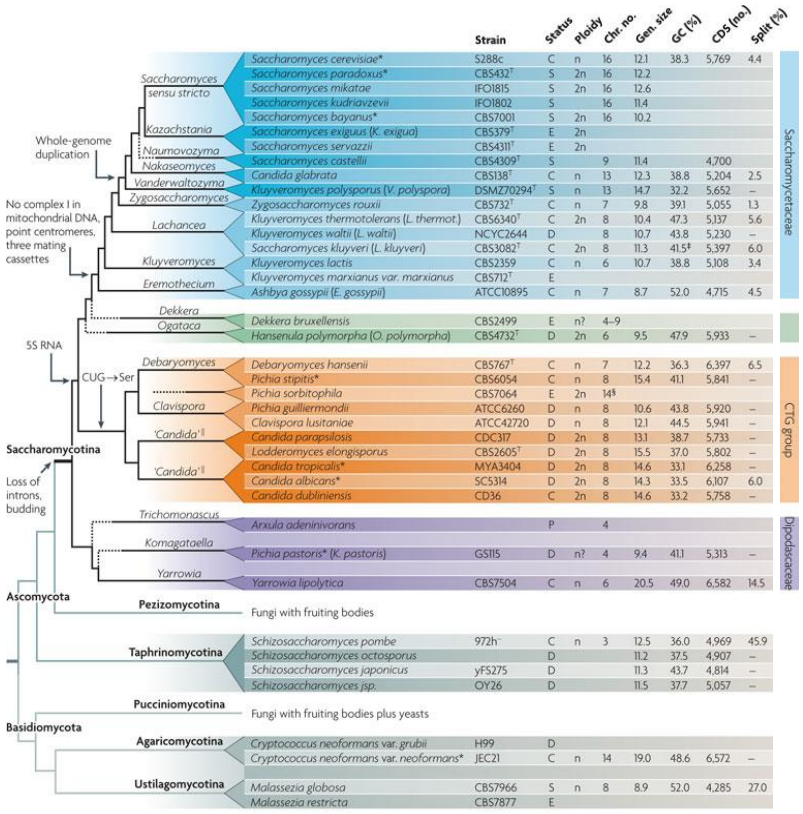


Figure I.1. Tree topology of the sequenced yeast genomes (adapted from Bernard Dujon, 2010).

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

2.1.2. The Saccharomyces genus

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

Saccharomyces mikatae, *Saccharomyces kudriavzevii* (Naumov *et al*., 2000), *Saccharomyces arboricolus* (Naumov *et al*., 2010) and *Saccharomyces eubayanus* (Libkind *et al*., 2011). *S. bayanus* includes two varieties: *uvarum* and *bayanus* (Rainieri *et al*., 2006).
The ecology of *Saccharomyces* species is diverse. Several species of this genus have been only found in natural environments, this is the case of *S. mikatae* (partially decayed leafs), *S. kudriavzevii* (decayed leafs, soils and oaks), *S. arboricolus* (oak trees), *S. cariocanus* (*Drosophila* sp.) and *S. eubayanus* (bark); whereas *S. cerevisiae*, *S. paradoxus* and *S. bayanus* have been found associated to both natural and biotechnological environments.
Saccharomyces species are used to produce a range of fermented beverages, including wine, cider and lager beer. The polyploidy nature, the capability of exchanging genetic material, the high genetic variability and the complexity of evolution in *Saccharomyces* yeasts, make species definition very troublesome. Additional genomic variation can arise from interspecific hybridization, which can occur between two or more *Saccharomyces* species (Barrio *et al*., 2006; Dujon, 2010). Some examples, *S. cerevisiae-S. kudriavzevii* hybrid wine and brewing yeasts (González *et al*., 2008; Peris *et al*., 2011; 2012a y b), *S. cerevisiae-S. uvarum* hybrid cider and brewing yeast (Rainieri *et al*., 2006), and the most well-known hybrid, the lager yeast *S. pastorianus*, which is an interspecific hybrid between *S. cerevisiae* and the recently described *S. eubayanus* (Libkind *et al*., 2011)

2.1.2. Reproduction and cell cycle

Yeast typically grow asexually by vegetative multiplication but can also reproduce sexually by forming ascospores (Figure I.2). The cell cycle in budding or vegetative multiplication consists of four distinct phases (G_1, S, G_2 and M). The sexual reproduction involves the formation of four haploid spores (two MATa and two MATα) and is induced during nutrient starvation (Taxis *et al*., 2005). During conjugation, two cells of opposite mating type (MATa and MATα) fuse to form a diploid zygote (Jackson & Hartwell, 1990). Strains that can be maintained stably for many generations as haploid are termed heterothallic. Strains in which sex reversals, cell fusion and diploid formation occur are termed homothallic. The large majority of *S. cerevisiae* industrial strains are homothallic.

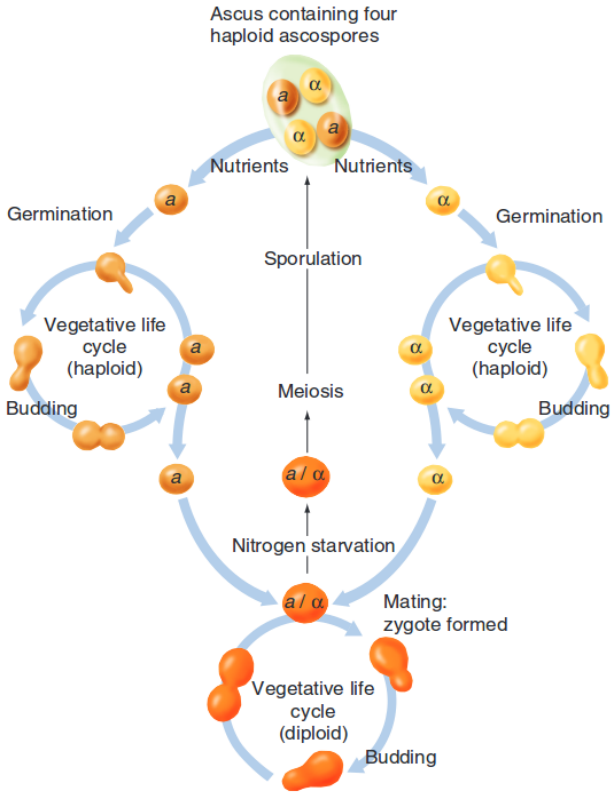


Figure I.2. Life cycle of yeasts. Yeast can grow vegetatively as either haploid or diploid cells. The transition from haploid to diploid occurs via mating, and the transition from diploid to haploid occurs via meiosis during sporulation.

Yeast population growth is the result of cell division and the progression though the cell cycle. Under optimal growth conditions, yeast growth kinetic follows the typical microbial growth curve, comprising three main phases: lag phase, exponential phase and stationary phase (Figure I.3). The **lag phase** reflects the time required for yeast cells to adapt to their new environment by synthesizing ribosomes and enzymes needed to establish growth at a higher rate. The duration of this phase depends on firstly the initial population size and secondly environmental conditions. Once the cell starts actively metabolizing, they begin DNA replication and shortly after the cells divide. This begins the second phase of growth

called the **exponential phase** of growth. This is the period in which the cells reproduce at maximum specific growth rate (μ_{max}). The time it takes the population to double is called generation time. Yeast strain, growth medium, and temperature are important factors in determining the generation time. Industrial fermentations aim to extend this phase for maximizing the output of biomass and metabolites production (López *et al*., 2004). The third phase in yeast growth is **stationary phase**; a period of no growth when metabolism slows and cell division is stopped. The factors that cause cells to enter stationary phase are related to change in the environment, such as nutrient deprivation, toxic metabolites and high temperatures. After prolonged periods in stationary phase, cells may die and autolysate.

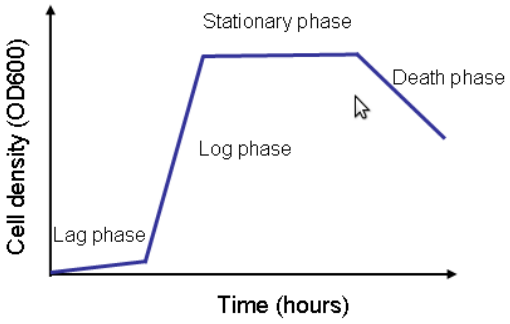


Figure I.3. Standard yeast growth curve

2.1.3. Importance in fermentative processes

Saccharomyces genus possesses a series of unique characters that are not found in other genera (Vaughan-Martini & Martini 1998). One of these unique characteristics is their high capability to ferment sugars vigorously, either in the presence or in the absence of oxygen, to produce ethanol. This ability allows them to colonize sugar-rich substrates (plant saps and fruits) and compete with other yeasts, which are not so tolerant to alcohol. The apparition of angiosperm plants with sugar-rich saps and fruits introduced a new ecological niche with a different selection regime that likely imposed altered physiological demands to the ancestors of *Saccharomyces* yeasts (Wolfe & Shields, 1997). Under such

circumstances, adaptive evolution took place in this new ecological context favouring the acquisition of such high fermentative capability.

This capability has unconsciously been used by humans to produce fermented foods and beverages, which introduced new selective pressures on these yeasts. Neolithic human populations likely observed that fruit juice spontaneously ferment producing an alcoholic beverage (Mortimer *et al*., 1994). Since then, the yeast *S. cerevisiae* and related species become an essential component of many important human activities including baking, brewing, distilling, and wine making.

In general, these industrial *Saccharomyces* strains are highly specialized organisms, which have evolved to utilize the different environments or ecological niches that have been provided by human activity. This process can be described as "domestication" and is responsible of the peculiar genetic characteristics of the industrial yeasts. During the last years, intensive researches have been focused on elucidating the molecular mechanisms involved in yeast adaptation to the industrial process, and the reshaping of genomic characteristics of the industrial yeast which have been unconsciously selected over billions of generations (Querol *et al*., 2003).

Among them, the most useful and widely exploited species are those from the *Saccharomyces* genus, especially *S. cerevisiae*. The ability to this genus to degrade carbohydrates has been unconsciously used by humans for thousands of years to ferment a broad type of beverages (cider, beer, wines, etc.; Querol & Fleet, 2006).

S. cerevisiae has been found associated to very diverse fermentation processes including baking, brewing, distilling, wine making, cider production, etc. and also in different traditional fermented beverages and foods around the world. It is also the principal model eukaryotic organism utilized for fundamental research (Mustacchi *et al*., 2006; Oliver, 2007), and is the yeast best adapted to grow at high temperatures within the *Saccharomyces* genus, with the highest optimum (32.3°C) and maximum (45.4°C) growth temperatures (Salvadó Z. *et al*., 2011). Also *S. cerevisiae* is the species with the highest ethanol resistance (Arroyo-López *et al*., 2010). Besides its traditional role in baking, brewing and wine making, *S. cerevisiae* is an attractive host for the production of heterologous proteins. Moreover, most commercial

fermentations for the production of bioethanol from hexoses regularly employ highly fermenting strains of *S. cerevisiae* (Goldemberg, 2007).

The cryophilic ***S. bayanus***, although has been found in natural habitat in Far East Asia together with strains of *S. cerevisiae* and *S. paradoxus*, also appears associated to different fermentation processes: winemaking (Demuyter *et al*., 2004, Le Jeune *et al*., 2007), cider production (Coton *et al*., 2006, Suárez Valles *et al*., 2007), brewing, and as grape must contaminants. The type strain of this specie, originally isolated from beer, has been described as a hybrid possessing also nuclear genome from *S. cerevisiae* (Nguyen *et al*., 2000, de Barros Lopes *et al*., 2002, Nguyen & Gaillardin 2005), which led to the proposal of the reinstatement of *S. uvarum*, a former taxon included in *S. bayanus*, as a distinct specie (Pulvirenti *et al*., 2000, Nguyen & Gaillardin, 2005) or as a different variety within *S. bayanus* (Naumov, 2000). Recently our group (Perez-Través *et al*., 2014) analysing the genetic variability of the 'uvarum' group showing a high intraspecific homogeneity, although a certain degree of interbreeding among the strains of this variety was shown. The situation of the 'bayanus' group is more complex. Among the *S. bayanus* strains, different levels of homozygosity, hybridization and introgression were found, all these strains are hybrids between *S. uvarum* and *S. eubayanus* and no pure *S. bayanus* var. *bayanus* strain was identified. These *S. bayanus* hybrids can be classified into two types according to the level of homozygous/ heterozygous, indicating that they have been originated by different hybridization processes.

The wild yeast ***S. paradoxus***, the closest relative to *S. cerevisiae*, according to phylogenetic reconstructions (Rokas *et al*., 2003), is a natural specie worldwide distributed with a fortuitous presence in fermentation processes. Strains of *S. paradoxus* have been isolated from natural environments usually associated with tree exudates, the phylloplane or with an unidentified species of *Drosophila* (Glushakova *et al*., 2007; Naumov *et al*., 1997, 1998; Phaff *et al*., 1956). However, also has been described as the predominant yeast in Croatian vineyards (Redzepovic *et al*., 2002).

S. kudriavzevii species has been mainly isolated in natural environments, like decaying leaves (Naumov *et al*., 2000) or oak barks (Sampaio & Gonçalves, 2008; Lopes *et al*.,

2010). Nevertheless, strains of the species have also been isolated in commercial fermentations in New Zealand and in Europe (González, 2006, Lopandic *et al*., 2007). Physiological characterization of *S. kudriavzevii* strains has showed up its cryotolerance, growing quite well at low temperatures (10-15ºC) (Belloch *et al*., 2008; Tronchoni *et al*., 2014). However, *S. kudriavzevii* participates in hybrid formation with *S. cerevisiae* and *S. bayanus* species, which are present in industrial fermentations in central Europe (Masneuf *et al*., 1998; González *et al*., 2006, 2007, 2008; Lopandic *et al*., 2007; Sipiczki, 2008; Dunn *et al*., 2008; Belloch *et al*., 2009; Horinouchi *et al*., 2010; Peris *et al*., 2012 a and b). Physiological data suggest that *Saccharomyces* hybrids might have inherited the ability to grow at high temperatures (30-37ºC) and ethanol tolerance from their *S. cerevisiae* parental and ability to grow at low temperatures (10-16ºC) from their *S. bayanus* and *S. kudriavzevii* parental (González *et al*., 2007; Gangl *et al*., 2009; Gamero *et al*., 2013). These physiological characteristics point out *Saccharomyces* hybrids as better adapted to respond the new winemarkers' trends, such as conducting wine fermentation at low temperatures, which causes wine aroma improvement (Lambrecht & Pretorius, 2000; Torija *et al*., 2003; Llauradó *et al*., 2002, 2005; Novo *et al*., 2003). Oenological characterization of hybrids between *S. cerevisiae* and *S. kudriavzevii* has demonstrated that they are well adapted to ferment at low and intermediate conditions of temperature, giving intermediate or higher amounts of glycerol, less acetic acid and higher amounts of higher alcohols with regard to reference strains of *S. cerevisiae* and *S. kudriavzevii* (Gangl *et al*., 2009; González *et al*., 2007; Lopandic *et al*., 2007). Nevertheless, these hybrids show intermediate ethanol tolerances when compared with their parental strains (Arroyo-López *et al*., 2009; Tronchoni *et al*., 2009; Arroyo-López *et al*., 2010). Concerning oenological characterization of natural hybrids between *S. bayanus* and *S. cerevisiae*, there is limited information in spite of having been described by some authors in wine and cider (Masneuf *et al*., 1998; Nguyen *et al*., 2000). However, artificial *S. cerevisiae* x *S. bayanus* hybrids have been constructed and characterized. These hybrids seem to have inherited the cryotolerance from *S. bayanus* (Kishimoto *et al*., 1994) and they produce intermediate glycerol concentrations with respect to their parental species (Zambonelli *et al*., 1997).

The rest of the species are not associated with fermentative environments. *S. arboricolus* was found associated with the bark of two tree species of the family *Fagaceae* in different regions of China (Wang & Bai 2008), *S. cariocanus* was isolated from a fruit fly (Drosophila sp.) in Brazil (Naumov *et al*., 2000a), *S. mikatae* was isolated from soil and decaying leaves in Japan (Naumov *et al*., 2000a) and *S. eubayanus* was found in in Nothofagus (Southern beech) forests in Patagonia (Libkind *et al*., 2011).

2.1.4. Genetic constitution: lab vs industrial strains

The genome of the laboratory strain *S. cerevisiae* S288c was the first completely sequenced from a eukaryote, and it was released in 1996 (Goffeau *et al*., 1996). The yeast genome is quite small, at only 12 Mb. It is highly packed, with about 6,000 genes distributed over 16 chromosomes. Each chromosome is a single DNA molecule with a length of about 200 to 2200 kilobases. *S. cerevisiae* also has two small cytoplasmic genomes: mitochondrial DNA and killer dsRNA. It is the best well-characterized eukaryotic genome and one of the simplest in terms of identifying open reading frames (ORFs). The sequence defines 5885 potential protein-encoding genes, approximately 140 genes specifying ribosomal RNA, 40 genes for small nuclear RNA molecules, and 275 transfer RNA genes (Goffeau *et al*., 1996). Its primary annotation was updated recently in its first major update since 1996 (Engel *et al*., 2014).

Industrial “domesticated” *S. cerevisiae* strains generally differ from laboratory strains, as well as from “wild” strains, in genetic and physiological properties (Bakalinsky & Snow, 1990; Fay & Benavides, 2005; Mortimer & Polsinelli, 1999; Porro *et al*., 2005; Shuller *et al*., 2007). While laboratory strains of *S. cerevisiae* can be grown stably under haploid or diploid states, industrial strains are usually diploid or aneuploidy and occasionally polyploidy (Pretorius, 2000). Furthermore, industrial yeast strains often sporulate poorly (Bakalinsky & Snow, 1990; Barre *et al*., 1993), and are highly specialized organisms, which have evolved to grow in the different environments or ecological niches that have been provided by human activity. These environments constitute much of the evolutionary framework of the species in the past centuries, and many genes that appear not to be associated with a specific

function in laboratory strains may be responsible for specific phenotypes in industrial strains. These phenotypes could be important in commercial processes, such as rapid and complete sugar fermentation, increased alcohol production and tolerance, formation of desired flavors and aromas, enhanced flocculation, ability to utilize disaccharides and trisaccharides, low foaming propensity, and other traits (Benitez *et al*., 1996; Bisson, 2004; Dequin, 2001). Industrial strains have a large capacity for genome reorganization through chromosome rearrangements (Bidenne *et al*., 1992; Rachidi *et al*., 1999; Puig *et al*., 2000), promoting rapid adaptation to environmental changes. This specialization has been associated with some genome characteristics which cover a wide range of phenotypic traits, such as diploid genome with the presence of aneuploidies or polyploidies, high level of chromosome length polymorphism, homotallism, high heterozygosis, genome renewal and allopolyploid/hybrid genomes (Mortimer *et al*., 1994; Querol *et al*., 2003). It was proposed that the ploidy of the wine yeast may confer advantages in adapting to variable external environments or increasing the dosage of some genes important for fermentation (Bakalansky & Snow, 1990; Salmon, 1997). The genome sequencing of a *S. cerevisiae* wine yeast revealed the presence of horizontal gene transfers that could be involved in adaptation to industrial environment (Novo *et al*., 2009). In addition, wine strains have been characterized by the presence of a set of duplication and depletion genes referred as “commercial wine yeast signature” (Dunn *et al*., 2005; Carreto *et al*., 2008).

Many other *Saccharomyces* species have been sequenced or molecular characterized for studying the genetic basis of phenotypic differences and for elucidating the evolutionary history and the population structure (Fay & Benavides, 2005; Liti *et al*., 2009; Schacherer *et al*., 2009; Warringer *et al*., 2011). In *S. cerevisiae* there are five lineages that exhibit the same phylogenetic relationship across their entire genomes, which are considered to be “pure”. These are strains from Malaysia, West Africa, Sake, North America and Wine/European (Liti *et al*., 2009). In *S. paradoxus* were found three populations depending on the geographic isolation: American (includes *S. cariocanus*), Far Eastern and European (Liti *et al*., 2006, 2009). In the case of *S. kudriavzevii* two different populations have been described: European and Japanese (Sampaio & Gonçalves, 2008; Hittinger *et al*., 2010;

Lopes *et al*., 2010; Peris *et al*., 2012). *S. bayanus* includes two varieties: *uvarum* and *bayanus*. *S. bayanus* var. *bayanus* strains have been shown to be hybrids between *S. cerevisiae* and other yeast close to *S. bayanus* var. *uvarum* (Rainieri *et al*., 2006). A "pure" strain of *S. bayanus* was described as the new specie *S. eubayanus* (Libkind *et al*., 2011). The *S. eubayanus* like-strain genome has been found in the former *S. pastorianus*, an alloploid hybrid of *S. cerevisiae* and *S. eubayanus*, which is found in lager-brewing fermentation (Libkind *et al*., 2011; Bing *et al*., 2014; Peris *et al*., 2014; Gibson & Liti, 2014).This genetic and physiological variability makes interesting the comprehension of the molecular basis of yeasts, and stresses the need for application of genomic approaches together with physiological data, to a wider knowledge of these molecular mechanisms.

2.2. Comparative genomics

2.2.1. *Genomic analysis with array CGH*

Conventional comparative genomic hybridization (CGH) for genome-wide detection of DNA sequences that vary in copy number among individuals has been developed in the early 90s. This approach, in which differentially labelled genomic DNA from a test and reference sample compete for *in situ* hybridization onto normal spreads, has proven useful in assessing chromosomal regions that are repeatedly gained or lost in *Saccharomyces* Gene copy number (GCN) is determined by the fluorescence ratio between corresponding DNA sequences from hybridized test and reference DNA. This technique can also give information on whole or partial chromosome aneuploidies, non-reciprocal translocations and isolated gene deletions or amplifications. Using *S. cerevisiae*-based microarrays, this technique it has also been employed to discover non-reciprocal chromosomal translocations that occurred in yeasts evolved to tolerate low glucose concentrations (Dunham *et al*., 2002). A number of previous papers have demonstrated that aCGH data accurately reflects genome changes. For example, Dunn *et al*., 2005; Carreto *et al*., 2008 using *S. cerevisiae*-based microarrays demonstrated genetic diversity among both commercial and wild *S. cerevisiae* wine yeast strains, and it has been hypothesized that this genetic diversity may, at least in part, be a root cause of their differing fermentative and

sensory qualities. Also, Dunham *et al*. (2002) used array-CGH to study rearrangements (with associated copy number changes) in *S. cerevisiae*. Microarray karyotyping has also been used to evaluate the genome composition of different *S. cerevisiae* × *S. kudriavzevii* natural hybrids isolated from wine and beer fermentations to infer their evolutionary origins and to figure out the potential role of common *S. kudriavzevii* gene fraction present in these hybrids (González *et al*., 2006, 2008; Belloch *et al*., 2008; Peris *et al*., 2012), as well as the genomic architectures (relative to *S. cerevisiae*) of the hybrid organism *S. pastorianus* (Bond, *et al*., 2004) and the *Saccharomyces* species (Edwards-Ingram *et al*., 2004). For yeast microarrays, standard PCR (Lashkari *et al*., 1997; Perez-Ortin *et al*., 2002) or quantitative real-time PCR (Bond *et al*., 2004; Belloch *et al*., 2008) has been used to validate either deletions or amplifications predicted by aCGH data. Also, Dunham *et al*., 2002 and Winzeler *et al*., 2003 have used DNA sequencing to validate rearrangements (with associated copy number changes) or single-nucleotide polymorphisms, respectively, to corroborate their aCGH results.

2.2.2. *Genome sequencing*

Over the past few years, genome sequences have become available from an increasing range of yeast species, which has led to remarkable advances in our understanding of evolutionary mechanisms in eukaryotes. Yeasts offer us an opportunity to examine how molecular and reproductive mechanisms combine to affect genome architectures and drive evolutionary changes over a broad range of species.

As mentioned above, the genomes of many *Saccharomyces* strains have been completely sequenced. S. cerevisiae: S288c, EC1118 (Novo *et al*., 2009), LalvinQA23, Kyokai7; *S. arboricolus* (Liti *et al*., 2013); VIN7 a hybrid *S. cerevisiae* x *S. kudriavzevii* (Borneman *et al*., 2012). Whole genome sequencing provides the most comprehensive collection of an individual's genetic variation. Recent sequencing efforts and experiments have advanced our understanding of genome evolution in yeasts, particularly the *Saccharomyces* yeast as we analysed before. The ancestral genome of the genus *Saccharomyces* has been subject to both whole-genome duplication (WGD) (Wolfe & Shields, 1997; Dietrich *et al*., 2004;

Kellis *et al*., 2004; Langkjær *et al*., 2003; Piskur & Langkjær, 2004), followed by massive sequence loss and divergence, and segmental duplication. WGD has consequences for gene dosage and could affect the protein interaction networks. WGD could explain successive deletions of genes from the initial polyploidy stage creating phenotypically disadvantaged intermediates that could be maintained and evolved under several steps of bottlenecks (Presser *et al*., 2008; Vinogradov & Anatskaya 2009). In *S. cerevisiae*, only ~550 duplicated pairs (ohnologues) have been retained (Byrne & Wolfe, 2005), and similar or lower number of duplicates are observed for other yeasts coming from the same duplication event. The first yeast genome to be assembled primarily from next-generation data was a haploid derivative of a wine yeast strain (AWRI1631) (Borneman *et al*., 2008) and there are another five, high quality commercial wine yeast genome assemblies currently available (Novo *et al*., 2009; Bornerman *et al*., 2011). In addition to wine strain sequencing, there are genome sequences for *S. cerevisiae* strains involved in all of the other major alcohol fermentation industries; beer (Bornerman *et al*., 2011), sake (Akao *et al*., 2011) and bioethanol (Argueso *et al*., 2009; Babrzadeh *et al*., 2012).

Many other *Saccharomyces* species have been also sequenced. *S. kudriavzevii* (IFO1802 and ZP591) (Scannell *et al*., 2011), *S. eubayanus* (Baker *et al*., 2015), and the yeast *S. arboricolus*, that is closely related to *S. cerevisiae*, has been also sequenced (Liti *et al*., 2013). In addition, recently genomewide population analyses have shown the origins of wine yeast domestication (Almeida *et al*., 2015). They have used whole-genome data of numerous *S. cerevisiae* strains and a combination of phylogenomics, population genomics, demographic models and genomic surveys of domestication fingerprints to analyse the relationship of the wine group with a recently discovered oak-associated Mediterranean *S. cerevisiae* population. This new population has been proposed to contain the wild genetic stock that originated the domesticated wine yeasts. It is possible that variation in preferences for fermentative attributes between regions and wine producers, and differences in wines in different regions, have selected distinct genotypes thus enhancing yeast diversity. All these data provide a framework for comparing the genomic attributes of strains between and within industries.

2.3. RNA-sequencing

For over a decade, microarrays were the dominant platform in the high-throughput analysis of gene expression (Marguerat & Bähler, 2010). After the genome sequence of *S. cerevisiae* was reported (Goffeau *et al*., 1996), many studies have been done on the genome-wide expression analysis using DNA microarrays to better understand winemaking processes (Rossignol *et al*., 2003; Varela *et al*., 2005), temperature influence on growth or aroma production (Beltran *et al*., 2006; Pizarro *et al*., 2008), the genes involved in aroma production (Rossouw *et al*., 2008), a general or sugar stress response (Marks *et al*., 2008; Erasmus *et al*., 2003; Ramirez-Córdova *et al*., 2012), the response to nitrogen depletion (Backhus *et al*., 2001), or the physiology of virulent *S. cerevisiae* cells during infection (Llopis *et al*., 2012). Despite several genome-wide expression studies in *S. cerevisiae* using DNA microarray technology, there was no equivalent information available on other species of the genus. To date, some studies using hybridization arrays have been performed to determine the gene expression profile of *S. kudriavzevii* and *S. uvarum* yeasts in our group (Combina *et al*., 2012; Tronchoni *et al*., 2014; Gamero *et al*., 2014). Similarly, gene expression levels of the hybrid yeast species *S. pastorianus* have been seen using microarrays (Horinouchi *et al*., 2010).

Sequencing based methods, such as SAGE (Serial Analysis of Gene Expresion) and MPSS (Massively Parallel Signature Sequencing), used to be the major alternative methods. One of the advantages of these methods is that they provided precise digital gene expression measures instead of analog expression measures provided by microarrays. These methods, however, were based on a conventional Sanger sequencing, and they were not as efficient as later developed methods based on *next-generation sequencing* (Wang *et al*., 2009).

There are several different technologies which correspond to the next-generation sequencing, but all of them have one thing in common - sequencing is done via massive parallelization. Although the three most popular next-generation sequencing technologies are Roche/454, Illumina and ABI SOLiD (Marguerat & Bähler, 2010), other NGS technologies have been and are being developed, and some have already been commercialized, the Ion Torrent System by Life Technologies and the Heliscope by Helicos

BioScience. Their ability to sequence transcriptome cost-effectively and in a high depth gave birth to a new technology for gene expression measurement RNA-sequencing (RNA-seq) (Wang *et al*., 2009). There are several studies that demonstrate the benefits of RNA-Seq over microarrays in transcriptome profiling (Marioni *et al*., 2008; Nookaew *et al*., 2012). Nevertheless, there has been only one RNA-Seq study in industrial yeasts to date, describing the changes in the transcriptomes of three genetically distinct *S. cerevisiae* strains during bread dough fermentation (Aslankoohi *et al*., 2013).

In this section, we review the basic principles of the next-generation sequencing technologies and the steps needed to take in order to conduct an RNA-seq experiment. Also, we give a brief overview of the computational methods involved in RNA-seq data analysis.

2.3.1 Technical principles of RNA-seq technology

The first key step in the next-generation sequencing is sample preparation. The procedure varies from technology to technology, but the basic principles remain the same: coding RNA (mRNA) has to be separated from the rest of the sample, reverse transcribed, fragmented and amplified (Wang *et al*., 2009).

For separation purposes, poly-A tail of the mRNA is often targeted by poly-T oligonucleotides attached to a given substrate. Next, the mRNA is reversed transcribed to cDNA and fragmented into sizes required by the specific protocol. The amplification can be carried out in a few different ways: 454 and SOLiD use *emulsion* PCR while Illumina uses *bridge amplification* (Moorthie *et al*., 2011). The end result for any sample preparation is the same: a number of short single-stranded DNA molecules separated into clusters or microscopic wells on a plate and ready to be sequenced (Moorthie *et al*., 2011).

The next-generation sequencing technology which we used in this study was SOLiD. This system is based on a methodology which is called synthesis by ligation (Moorthie *et al*., 2011; SOLiD. SOLiD Sequence Technology http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/solid-next-generation-sequencing.html, 2012). First, a universal primer and 8-base long fluorescently

labelled oligonucleotide probes are added to the reaction. Of these 8 bases only the first two are meaningful, the rest of them are *degenerate*, meaning that they can pair with any other base. The oligonucleotide probe binds to the DNA strand being sequenced and a *DNA ligase* enzyme links the oligonucleotide to the growing strand. The unlinked oligonucleotides are washed away and the fluorescent label is read by a scanner. Next, three trailing degenerate nucleotides are cleaved and new oligonucleotides are added. This process continues until the new strand is fully synthesized. After this, the whole new strand is denatured and the whole process is repeated with the only difference that a new primer is one nucleotide shorter than the previous one. As a result new bases are read during the process. The whole reaction is repeated five times, to ensure that each nucleotide on the strand is interrogated twice. In this system, only 4 fluorescent colours are used to label 16 types of oligonucleotide probes, but this is sufficient, because the sequence can be later inferred based on a set of logical rules, known as 4-color coding scheme (Moorthie *et al*., 2011).

After the sequencing is completed, the data has to be computationally analyzed. The exact type of analysis depends on what kind of experiment we want to conduct. In addition to gene expression profiling, RNA-seq data can be used for non-coding RNA discovery and detection, transcript rearrangement discovery or single-nucleotide variation profiling (Morozova *et al*., 2009). We, however, will focus only on RNA-seq applications for the gene expression profiling.

2.3.2 Computational analysis of RNA-seq data

In the computational analysis description we will assume that the genome of the species being investigated is known. In that case, the first step is to map the sequencing reads to the reference genome in order to know where they have originated from. This process is not complicated for individual reads, but the problem arises because of the huge amount of reads that need to be mapped. Conventional alignment programs such as BLAST would simply be too slow for this task (Wilhelm & Landry, 2009). Hence, new alignment tools have been developed which are aimed at aligning a large amount of short read data. One of

the most popular tools used for this task is Bowtie (Langmead *et al*., 2009; Langmead, 2010), a program which aligns the short reads in a very fast and memory efficient way.

Another problem or read mapping might be caused because of repetitive regions in a genome. Reads originating from these regions usually cannot be mapped unambiguously. In higher eukaryotic organisms, these regions constitute almost 50% of the genome (Wilhelm & Landry, 2009), so discarding all of those reads would result in a substantial loss of the data. Therefore, many studies use *pair-ended* reads. The idea is that a DNA fragment is sequenced from both of its ends giving rise to two reads with approximately known gap length between them. In this way, aligning one of the paired reads could help to align the other one unambiguously. Nowadays, pair-ended reads are supported by most of the sequencing platforms and alignment tools (Wilhelm & Landry, 2009).

Yet another problem is caused by reads originating from the locations of splice junctions. These reads cannot be straightforwardly mapped to the original genome, because the read sequence is split into two parts and separated by an intron sequence in the original genome. Some of the alignment programs take into consideration the existing transcriptome annotations or even try to find novel splice junctions in order to map such reads. One of the most popular programs among these is TopHat (Trapnell *et al*., 2009). Original Bowtie software was not able to deal with such reads, but the problem is addressed in Bowtie 2 (Langmead & Salzberg, 2012), a new version of the tool.

After the reads are mapped, the gene expression levels can be inferred simply by counting how many mapped reads fall into the regions of known genes (Mortazavi *et al*., 2008). In the fragmentation step, longer genes get more fragments for sequencing; therefore, the counts for each gene have to be normalized by gene lengths. Moreover, these counts have to be normalized by the total number of mapped reads for a sample, because some samples might have more mapped reads than the others which would result in a sample bias. A popular gene expression measure which follows these principles is called RPKM (Reads Per Kilobase of exon model per Million mapped reads) (Mortazavi *et al*., 2008). It normalizes read counts by gene lengths in kilobases and by millions of mapped reads for a sample. When data is originated from paired-end sequencing, our case, the measure is FPKM

(Fragments Per Kilobase of exon per Million fragments mapped), for both gene and isoform quantification, which is analogous to RPKM.

More advanced methods, such as Cufflinks (Trapnell *et al*., 2010) measure gene expression levels not on the gene level, but on the isoform level. Having isoforms level expressions, gene expression can be derived by summing up all of the isoform expressions belonging to a single gene.

The downstreme analysis often includes identification of differentially expressed genes. One of the most popular tools used for this task is Cuffdiff. Differential expression analysis tool calculates $\log_2$ fold change values, p-values and FDR values.

3. BACKGROUND & JUSTIFICATION

This PhD thesis has been carried out within the Department of Biotechnology, group of Systems biology in yeast of biotechnological interest, located at the Institute of Agrochemistry and Food Technology (IATA) from the Spanish Scientific Research Council (CSIC). This research group is focused in different topics of wine microbiology and biotechnology, such as the application of molecular techniques for the identification and characterization of industrial yeasts, selection of starter cultures for their use in industrial fermentations, functional analysis of the yeast under fermentation conditions and genetic improvement of yeasts by classical and recombinant methods. The current PhD work has been focused in the analysis of the genome structures by using a "microarray karyotyping" (also known as "array-CGH" or "aCGH") technique, and the study of gene expression using the RNA-seq technology of diverse *Saccharomyces* species isolated from different fermentative environments.

The yeasts within the *Saccharomyces* genus are responsible for numerous biotechnological processes such as the production of beverages and fermented foods. The adaptation to their environment has been different in diverse species of the genus. The sequence variability and the expression level of genes are responsible for the heterogeneity in metabolic pathways related with the particular and enological characteristics of each specie and strain.

Many technologies have been used over the years for the purpose of measuring gene expression, but the recent availability of next-generation sequencing (NGS) methods has opened up new horizons at the level of gene expression analysis. Thus, the main objectives of this thesis have been the genomic characterization of different *Saccharomyces* species and to explore the transcriptome of *Saccharomyces cerevisiae*, *Saccharomyces kudriavzevii*, *Saccharomyces bayaus* var. *uvarum* and *Saccharomyces paradoxus* strains, isolated from diverse environments with the purpose to detect differences in the expression levels of the homologous genes and specific genes of the new wine species that are not found in *S. cerevisiae*. In the first objective, we performed comparative genomic hybridization analysis (aCGH). In the second objective we used a massively parallel mRNA sequencing platform (RNA-Seq), based on next-generation sequencing technology, to map

and quantify the transcriptome of the *Saccharomyces* yeast at the genome scale under fermentation conditions. Through this doctoral thesis the gene expression of these *Saccharomyces* species has been evaluated in order to better understand the differences of all them and, by this, to help the wine industry to offer new strains for making better wines.

4. MATERIALS & METHODS

4.1. Yeast strains

The yeasts used in the present thesis belong to different species of the genus *Saccharomyces*. A total of twelve yeast strains, seven *S. cerevisiae* (*Sc*), four *S. bayanus* var. *uvarum* (*Sb*), one *S. kudriavzevii* (*Sk*) and one *S. paradoxus* (*Sp*), were used in this work (Table M.1). Some of them were isolated from wine fermentations in different countries (Spain and Croatia) and two strains are currently commercialized as active dry yeasts. The Spanish *S. kudriavzevii* have been isolated from our lab in Ciudad Real and Castellón (Lopes *et al*., 2010).

In all the experiments done in this PhD work, the strain *Sc*S288c, is used as control strain.

4.2. Culture media

4.2.1. *GPY*

Glucose	20g
Peptone	5g
Yeast extract	5g
*Agar	20g
H_2O (distilled)	1000ml

Autoclave at 120º C for 20 minutes.

* GPY solid medium preparation: Follow the same technique to make solid agar plates (media contains 20 g/l agar). Once the media has cooled to about 70° C, it can be poured directly from the bottle into sterile plastic petri dishes in sterile conditions. Fill petri dishes to about 1/3 capacity.

Table M1. Yeast strains used in the present study and isolation source.

Yeast reference†	Species	Commercial and non-commercial strains	Source
*Sc*T73	*S. cerevisiae*	Lalvin T73††	Wine (Spain)
CECT10131	*S. cerevisiae*	CECT10131	*Centaurea alba* (Spain)
Temohaya-26	*S. cerevisiae*	Temohaya-26	Agave juice (Mexico)
PE35M	*S. cerevisiae*	PE35M	Masato (Peru)
CPE7	*S. cerevisiae*	CPE7	Cachaça (Brazil)
Kyokai nº7	*S. cerevisiae*	Kyokai nº7	Sake (Japan)
GB-FlorC	*S. cerevisiae*	GB-FlorC	Byass wineries (Spain)
BMV58	*S. bayanus var.uvarum*	BMV58††	Wine (Spain)
CECT12600	*S. bayanus var uvarum*	CECT12600	Sweet wine (Spain)
CECT1969	*S. bayanus var.uvarum*	CECT1969	Black currant juice (The Netherlands)
NCAIM789	*S. bayanus var.uvarum*	NCAIM789	Oak (Hungary)
*Sp*54	*S. paradoxus*	54	Wine (Croatia)
*Sk*CR85	*S. kudriavzevii*	CR85	Oak tree (Spain)

†† Strains currently commercialized as active dry yeasts by Lallemand.

4.2.2. *Synthetic complete (SC) medium*

Difco Yeast Nitrogen Base	6,7%
Glucose	2%
H2O (distilled)	100 ml
* Difco Yeast Nitrogen Base (w/o amino acids & ammonium sulphate)	1,7%
Proline/Arginine	0,1%
Glucose	2%
H2O (distilled)	100ml

Sterilized by filtration (0.2 µm)

* Minimal media with different nitrogen sources

4.2.3. *Synthetic must media*

Synthetic must media reproduces a standard natural must composition. This media is very useful to make lab micro-fermentations in a reproducible manner. Adapted from Bely *et al*. (1990). Micro-fermentations were done at controlled temperature (22ºC).

Media composition for 1L consist in:

Sugars:	
Glucose	100 g
Fructose	100g
Organic acids:	
Malic	5g
Citric	0,5g
Tartaric	3g
Minerals:	
KH_2PO_4	0,75g
K_2SO_4	0,5g

$MgSO_4 7H_2O$	0,25g
$CaCl_2 2H_2O$	0,155g
NaCl	0,2g
NH_4Cl	0,46g

Weight the different substances and add distilled water (up to 1 L).

Autoclave at 121º C for 15 minutes.

Add the previous prepared stocking solutions of:

amino acids	13,09 mL
trace elements	1 mL
vitamins	10 mL

pH = 3,3

Filter the whole volume using an antimicrobial filter.

Amino acids stocking solution (1 L)

Tyrosine (Tyr)	1,5g
Tryptophan (Trp)	13,4g
Isoleucine (Ile)	2,5g
Aspartic Acid (Asp)	3,4g
Glutamic Acid (Glu)	9,2g
Arginine (Arg)	28,3g
Leucine (Leu)	3,7g
Threonine (Thr)	5,8g
Glycine (Gly)	1,4g
Glutamine (Gln)	38,4g
Alanine (Ala)	11,2g
Valine (Val)	3,4g
Methionine (Met)	2,4g
Phenylalanine (Phe)	2,9g

Serine (Ser)	6g
Histidine (His)	2,6g
Lysine (Lys)	1,3g
Cysteine (Cys)	1,5g
Proline (Pro)	46,1g
Keep at -20 ºC	
Vitamins stocking solution (1 L)	
Myo-inositol	2g
Calcium pantothenate	15g
Thiamine hydrochloride	0,025g
Nicotinic acid	0,2g
Pyridoxine	0,025g
*Biotin3 mL	
*(stocking biotin solution 100 mg/l)	
Keep at -20 ºC	
Trace elements stocking solution (1 L)	
$MnSO_4$, H_20	4g
Zn SO_4, $7H_2O$	4g
$CuSO_4$, $5H_2O$	1g
KI	1g
$CoCl_2$, $6H_2O$	0,4g
H_3BO_3	1g
$(NH_4)6Mo_7O_{24}$	1g
Keep at -20 ºC	

4.3. Growth kinetics in nitrogen sources

The basal growth media selected for the experiments was SC. To study the effect of having proline or arginine as a nitrogen sources SC-pro or SC-arg media were prepared. Yeast growth was carried out in microtiter plates on a reader model POLARstar Optima

(BGM Labtech, Offenburg, Germany). The wells were filled with the appropriate inoculum and 0.50 ml of YNB medium (with or without proline), reaching an initial OD of approximately 0.2 (corresponding to a starting cell number of ~106 cells/ml). Uninoculated wells for each experimental series were also included in the microplate to determine, and consequently subtract, the noise signal. Growth was monitored by optical density (OD) changes at a wavelength of 600 nm. Measurements were done every 30 min for 72 h at 28 °C (until yeast cells reached the stationary phase), after a pre-shaking of 20s. All experiments were carried out in triplicate. A total of 42 growth curves (2 mediums × 7 yeast strains × 3 replicates) were obtained and analyzed. The μmax from each growth curve was determined as described in Arroyo-Lopez *et al*. (2009).

4.4. Analytical techniques

4.4.1. *Glucose and fructose determination*

Glucose and fructose concentration were determined by HPLC (High Performance Liquid Chromatography) consisting of a GP40 gradient pump, an ED40 pulsed electrochemical detector and an AS3500 autosampler system (Dionex Corporation, Sunnyvale, CA, USA). The mobile phase consisted of water and sodium hydroxide 1M (52:48, V/V) at a flow rate of 0.4 ml/min. The anion-exchange CarboPac MA1 column (Dionex, 4 x 250nm) with guard (4 x 148 50nm) was used for chromatographic separation. in duplicate. Fermentations for the experiments were considered finished when sugars concentration was 100 g/l (50% sugars consumption at early-exponential growth phase).

4.4.2. *Mannoproteins extraction & quantification*

Total mannoproteins released during fermentation in synthetic must were quantified at the middle of the fermentative process. The relative mannoprotein content of the yeast cell wall was also determined. Yeast cells were collected at early-exponential growth face during fermentation at 22ºC, from three independently cultured replicates. 2ml of each culture were used to calculate the dry cell weight per litter, in order to standardize each measure with the cell growth.

We used the method for mannoprotein quantification described by Quirós *et al*. (2012). 3ml of supernatant were gel filtered through 30 x 10 mm Econo-Pac® 10 DG disposable chromatography columns (Bio-Rad Laboratories, Hercules, CA) and eluted with 4 ml distilled water in order to isolate the non-retained macromolecular fraction. Then, 3ml of the eluted fraction were filtered again using the same type of columns and eluted with 4ml of distilled water. Two aliquots of 2ml were concentrated in 2ml screw-capped microtubes (QSP, USA) using a Concentrator Plus (Eppendorf, Germany) at 60ºC until complete evaporation. In order to obtain an indicative value of the mannoprotein content in the yeast cell wall, 2ml of each fermentation were centrifuged and cells washed with 1 mL sterile distilled water. Resulting pellets were carefully resuspended in 100 µL of 1 M H_2SO_4. Tubes were tightly capped and placed in a bath at 100 ºC for 5.5 hours to undergo acid hydrolysis. After this treatment, tubes were briefly spun down, 10 fold diluted using 900 µL of miliQ water, filtered through 0.22 µm pore size Nylon filters (Micron Analitica, Spain) and subjected to HPLC analysis for quantification of the glucose and mannose released during hydrolysis. Analytical HPLC was carried out in a Surveyor Plus Chromatograph (Thermo Fisher Scientific, Waltham, MA) equipped with a refraction index detector. A total volume of 25 µL was injected into a HyperREZ XP Carbohydrate H+8 µm column (Thermo Fisher Scientific) assembled to its correspondent guard. The mobile phase used was 1.5 mM H_2SO_4 with a flux of 0.6mL/min and a column temperature of 50ºC. For the preparation of a standard curve, serial aqueous dilutions of commercial mannan from *S. cerevisiae* (Sigma-Aldrich: Fluka) containing 10 different concentrations, ranging from 400 to 1 mg/L, were prepared and subjected to the double hydrolysis described above.

4.4.3. *Gas chromatograph: volatile aroma compound analysis*

Higher alcohols and esters were analyzed based on a headspace solid phase microextraction (SPME) technique using a 100 µm poly-dimetylsiloxane (PDMS) fiber (Supelco, Sigma-Aldrich, Spain). Aliquots of 1.5 mL of the sample were placed into 15mL vials and 0.35 g of NaCl and 20 µL of 2-heptanone (0.005%) as an internal standard were added. Vials were closed with screwed caps and 13 mm silicone septa. Solutions were

stirred for 2 h to obtain the required headspace-liquid equilibrium. Fibers were injected through the vial septum and exposed to the headspace for 7 min and then were desorbed for 4 min in a gas chromatograph (TRACE GC Ultra, Thermo Scientific) with a flame ionization detector (FID), equipped with an HP INNOWax 30 m x 0.25 mm capillary column coated with a 0.25-m layer of corss-linked polyethylene glycol (Agilent Technologies). The carrier gas helium (1mL/min) and the oven temperature program utilized was: 5 min at 35ºC, 2ºC/min to 150ºC, 20ºC/min to 250ºC and 2 min at 250ºC. The injector and detector temperatures were maintained at 220ºC and 300ºC, respectively. A chromatographic signal was registered by the ChromQuest program. Volatile compounds were identified by comparing the retention time for reference compounds. Volatile compound concentrations were determined using the calibration graphs of the corresponding standard volatile compounds. 2-heptanone (0.005% w/v) was used as an internal standard.

4.5. Molecular techniques

4.5.1. *Pulsed-field electrophoretic karyotyping*

Chromosomal profiles were determined by the contour-clamped homogeneous electric field electrophoresis technique with a CHEF-DRIII equipment (Bio-Rad Laboratories). Chromosomal DNA was prepared in agarose plugs as described by Carle and Olson (1985). *S. cerevisiae* chromosomes (Bio-Rad Laboratories, Hercules, CA) were used as standard markers. Yeast chromosomes were separated on 1% agarose gels in two steps as follows: a 60-s pulse time for 14 h and then a 120-s pulse time for 10 h, both at 6 V/cm with an angle of 120º. The running buffer used was 0.5X TBE (45 mM Trisborate, 1mM EDTA) cooled at 14 ºC. Once the electrophoretic run was stopped, gels were stained with ethidium bromide, visualised under UV light and photographed with an Image Capture System.

4.5.2. *Mating type*

The mating type was determined by PCR procedure described by Huxley *et al*. (1990). These authors described a rapid and unambiguous approach to determine mating

type, which utilizes three oligonucleotides: one generic primer and two specific primers. MAT primer (5'-AGTCACATCAAGATCGTTTATGG) corresponds to an external sequence at the right of MAT locus, Matα primer (5'-GCACGGAATATGGGACTACTTCG) corresponds to a sequence within the α-specific DNA located at MATα and MATa primer (5'-ACTCCACTTCAAGTAAGAGTTTG) corresponds to a sequence within the a-specific DNA located at MATa. When these three oligonucleotides are used in a single PCR, DNA at MATα generates a 404 bp product, whereas DNA at MATa generates a 544 bp product. Finally, diploid colonies yield both products.

Sporulation was induced by incubating cells on acetate medium (1% CH3COONa, 1% Glucose, 0.25% yeast extract and 1.5% agar) for 7-10 days at 28 ºC. Following preliminary digestion of the ascus walls with glucuronidase (Roche) adjusted to 5 mg/ml, spores were dissected using a Singer MSM Manual micromanipulator in GPY agar plates and incubated at 28 ºC during 3-5 days to determine spore viability. A representative number of developed yeast colonies (F1 segregants) were selected and subjected to direct PCR amplification of the MAT locus, as described above, to determine the homo/heterothallism of the parental strains.

4.5.3. *Flow cytometry*

For flow cytometry analysis, cells were grown in GPY at 28ºC for 48 h. Approximately 1x106 cells were recovered by centrifugation (4000 x g during 5 min at room temperature) and washed with 1mL of 1x PBS buffer (8 g/L NaCl, 0.2 g/L KCl, 1.44 g/L Na2HPO4, 0.24 g/L KH2PO4, pH 7.4). To prevent the formation of cellular aggregates, 2 µl of Tween 80 were added. Cells were centrifuged in the same conditions and fixed with 1 mL of cold 70% ethanol. Samples were incubated at -20 ºC for at least 30 minutes. After centrifugation, cells were washed with 1 ml of PBS buffer, centrifuged again and resuspended in 200 µl of PBS containing 0.5 mg/mL RNAse A. Samples were incubated overnight at 37 ºC. After centrifugation in the same conditions, cells were suspended in 500 µL of PBS buffer and sonicated to disrupt aggregates (8-10 seconds). For cell staining, 5 µL of 50 µg/mL propidium iodine was added to each sample, which were then incubated at

37ºC during 15-20 minutes in darkness. Yeast cell DNA content was determined using a FACScan cytometer (Becton Dickinson Immunocytometry Systems, Palo Alto, California, United States). Fluorescence intensity was recovered at a 1039V in FL3 channel. DNA content values were scored on the basis of the fluorescence intensity compared with the S. cerevisiae haploid (S288c) and diploid (FY1679) reference strains. DNA content value reported for each strain is the result of two independent measures. Results were tested by one way ANOVA and Tukey HSD test ($\alpha = 0.05$, $n = 2$).

4.5.4. *Array Comparative Genomic Hybridization (aCGH)*

The experiments were carried out following the methodology described by Peris *et al*. (2012). All experiments were performed using duplicate arrays, and Cy5-dCTP and Cy3-dCTP dye-swap assays were performed to reduce dye-specific bias. Array slides were scanned with an Axon GenePix 4100A scanner (Axon Instruments), and the images analysed using the program GenePix Pro 6.0 software (Molecular Devices Corp., Union City, CA, USA). Using Acuity 4.0 software (Molecular Devices Corp., Union City, CA, USA), manually flagged bad spots were eliminated and the local background was subtracted before averaging the replicate features on the array. Log2 intensity ratios (M values) were then Median normalized to correct for differences in genomic DNA labelling efficiency between samples. The relative hybridization signal of each ORF was derived from the average of the two dye-swap hybridizations performed for each strain. Raw and normalized microarray data are available in GEO (Barrett et al., 2009) under Accession number GSE46165. The normalized log2 ratio (M value) was considered as a measure of the relative abundance of each ORF relatively to that of the reference strain S288c. Deviations from the 1:1 R/G ratio were taken as indicative of changes in gene copy number (GCN). If there are more copies of the gene in the experimental strain relative to the reference strain, the R/G ratio will thus be higher than 1:1 (2:1 for a duplication, 3:1 for a triplication, etc.). Likewise, if there has been a deletion of the gene in the experimental strain relative to the reference strain, or if there are more copies of the gene in the reference strain than in the experimental strain (i.e., the experimental strain has a “depletion” of the gene in terms of

copy number), the R/G ratio will be less than 1:1. Thus, "higher" or "lower" gene copy number refers to increased or decreased number of copies of a gene in comparison with the reference strain S288c. Data imported from Acuity was manipulated and clustered, using established algorithms implemented in the software program Genesis. Average linkage clustering with centered correlation was used to generate visual representations of clusters.

4.5.5. *qRT-PCR*

To validate our microarray results, qPCRs were performed using a LightCycler® 2.0 System (Roche Applied Science, Germany). Oligonucleotide primers (Table M.2) were designed using the Primer-Blast (NCBI) website, according to the available genome sequences of the laboratory strain *S. cerevisiae* S288c. Specificity, efficiency, and accuracy of the primers were tested and optimized by standard PCRs, using DNA from the different strains. Primers showing amplification were used in the subsequent quantitative real-time PCR (qRT-PCR) analysis. The amplification of gene fragments from different yeast strains was determined by qRT-PCR using a standard curve method (Wilhelm et al., 2003). DNA from overnight stationed precultures was extracted in triplicate as described in Querol *et al*. (1992a). qRT-PCR was performed with gene-specific primers (50µM) in a 10µl reaction mixture, using the LightCycler FastStart DNA MasterPLUS SYBR green (Roche Applied Science, Germany). All samples were processed for melting curve analysis, amplification efficiency, and DNA concentration. A mix of all samples and serial dilutions (10^{-1} to 10^{-5}) were used as standard curve. The copy number for each gene was estimated by comparing the DNA concentration of S288c (haploid *S. cerevisiae* strain) with the different strains.

Table M2. Primers used for quantitative PCR in Chapter 1.

Gene	Forward 5' > 3'	Reverse 5' > 3'
HAP3	GCGCGACTCATGAAGAAT	TCTTTCTTTTGTCAGCAGC
SNF4	CGTTGCTTTAAATTGCAGG	GAATGACGTCTATGACCG
VMA5	TATTGTCAAGCCCGAAGAC	ATGCTGGTACAACGTTCTT
PUT1	TGGTTTCCACTATTGACGC	CCCTTAGGTACAACTGCCA

4.5.6. *RNA-sequencing*

4.5.6.1. *RNA preparation and sequencing*

Samples for RNA extractions were taken in the early-exponential phase (50% sugars consumption). Samples were taken in three biological replicates. The withdrawn sample was immediately cooled on ice and the pellet was harvested by centrifugation at 4ºC, washed with DEPC-treated water and the biomass stored at -80ºC until further treatment. The total RNA was extracted from cells through mechanical disruption with glass beads, digested with DNAse and purified using the conventional phenol-chloroform method. PolyA$^+$ RNA was isolated from total RNA using Poly(A) purist kit (Ambion). The quality of the RNA was assayed using a BioAnalyzer (Agilent Technologies, Palo Alto, CA, USA). In total, 250 ng of the total RNA was used to synthesize cDNA using Affymetrix 30 IVT Express kit and successively cRNA was synthesized (Affymetrix Inc., Santa Clara, CA, USA). The same high quality RNAs were submitted for constructing the library that was used for sequencing. RNA sequencing was performed on the SOLID v4 from Applied Biosystems (Life Technologies, Inc) following standard procedures.

4.5.6.2. *RNA-seq analysis*

Due to paired-end sequencing, two RNAseq datasets resulted from each RNA sample, a dataset of reads in the 5'-3' direction and a dataset in the 3'-5' direction, for a total of six data sets. After signal processing and trimming, and conversion from the SOLiD encoding color space into double-encoded colors in the FASTQ format, available reads were aligned to the *S. cerevisiae* S288c reference genome using TopHat, version 2.0.3 (Trapnell *et al*., 2009), which utilized Bowtie aligner, version 0.12.8 (Langmead *et al*., 2009) and SAMtools, version 0.1.18 (Li *et al*., 2009). Default alignment parameters were used, but with the following options: mate inner distance of 200 (for paired end runs); Color space reads. SAMStat, v 1.08 (Lassmann *et al*., 2011), was used to extract mapping quality information from the SAM files generated by Tophat. The mapping results were viewed with the Integrated Genome Viewer (IGV) (Thorvaldsdottir *et al*., 2012) software, version 2.0.30. The read alignments from Tophat were processed by Cufflinks, v1.3.0 to generate the

transcripts (Trapnell *et al*., 2010). Cufflinks measures transcript abundances in Fragments Per Kilobase of exon per Million mapped reads (FPKM), which is analogous to single-read "RPKM" (Mortazavi *et al*., 2008). An FDR<0.05 was considered to have significant expression abundance. We only had technical replicates, so we used Cuffdiff to get the differential expression. Cuffdiff was run with default parameters and provided the saccharomyces_cerevisiae.gff file for reference annotations. CummeRbund, v2.2.0, an R package (Goff & Trapnell, 2011), was used for viewing and managing Cufflinks data.

4.5.6.3. *De novo assembly for unmapped reads*

To find sequences not present in the lab strain S288c and sequences from other *Saccharomyces* species, we considered unmapped reads and carried out *de novo* assembly with the Velvet software, v1.2.08 (Zerbino & Birney, 2008). We performed Basic Local Alignment Search Tool (BLASTn) of contigs longer than a 100bp against the *Saccharomyces* translated nucleotide database.

5. RESULTS & DISCUSSION

CHAPTER 1

Comparative genome analysis of *Saccharomyces cerevisiae* yeast isolated from traditional fermentations unveils different strategies.

Results published:

Ibáñez C[1], Pérez-Torrado R[1], Chiva R[1], Guillamón JM[1], Barrio E[1,2], Querol A[1]. Comparative genome analysis of *Saccharomyces cerevisiae* yeast isolated from traditional fermentations unveils different adaptive strategies. Int J Food Microbiol. 2014 Feb 3;171:129-35. http://dx.doi.org/10.1016/j.ijfoodmicro.2013.10.023

[1]Departamento de Biotecnología de los alimentos, Instituto de Agroquímica y Tecnología de los Alimentos (CSIC), Avda. Agustín Escardino,7, E-46980-Paterna, Valencia, Spain

[2]Departamento de Génetica, Universidad de Valencia, Spain

1. Introduction

All over the world, traditional fermented foods and beverages have been produced by humans since the beginning of the first Neolithic civilizations. These foods and beverages are produced from different sources of sugars, in different climates and with different levels of technological complexity. However, in all of them, *Saccharomyces cerevisiae* is the main responsible of the fermentation processes. *S. cerevisiae* yeast strains isolated from different origins show different physiological properties, indicating their ability to adapt to different environmental conditions. Different molecular methods also revealed a high genetic variability among these *S. cerevisiae* yeasts, which has been correlated to their geographic origin and sources of isolation (Aa *et al*., 2006; Fay & Benavides, 2005; Legras *et al*., 2007; Liti *et al*., 2009).

The characteristics of the diverse traditional fermentations are very different attending to their sugar composition, temperature, pH or nitrogen sources. During the adaptation of yeasts to these new environments provided by human activity, their different properties imposed selective pressures that shaped the *S. cerevisiae* genome. Molecular studies have shown that adaptive mutations include not just the generation of new alleles by nucleotide substitution (Cubillos *et al*., 2011; Salinas *et al*., 2012), but also gene duplication (Barrio *et al*., 2006) and other genome rearrangements (Pérez-Ortín *et al*., 2002b).

DNA-array based Comparative Genome hybridization (aCGH) technique has proven as very useful to characterize yeasts strains at the genomic level. This technology has most commonly been used to detect chromosomal amplifications and deletions in cancer cells (Pollack *et al*., 1999). It has also been used to explore the genomic diversity of different *S. cerevisiae* strains (Carreto *et al*., 2008; Dunn *et al*., 2005; Hauser *et al*., 2001; Infante *et al*., 2003; Pérez-Ortín *et al*., 2002a; Winzeler *et al*., 2003) as well as the genomic structure of the lager yeast strains (*S. pastorianus*) and the species of the genus *Saccharomyces* (Dunn *et al*., 2012; Edwards-Ingram *et al*., 2004). These studies revealed that, in addition to variability in the copy numbers of transposable elements, *S. cerevisiae* strains show gene copy number variability in dispersed (*CUP1, ASP3, ENA1/2/5* and *HXT6/7* loci) and subtelomeric (such as *SNO/SNZ*) gene families.

In the present work we perform a comparative genomic hybridization analysis to explore the genome constitution of seven *S. cerevisiae* strains isolated from different traditional fermentations. Our final goal is to determine the adaptive evolution of *Saccharomyces* yeasts properties of biotechnological interest.

Many gene copy number (GCN) changes were observed, especially in genes associated to subtelomeric regions and transposon elements. Among the fermentation strains, differential GCN was mainly observed in genes related to sugar transport and metabolism. An outstanding example of GCN changes is the gene *PUT1*, involved in proline assimilation, which correlated with the adaptation of the strains to the presence of this nitrogen source in the media.

2. Results

2.1. *Strain diversity and genome variability*

A total of 7 *S. cerevisiae* strains isolated from different sources and the reference laboratory strain S288c were analysed by aCGH in this study (Table 1.1). These yeast strains were isolated from very different environments, including different traditional fermentation processes. We selected a traditional wine strain, T73, that has been used as a classic wine yeast model (Gómez-Pastor *et al*., 2010; Querol *et al*., 1992b); sake strain Kyokai nº 7; strain PE35M, isolated from Masato fermentation a traditional drink from the Peruvian Amazonia obtained from the fermentation of cassava (*Manihot esculenta* Crantz) root juice,; strain CPE7 isolated from sugarcane juice fermentation used for cachaça distillation in Brazil; strain Temohaya-26 isolated from maguey plant (*Agave* sp.) juice fermentation used for Mezcal production; strain GB-FlorC isolated from the biofilm (flor) developed on the surface of fortified sherry wines during their biological aging; and a strain, CECT10131, isolated from flowers in the wild.

Before the aCGH analysis, we performed sporulation and spore viability analyses for the strains, as well as their homo/heterothallism (Table 1.1). Strains CECT10131, PE35M, T73 and Temohaya-26 produced abundant tetraspored asci after 7-10 days of incubation on acetate agar plates, and strains CPE7, Kyokai nº7 and GB-FlorC did not sporulate at all. For

those strains capable to sporulate, ascospores were manually dissected using a micromanipulator and deposited on GPY agar plates to measure spore viability. They exhibited a high variability in the spore viability which ranged from null viability (0%) in the CECT 10131 to a high viability (84%) of the wine strain T73. There are several possible causes that can explain absence of sporulation and spore inviability, the most important are haploidy, autopolyploidy, alloploidy (interspecies hybridization), the presence of aneuploidies, and heterozygosity for translocations. Alloploidy was discarded because these strains were previously characterized by multilocus sequencing and they corresponded to pure *S. cerevisiae* strains (Arias, 2008). Therefore, we decided to evaluate the ploidy of all the isolates by flow cytometry. Moreover, ploidy estimates also are very important to interpret aCGH data because hybridization signals are normalized with respect to those of the reference haploid strain S288c.

The ploidy estimates (Table 1.1) indicate that the four strains capable to sporulate (CECT10131, PE35M, Temohaya-26 and T73) were diploids or almost diploids (from 2,00 + 0,05 to 2,15 + 0,10), two strains (CPE7 and Kyokai nº7) were very close to the haploidy (~1.5x), and the flor yeast GB-FlorC appeared as more than triploid (~3.5x).

Table 1.1. Genetic analysis of the yeast strains examined.

Strain designation	Source/origin	Genotype	Spore viability (%)	Segregants genotype	Thallism*	Ploidy [a]
CECT10131	*Centaurea alba;* CECT, Spain	MAT a/α	0			2,00 ± 0,05[c]
PE35M	Masato; Greater San Marcos University, Lima, Peru	MAT a/α	20.45	MAT a/α	HOM	2.15 ± 0.10[c d]
CPE7	Cachaça; Federal University of Minas Gerais. Brazil	MAT a	0		HET	1.59 ± 0.1[b]
Temohaya-26	Agave juice; Technological Institute of Durango. Mexico	MAT a/α	80.77	MAT a/α	HOM	2.00 ± 0.12[c]
Kyokai nº 7	Sake; Japan	MAT a	0		HET	1.43 ± 0.3[a b]
Lalvin T73	Wine; Alicante. Spain	MAT a/α	83.93	MAT a/α	HOM	2.08 ± 0.06[c d]
GB-FlorC	Jerez wine; González-Byass wineries. Jerez. Spain	MAT a/α	0			3.28 ± 0.04[e]

*HOM: homothallic. HET: heterothallic.
[a]Values expressed as mean ± standard deviation. Values not shearing the same superscript letter are significantly different (ANOVA and Tukey HSD test. α=0.05. n=2).
Culture collection abbreviations: CECT (Spanish Type Culture Collection. University of Valencia).

The possible presence of aneuploidies or chromosomal rearrangements was qualitatively determined by pulsed-field gel electrophoresis (PFGE). This technique allowed us to genotype the different strains based on their chromosomal pattern (Figure 1.1). No clear differences that could be associated to gross chromosomal rearrangements (translocations) were observed among the different strains. However, some chromosome number variations, that could explain the aneuploid values obtained by flow cytometry, as well as chromosomal length variations were evident. In general, *S. cerevisiae* exhibits a rich chromosomal length polymorphism, observed predominantly in the smaller chromosomes. This is mostly due to variations in telomeric repeats, to changes in copy numbers of the ribosomal genes located in chromosome XII, and, in a lesser extent, to changes in the copy numbers of genes in gene families mainly located in subtelomeric regions (Barrio *et al*., 2006). Accordingly, we decided to perform comparative genome hybridization assays, a powerful technique that

permits to infer genomic differences due to GCN changes that cannot be clearly identified with the CHEF technique.

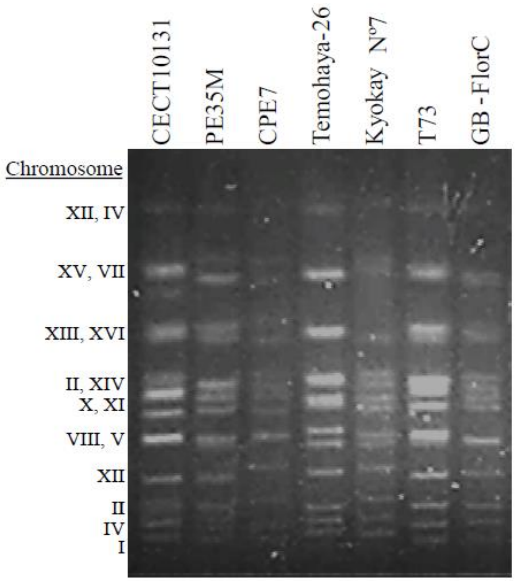


Figure 1.1. Chromosome profiles of the *Saccharomyces cerevisiae* strains used in this study obtained after chromosome DNA electrophoresis on agarose gels by using the CHEF method.

2.2. *Shared genomic differences with the laboratory strain S288c*

For aCGH analysis, genomic DNA from each strain was fluorescently labelled and hybridized with genomic DNA from the reference strain S288c, with duplicate experiments in reverse Cy-dye labelling (dye-swap) design (see Methods). Significant data are presented in Figure 1.2 where we observed a genomic distribution of the gene copy number (GCN) changes. Among the total of 632 genes exhibiting GCN changes at least in one strain (Table S1.1), we observed that most of them correspond to subtelomeric regions (25.9 %) and to transposons (3.7 %). When we determined the molecular function ontology of the ORFs, most of these genes are described as encoding hypothetical proteins (20.4 %), homologous or similar to another protein (25.2 %), of unknown function (6.6 %), putative protein (6.0%)

or questionable ORFs (7.1 %). Also, with respect to their biological process ontology, a significant number of genes have been classified as unknown biological process (45.5 %). GO term analysis confirmed the high frequency of GCN changes in genes related to transposition process (transposition, GO:0032196, p-value<0.0038) and also to sugar transport and fermentation functions (carbohydrate transmembrane transporter activity, GO:0015144, p-value<0.0051; aryl-alcohol dehydrogenase (NAD+) activity, GO:0018456, p-value<0.0030). When we focused on genes with altered copy number respect to control lab strain S288c we observed three groups of depleted genes in all the experimental strains, two located in subtelomeric regions (*YFL050-68*, and *YOL162-164*) and one (*ASP3*) near the ribosomal cluster in chromosome XII. In addition, *HXT12*, *HXT15*, *HXT16*, *MPH2*, *MPH3*, *CUP1-1* and *CUP1-2* are depleted in most of the strains. Thus, the common genomic differences among all the experimental strains in comparison to the reference strain correspond to genes, involved in sugar and amino acid metabolism pathways, mainly transporters. To validate these microarrays results, the copy numbers of four representative genes were determined by quantitative real-time PCR (qRT-PCR) in some of the experimental strains (Table S1.2). Generally, the copy number determination by both methods showed similar values (Tables S1.1 and S1.2), confirming the robustness of the data obtained in the aCGH analysis.

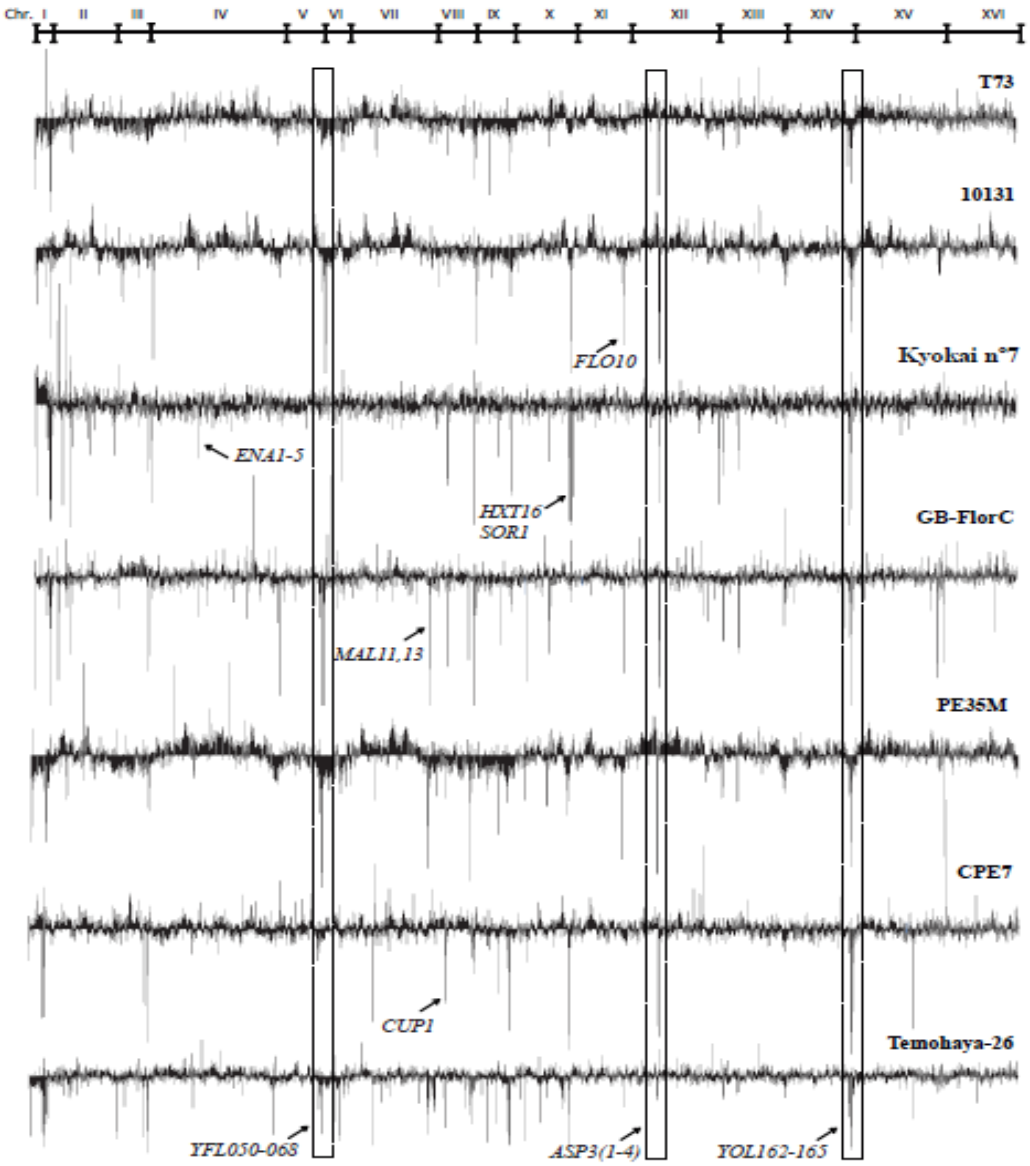


Figure 1.2. Genomic profiles of the *S. cerevisiae* strains obtained after CGH assay comparing gene copy number (GCN) variation in every chromosome with S288c strain. Vertical boxes highlight gene clusters with altered GCN in most of the strains analysed. Specific GCN are indicated with arrows.

We performed a hierarchical clustering to observe the similarities among the different strain analysed in this study (Figure 1.3A). Interestingly, the wine yeasts T73, CECT10131 and GB-FlorC, did not form a specific cluster. In fact, T73 clusters with PE35M, strain isolated in Masato (traditional fermentation), a different environment compared to vineyards. We also compared the aCGH results from our strains with those from other clinical and wine strains analysed by Carreto *et al*. (2008). This comparison (Figure 1.3B) showed that our strains formed two clusters (T73, CECT10131, PE35M, Temohaya-26 in one cluster and GB-FlorC, CPE7, Temohaya-26 in the other one) separated from the clusters including clinical strains and other wine strains. Interestingly, wine strain T73 did not cluster with other wine strains from Carreto *et al*. (2008). All these results suggest that the global genomic fingerprint is not clearly correlated with the origin of the strains. To go into depth in specific genomic differences among the fermentative yeast strains, we studied variability of specific group of genes related to fermentative metabolism.

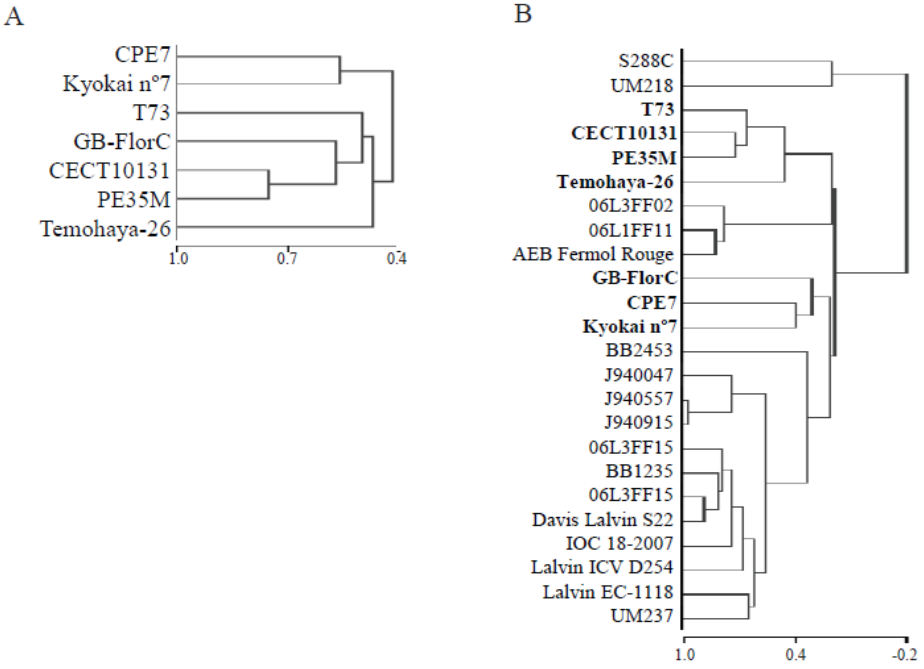


Figure 1.3. Hierarchical clustering of CGH profiles. A) The strains used were grouped according to their CGH profiles. For this, a hierarchical clustering analysis using Genesis program was performed. These clusters identify strains that shared similar ORF copy number changes. B) Comparison with Carreto *et al*. (2008) CGH dataset.

2.3. *Genomic differences between fermentative yeast strains*

The flocculin gene family "Flo" responsible for flocculation is of great importance in the wine industry since it is necessary to form the velum in biological aged wines (Zara *et al*., 2005). In the GB-FlorC strain there is no decreased copy numbers for Flo genes as observed in T73 wine strain for *FLO9* or in strain CECT10131 for *FLO10*. On the other hand *FLO1* showed decreased GCN in all the strains. Other "Flo" related genes such as *YAL065C*, *YAR061W*, *YAR062W*, *YFL051C* and *YHR213W* also showed variability in GCN among the studied strains. Other interesting genes are those encoding maltose transporters because they are important for maltose assimilation and MAL13 is non-functional in the laboratory S288c strain (Charron *et al*., 1986). Surprisingly, the genes *MAL13*, *MAL11* and *YGR290W* (partially overlaps MAL11/YGR289C, a high-affinity maltose transporter) are depleted in GB-FlorC, PE35M and Temohaya-26 suggesting that these strains are not capable of fermenting maltose. Also *MAL31* exhibited lower GCN in CECT10131, CPE7, GB-FlorC and Kyokai nº7.

Genes such as *MPH2*, which codes for a sugar transporter, and *SDL1*, an L-serine ammonia-lyase, were depleted in the genome of CECT10131. Also ORFs *YER187W* and *YKR104W* (an oxidoreductase), were depleted in several strains. The absence of *SOR1, SOR2, AGP3* and *AYT1* genes was observed in traditional fermentative strains (CPE7, GB-FlorC, Kyokai nº7, PE35M and Temohaya-26) but it is present in the wine strain T73. We thus believe that these genes are a part of the "wine strain signature". Also a decreased GCN of *CAR2*, involved in arginine catabolism, was observed in the CPE7 strain.

2.4. *Differential adaptation to different amino acids availability*

A high variability in GCN among strains was observed for *PUT1*, which encodes a proline oxidase, the first step of the pathway for the assimilation of this amino acid. This way, *PUT1* showed increased GCN in T73, PE35M and CECT10131 whereas no change was observed in Temohaya-26, CPE7, GB-FlorC or Kyokai nº7. As explained above, copy numbers of this gene were confirmed by qRT-PCR in most of these strains (Table S1.2). These results prompted us to investigate the ability of this yeast to grow in media with proline as nitrogen source. We evaluated the maximum growth rate in standard minimal media and in a minimal media with proline as a unique nitrogen source. The results (Figure 1.4A) showed a fine correlation with the GCN changes of *PUT1*. The two strains with the highest ability to grow using proline as a nitrogen source were PE35M and T73, two strains with increased *PUT1* GCN. Strains with similar *PUT1* GCN to the reference strain showed intermediate levels of relative growth in proline. The strain with the worst relative growth rate was Kyokai nº7. The only exception was the strain CECT10131 since it showed increased GCN for *PUT1* but intermediate levels of relative growth in proline. We also observed high variability in GCN among strains for *CAR2*, which encodes a L-ornithine transaminase, the second step of arginine degradation pathway. *CAR2* showed decreased GCN in CPE7 and increased GCN in GB-FlorC whereas no changes were observed in T73, CECT10131, Temohaya-26, PE35M or Kyokai nº7. We determined the maximum growth rate in standard minimal media and in a minimal media with arginine as a unique nitrogen source. The results (Figure 1.4B) fitted with the GCN changes of *CAR2*. The strain with the highest ability to grow using arginine as a nitrogen source was Temohaya-26 and the strain with the lowest relative growth rate was CPE7.

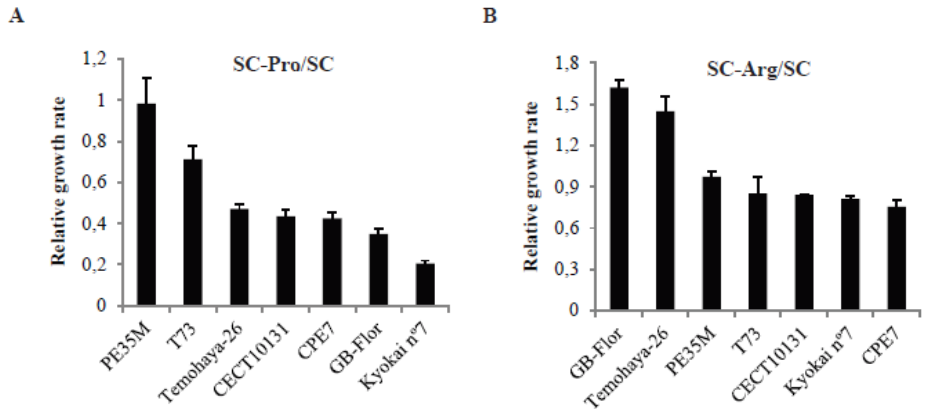


Figure 1.4. Capability of *S. cerevisiae* strains isolated from different traditional fermentations to grow with different nitrogen sources. A) Capability to grow with proline. Strains were inoculated in minimal media with YNB (SC) or proline as nitrogen source (SC-pro) and grow was followed in microtiter plates. Maximum growth rate in proline was calculated and normalized (μSC-pro/μSC). B) Capability to grow with arginine. Strains were inoculated in minimal media with YNB (SC) or arginine as nitrogen source (SC-pro) and grow was followed in microtiter plates. Maximum growth rate in proline was calculated and normalized (μSC-arg/μSC). Error bars represent the standard deviation of the mean reduction of at least 3 independent experiments.

3. Discussion

In this work we have studied strains isolated in different traditional fermentations from different regions of the world. Molecular and genetic approaches were performed to elucidate their characteristics and search for common features that define this group of yeasts.

We have found a high degree of variability in the GCN among the strains consistent with the adaptation to different environments. Our aCGH data showed that rearrangements in subtelomeric regions and Ty elements insertion (duplication and deletion) are the main common genomic diversity features as observed in previous studies (Carreto *et al*., 2008;

Dunn *et al*., 2012). Most of the genes showing GCN changes corresponded to subtelomeric (*MAL*, *FLO*, *HXT* and *SOR*), and intrachromosomal (*ASP*, *ENA* and *CUP*) gene families involved in metabolic functions related to cellular homeostasis, cell-to-cell interactions, or transport of solutes such as ions, sugars and metals. It is well known that subtelomeres are highly unstable showing high levels of strain variability (Winzeler *et al*., 2003). Many of these genes are not essential but they can play an important role in the adaptation to new environmental conditions (Barrio *et al*., 2006). This way, recent studies on quantitative trait mapping in *S. cerevisiae* yeasts showed that a third of the QTLs detected mapped to subtelomeric regions, which support their importance in adaptive evolution (Cubillos *et al*., 2011).

Saccharomyces yeasts have recently re-emerged as a prime model organism for genetics and evolutionary biology in general (Hittinger, 2013; Scannell *et al*., 2011) and for the analysis of complex quantitative traits in particular (Liti & Louis, 2012). Since the first aCGH studies, researchers have been looking for specific genomic patterns of the strains belonging to the different *S. cerevisiae* types. This interesting idea would provide the locus or group of genes that matters to adapt to a specific environment. But this idea has been only demonstrated for a specific type of strains, the bioethanol producing strains (Stambuk *et al*., 2009). Some effort was done to find "commercial wine yeast signatures" (Dunn *et al*., 2005), and *AGP3* and *DAK2* among other genes were suggested as the most important genes present in wine yeast respect to laboratory strains (Dunn *et al*., 2012). However, our data shows that these are genes with a high degree of variability in their copy numbers. *AGP3* was present in sake strain Kyokai nº7 but showed lower GCN in sherry wine strain GB-FlorC and *DAK2* showed decreased GCN in all strains but CPE7 and GB-FlorC. Other studies pointed out that wine strains have specific higher number of copies for *IMD* and *PHO* genes and lower for *MAL* genes (Carreto *et al*., 2008). However, we observed lower GCN of *MAL* genes in all strains, not only wine, and *IMD* or *PHO* genes did not show important GCN changes. Although there are some groups of genes that seem to be altered in fermentative strains, our data suggest that GCN patterns are very variable and is difficult to find a specific feature for yeasts of the same type. This conclusion is not surprising since

gene deletion/duplication is one of the most common forms to produce genetic variability, being five times more frequent than point mutations (Lynch *et al*., 2008).

An interesting depletion, observed in strains GB-FlorC and PE35M, is the *AYT1* gene, which encodes an acetyltransferase that was first characterized in *Fusarium* and subsequently identified in *S. cerevisiae* by homology. It plays a role in detoxifying endotoxins of the tricothecene family in *Fusarium*, but its presence in a non-trichothecene producer organism, which needs no self-protection against endogenic trichothecenes is really striking.

Although it has been shown that the *S. cerevisiae AYT1* gene product can acetylate tricothecene in vivo, cells that are deficient for this gene show no clear phenotypic effect (Dunn *et al*., 2005).

Nitrogen is the main growth-limiting factor in wine fermentations, since its deprivation produces a nutritional stress on metabolic activities (Pretorius, 2000). Depletion of the *AGP3* gene which play an essential role in amino acid transport suggest that this nitrogen source is not important in natural niches from where the traditional fermentative strains were isolated. However, we have observed a strong adaptation to nitrogen sources, illustrated by the GCN changes observed for *PUT1*, the main player in the assimilation of proline as a nitrogen source. Proline is the major amino acid in grape must. However, this amino acid is hardly used by wine yeast during the anaerobic conditions exerted during wine fermentation (Salmon & Barre, 1998). In spite of this contradiction between abundance of this amino acid in grape-must and poor adaptation of wine yeast to utilize proline, the T73 strain, adapted to grape must fermentation, showed a *PUT1* GCN increase. *PUT1* GCN also increased in PE35M, a strain adapted to cassava (Manihot esculenta) juice fermentation, and both strains showed a direct correlation between their higher GCN and their ability to grow with proline as a nitrogen source. Interestingly, the proline is also an important nitrogen source in cassava (Bradbury & Holloway, 1988). On the contrary, the strain Kyokai nº7, showing the worst capacity to grow with proline as a nitrogen source, is adapted to sake mash fermentation where proline is absent (Lemura *et al*., 1999). Another interesting piece of evidence is the lower GCN of *CAR2* gene, involved in arginine catabolism, in the CPE7 strain. This strain was isolated from sugarcane juice fermentation, in which arginine is a rare

amino acid (Wiggins & Williams, 1955). All these GCN data are nice examples of yeast genomic variation involved in the adaptation to the diverse nitrogen sources found in the different environments, suggesting that nitrogen metabolism has played an important role in the adaptive evolution of *S. cerevisiae*.

4. Conclusions

We have isolated and genotyped several yeast strains from natural environments and carried out an aCGH analysis. Genomic variability was identified between the strains, in particular in subtelomeric regions and intrachromosomal gene families involved in metabolic functions related to cellular homeostasis, cell to cell interactions, and transport of solutes such as ions, sugars and metals. However, the most interesting result is the association observed between Gene Copy Number changes in genes involved in the nitrogen metabolism and the availability of nitrogen sources in the different traditional fermentation processes. This suggested that this type of genome variability is the main source of genetic diversity and nitrogen metabolism has played an important role in the adaptive evolution of *S. cerevisiae* strains.

CHAPTER 2

Identification of genomic copy number variants in *Saccharomyces uvarum* using array Comparative Genomic Hybridization.

Results partially published:

Ibáñez C[1], Pérez-Torrado R[1], Guillamón JM[1], Querol A[1]. Identification of genomic copy number variants in *Saccharomyces uvarum* using array Comparative Genomic Hybridization. Biotechnol Prog.

[1]Departamento de Biotecnología de los alimentos, Instituto de Agroquímica y Tecnología de los Alimentos (CSIC), Avda. Agustín Escardino,7, E-46980-Paterna, Valencia, Spain

[2]Departamento de Génetica, Universidad de Valencia, Spain

1. Introduction

Despite the fact that *Saccharomyces cerevisiae* is the predominant species responsible for alcoholic fermentation, other species of the genus *Saccharomyces* seem to have an important role during fermentation processes (Blondin *et al*., 2009). A total of 7 natural species are considered nowadays in the *Saccharomyces* genus: *S. arboricolus, S. cerevisiae, S. eubayanus, S. kudriavzevii, S. mikatae, S. paradoxus* and *S. uvarum* (Almeida *et al*., 2014; Boynton & Greig, 2014). Additionally, two well-studied hybrid groups including the lager yeast *S. pastorianus* (*S. cerevisiae* x *S. eubayanus*) and *S. bayanus* (*S. uvarum* x *S. eubayanus* with minor contributions from *S. cerevisiae* in some cases) as well as some minor number of hybrids including portions of *S. kudriavzevii* (Lopes *et al*., 2010; Peris *et al*., 2012) also integrate the complexity of *Saccharomyces* genus (Libkind *et al*., 2011). *S. bayanus* has been isolated from wild habitats in different regions of the planet, and fermentation processes. The taxonomic status of this species is controversial because it includes strains with heterogeneous properties, isolated from beer, cider or wine. Some authors proposed to keep all of these strains grouped in the taxon *S. bayanus* as two varieties, var. *bayanus* and var. *uvarum* (Naumov *et al*., 2000), and others proposed to be considered as two distinct species (Rainieri *et al*., 1999, Pulvirenti *et al*., 2000).

S. uvarum has been associated with wild habitats and cider or wine fermented at low temperatures (Serra *et al*., 2005). Despite much less information being available about *S. uvarum* than for *S. cerevisiae*, several studies have shown some interesting enological features of this species. The optimum growth temperature of *S. uvarum* is lower than *S. cerevisiae* which makes the strains of this species good candidates for low-temperature fermentations (Salvadó *et al*., 2011).

S. uvarum strains are used as starters in winemaking for low-acid musts because they synthesize malic acid and succinic acid (Rainieri *et al*., 1998), inhibit malolactic fermentation (Caridi & Corte, 1997) and produce more glycerol than *S. cerevisiae*, and less acetic acid and ethanol (Bertolini *et al*., 1996; Castellari *et al*., 1994; Giudici *et al*., 1995; Sipiczki, 2002). The use of these yeasts makes it possible to decrease the quantity of sulphites added to stabilize wines (Caridi & Corte, 1997), a considerable advantage for sulphite-sensitive

consumers. These strains have also been described to produce a high concentration of volatile fermentative compounds such as ß- phenylethanol and its acetate. Thus wines are more aromatic (Eglinton *et al*., 2000; Naumov *et al*., 2001; Gamero *et al*., 2011).

S. uvarum is less commonly used than *S. cerevisiae* in wine and cider, and only appears as being predominant in European regions with an oceanic or continental climate, where fermentation traditionally takes place at lower temperatures. These strains are present in the spontaneous fermentation of various white variety musts from the Val de Loire area (Masneuf *et al*.,1996), Sauternes, Jurançon or Champagne (Naumov *et al*., 2000, 2001), Alsacia (Demuyter *et al*., 2004), Txakoli (Rementeria *et al*., 2003), Muscat and Amarone (Naumova *et al*., 2010), Valpolicella from Italy (Torriani *et al*., 1999); Tokai from Hungary and Slovakia (Sipiczki et al., 2001, Naumov *et al*., 2002, Antunovics *et al*., 2005), and wines from Ukraine (Naumov & Nikonenko, 1989).

The identification and molecular characterisation of the *S. uvarum* among wine yeasts offer the possibility of using its gene pool in breeding programs. The aim of the present study was to investigate the genotypic similarity and differences between four *S. uvarum* strains isolated from mistela, a spontaneous wine fermentation, tree exudate and currant juice, to confirm their relevance during different fermentative processes.

To this end, the genomic DNA of the four strains was analyzed by DNA microarrays. The accuracy of the microarray analysis was further investigated by performing physiological analysis on a selected set of genes that exhibited different hybridization in the microarray analysis.

2. Results and Discussion

2.1. *Determining variable genes*

Genetic differences between the different yeast strains may account for some of the variation seen in their fermentation properties and may also produce differing sensory characteristics in the final product itself. To search for the specific features of the different *S. uvarum* strains, we studied their genome by microarray CGH. This technique allowed us to

identify the genes with gene copy number (GCN) changes (Table S2.1). All the strains were diploid or came close to diploidy.

2.2. *Shared genomic differences with the laboratory strain S288c and other S. cerevisiae*

The labeled DNA from the four strains was hybridized with DNA from control strain S288c to a standard cDNA microarray containing cDNA of the S288c genome. The obtained signals were filtered and normalized. The whole dataset is deposited in GEO Database (Barrett *et al*., 2009) under Accession number GSE67917. Figure 2.1A shows a graphical representation of the microarray karyotyping data for the *S. uvarum* strains presented as "karyoscope" diagrams. To produce a karyoscope diagram the hybridization ratio for each gene is mapped onto its corresponding position on each chromosome of the reference strain of *S. cerevisiae*. The height of each vertical bar is proportional to the log2 of the red:green (R/G) ratio for a gene; if the ratio is greater than1 (i.e., a positive $\log_2$ value), the bar will be drawn above the chromosome; this bar thus represents an over-representation (amplification) of that gene in the wine strain relative to S288c. For R/G ratios less than 1 (i.e., a negative $\log_2$ value), the bar will be drawn below the chromosome; this bar thus represents an under-representation (deletion or depletion, i.e., lower copy number) of that gene in the *S. uvarum* strain relative to S288c.

When we analyzed the absolute signal intensity of the genes in relation to the distances to telomeres (Figure 2.1B) we observed that GCN variation are distributed along the chromosomes, but more significantly in the first 30kb, which matches with the subtelomeric regions. These observations are consistent with previous CGH studies which compared the genome of *S. cerevisiae* strains of different origins (Carreto *et al*., 2008; Dunn *et al*., 2012; Ibáñez *et al*., 2014).

There were more down regulated genes than up regulated genes in all the strains studied (Figure 2.1C). Seventy-nine genes were induced or repressed only in the strain CECT1969 while the other three strains had much lesser variation in GCN.

Only fifty-three genes with different copy number were common to the four strains (Figure 2.1D). We performed an enrichment analysis of this set of genes to better understand the

underlying biological processes. The genes associated with transposition nitrogen metabolism (*ASP3* genes) and stress responses (*CUP1* genes) were significantly enriched in our dataset. GCN variations (fewer or no copies) of the *CUP1* and *ASP1* genes have been observed previously in wine strains (Perez-Ortin *et al*., 2008). As these sequences are all moderately repetitive and highly conserved, the larger number of copies of these genes in S288c is a private feature of this strain, which is not shared by most of the *S. cerevisiae* and *S. uvarum* wine strains, in which their GCN variation has been previously assessed.

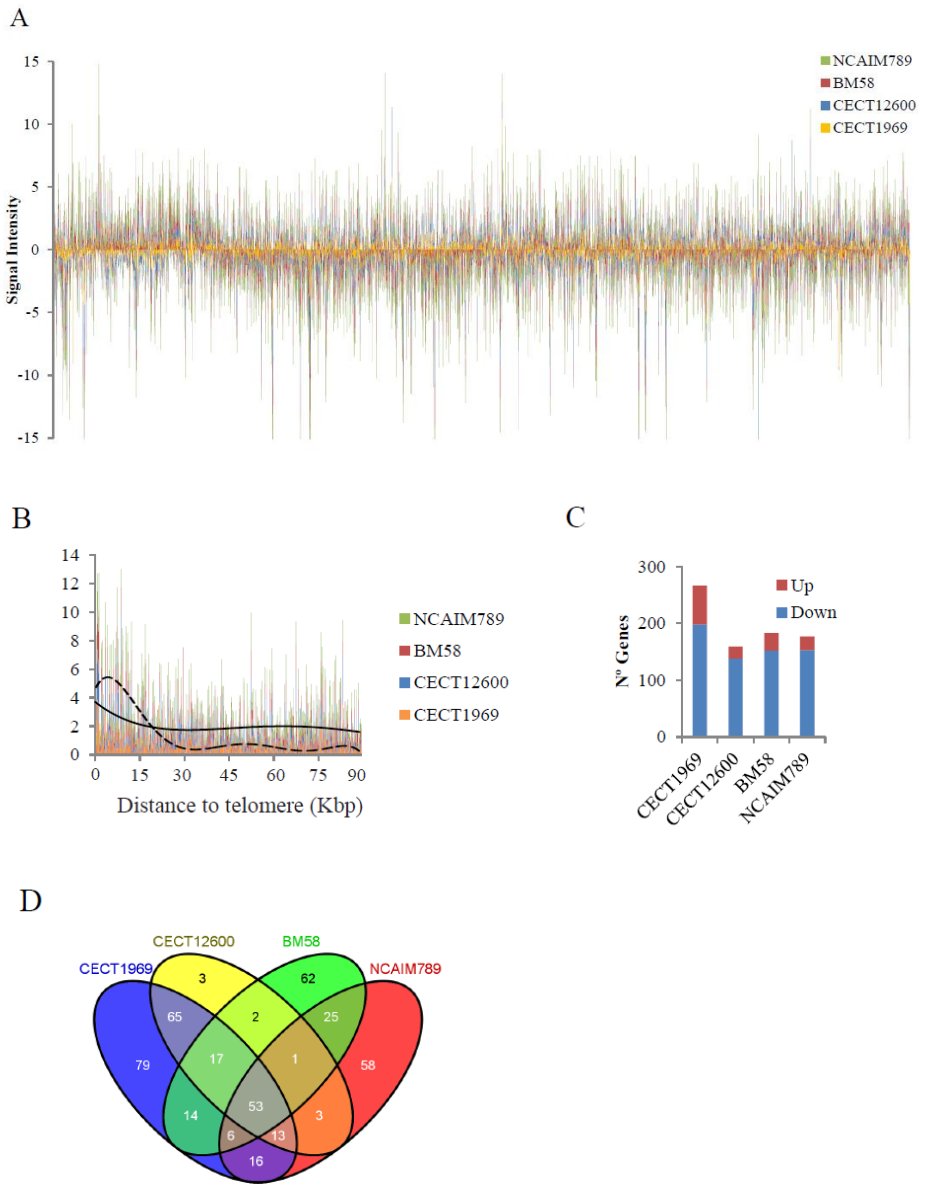


Figure 2.1. Comparative genomic hybrization of *S. uvarum* strains. **A)** Log2 of normalized and filtered data of the four strains is represented in stack bar graph. **B)** Absolute signal intensity of genes is represented against distance to telomere. Discontinuous line represents sliding window average (10x) of values in 1 Kbp. **C)** Number of genes in

increased or decreased CNV represented for each strain. **D)** Venn diagram showing the number of specific or common genes for strains CECT1969, CECT12600 BMV58 and NCAIM789.

To observe similarities among the strains analysed in this study and among those from other wine and clinical *S. cerevisiae* strains analysed by our group (Ibáñez *et al*., 2014) and Carreto *et al*. (2008) we performed a hierarchical clustering comparing all datasets. This analysis (Figure 2.2) showed that the *S. uvarum* strains formed a cluster separated from the *S. cerevisiae* cluster, but the four *S. uvarum* strains were divided into two groups, one that grouped the strains CECT1969 and CECT12600 (wine and currant juice) and another that grouped BMV58 and NCAIM789 (wine and oak). This finding indicated no correlation between the origins of the strains.

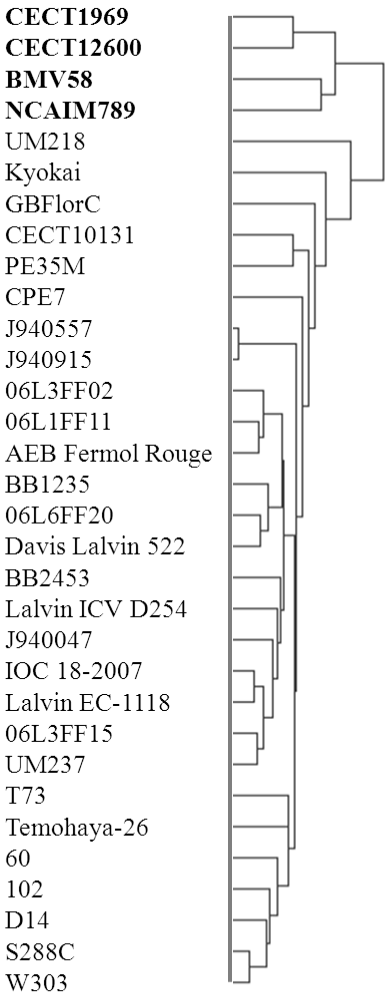


Figure 2.2. Hierarchical clustering of CGH profiles. A) The strains used were grouped according to their CGH profiles. For this, a hierarchical clustering analysis using Genesis program was performed. These clusters identify strains that shared similar ORF copy number changes. B) Comparison with Carreto *et al*. (2008) CGH dataset.

When we focused on genes with an altered copy number in *S. uvarum* to control lab strain S288c, we observed different groups: one formed by *VEL1* and its paralog *YOR387C*

subtelomeric genes involved in zinc metabolism; one constituted by *TDH2*, *TDH3, PDC1*, *PDC5*, *HAP3*, key genes in metabolic pathways for utilization of non-fermentable carbons. *HOT13*, a mitochondrial protein involved in zinc ion binding that influences the redox state, was also depleted in most strains.

Thus, the common genomic differences among all the experimental *S. uvarum* strains compared to the reference *S. cerevisiae* strain correspond to the genes involved in carbon metabolism pathways and ion homeostasis, which are mainly transporters.

2.3. *Differential adaptation to non-fermentable carbon sources*

S. cerevisiae preferentially uses glucose as a carbon source. Yet following its depletion, it can utilize a wide variety of other carbons including non-fermentable compounds such as ethanol, glycerol or acetate. A shift to a non-fermentable carbon source results in massive gene expression reprogramming, including the genes involved in gluconeogenesis, the glyoxylate cycle, and the tricarboxylic acid cycle. This shift also increases the transcription of the genes that encode the mitochondrial enzymes required for respiration. Given the different numbers of copies of the key metabolic genes involved in non-fermentable carbon utilization, we decided to test their ability to grow in different carbon sources.

To evaluate the growth ability of the different strains studied in various carbon sources we measured OD_{600} increases in several non-fermentable compounds (acetate, ethanol and glycerol).

While all the *S. uvarum* strains displayed similar growth patterns with ethanol as a carbon source, except for strain BMV58 which grew better than the other obtained values that were the closest to *S. cerevisiae*. Strain *S. cerevisiae* T73 exhibited an increased ability to grow when ethanol was added to GPY (Figure 2.3), which indicates its best fitness for the assayed ethanol condition. Regarding growth when acetate was added to the medium, strains *S. uvarum* NCAIM789 and *S. cerevisiae* T73 achieved higher values and grew similarly, while the other three *S. uvarum* strains obtained lower values. When glycerol was added to GPY, CECT12600 and T73 grew better compared to all the other strains, where

CECT1969 grew the least. Although strains displayed differences in growth, no growth defect on the non-fermentable carbon sources was observed. These results correlated with the GCN changes observed for the *TDH2*, *TDH3, PDC1*, *PDC5* and *HAP3* genes, involved in the utilization of non-fermentable carbon sources. The extra copies of these genes might help speed up the rate at which a species evolves since 'spare' copies are free to adapt to new roles.

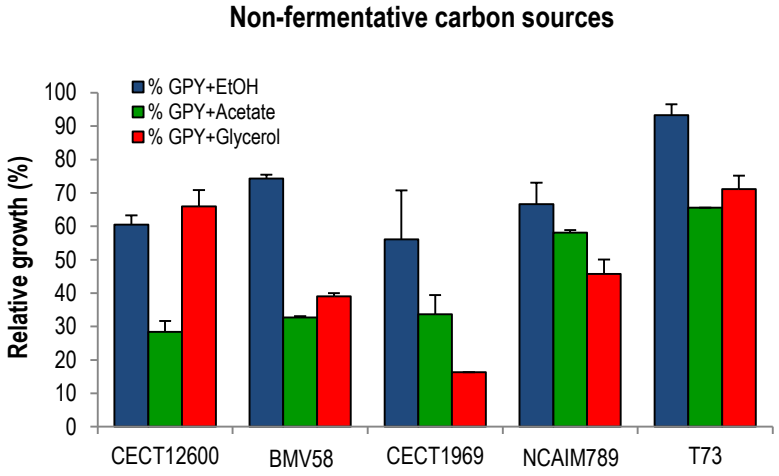


Figure 2.3. Capability of *S. uvarum* strains to grow with different non-fermentable carbon sources (ethanol, acetate and glycerol) compared with *S .cerevisiae* T73 strain. Strains were inoculated in GPY without glucose modified with 1% sterile pure ethanol, 3% glycerol or 1% potassium acetate as carbon sources, and growth was followed in microtiter plates. Error bars represent the standard deviation of the mean reduction of at least 3 independent experiments.

2.4. *Analysis of genes with apparent relation to growth in low zinc conditions*

Zinc is needed for all organisms to grow because it acts as a required cofactor for many different proteins. Zinc limitation might perturb cellular function by reducing the

amount of cofactor available to enzymes, transcription factors, and other proteins that require the metal.

The microarrays analysis revealed that strains CECT12600 and BMV58 exhibited a depletion of *VEL1* and its paralog *YOR387C*, involved in zinc metabolism. Induction of the *YOR387C* gene in response to zinc deficiency suggests that this gene is also involved in the first-line defense against zinc limitation. The function of this protein is not yet known, but its pattern of regulation by zinc suggests that it may play a role in zinc uptake or vacuolar zinc export (Wu Chang-Yi *et al*., 2008).

To evaluate the growth ability of the different strains studied in the presence or absence of Zn, we measured OD600 dynamics under these conditions (Figure 2.4). Zinc limitation conditions (LZM) resulted in diminished growth in all the *S. uvarum* strains in comparison to the non-limiting zinc condition (LZM+Zn). Conversely, both the assayed *S. cerevisiae* (the wine T73 and lab BY4743 strains) grew better under low zinc conditions (LZM), which indicates that *S. uvarum* strains show increased sensitivity to limiting zinc compared to *S. cerevisiae*, which is in agreement with the smaller copy number of the zinc metabolism genes in this species.

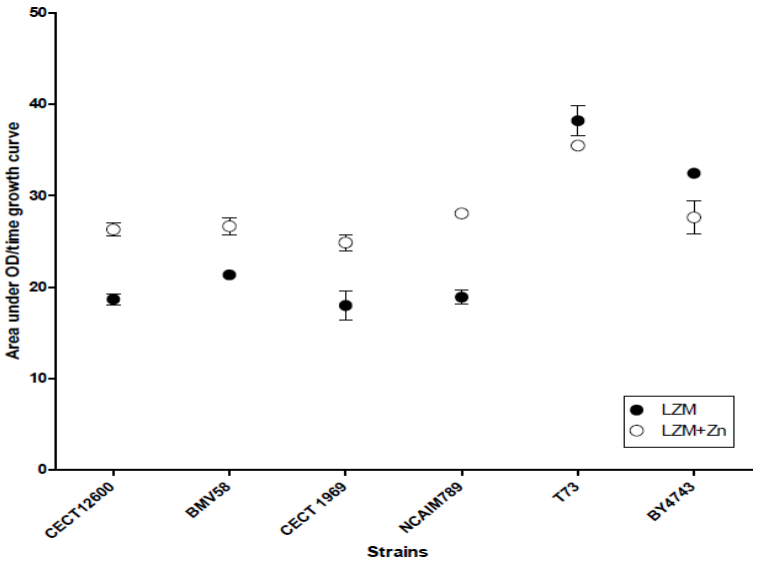


Figure 2.4. Effects of zinc availability on growth. *S. uvarum* and *S. cerevisiae* strains were grown in LZM medium. LZM+Zn containing the indicated zinc concentration. Cultures were incubated and A_{600} was determined.

3. Conclusions

This work characterizes a number of *S. uvarum* yeast strains derived from winemaking and other human activities. The relatively small number of differences seen by microarray karyotyping between the strains suggests that the differences ascribed to these different strains may arise from a small number of genetic changes. Genomic variability has been identified in subtelomeric regions and at Ty-element insertion sites in various studies (Dunn *et al*., 2005; Carreto *et al*., 2008; Peris *et al*., 2012; Ibáñez *et al*., 2014), and suggest that this type of genome variability is the main source of genetic diversity in yeast populations. Genes that display copy number variation are enriched for functions related to interactions with the external environment (sugar transport and metabolism, flocculation, metal transport and metabolism) (Bergström *et al*., 2014). In our study the genes identified with a variable copy number corresponded to intrachromosomal (*ASP* and *CUP*) gene

families involved in metabolic functions related to cellular homeostasis, or to the transport of solutes such as ions, sugars and metals. We also observed other genes (*VEL1*/*YOR387C, TDH2*, *TDH3, PDC1*, *PDC5*, *HAP3, HOT13*) whose copy numbers vary. Despite only minor variations in gene copy numbers between the different yeast strains, *S. uvarum* displayed greater diversity in the presence and absence, and also in the copy number of genetic material. Other studies have corroborated and extended this finding for *S.cerevisiae* and *S. paradoxus* (Bergström *et al*., 2014). Therefore, other evolutionary processes that involve different forms of genetic variation might be responsible for the genetic variants that underlie phenotypic variation in yeast. We also believe that these differences could represent potential selective markers for those strains that may be important to the wine industry.

CHAPTER 3

RNAseq-based transcriptome comparison of *Saccharomyces cerevisiae* strains isolated from diverse fermentative environments.

Ibáñez C[1], Pérez-Torrado R[1], Toft C[1,2], Barrio E[1,2], Querol A[1].

[1]Departamento de Biotecnología de los alimentos, Instituto de Agroquímica y Tecnología de los Alimentos (CSIC), Avda. Agustín Escardino,7, E-46980-Paterna, Valencia, Spain

[2]Departamento de Génetica, Universidad de Valencia, Spain

1. Introduction

Saccharomyces cerevisiae is the predominant species in most industrial fermentative processes, such as dough production, brewing, winemaking, cider production, sake, and also in traditional fermented beverages around the world like pulque, masato, chicha, sorghum beer, palm wine, etc. This species has been exploited by humans since ancient times. From an economic point of view, it may be considered the most important microorganism. Strains of biotechnological interest are highly specialized organisms that have evolved under stringent environmental conditions in different environments created by humans. Thus the physiological and genetic variability of yeasts strains isolated from different processes also present differences according to the origin and conditions of fermentations: temperature, pH or nitrogen sources. Sugar composition (glucose, fructose, maltose, sucrose, etc.) is extremely variable in nature and has important consequences on adaptation of fermentative yeasts (Querol *et al*., 2003; Barrio *et al*., 2006). The physiological differences of fermentative yeasts strains have also been found at the molecular level, which have revealed broad genetic variability among *S. cerevisiae* strains that has been correlated with their geographic origin and sources of isolation (Fay and Benavides, 2005; Liti *et al*., 2009).

Physiological and genetic diversity have been well-studied in strains associated with industrial processes like wine (Querol *et al*., 1994; Querol *et al*., 2003; Alba-Lois & Segal-Kischinevzky, 2010; Dequin & Casaregola, 2011; Schuller D. *et al*., 2012; Franco-Duarte R. *et al*., 2014), brewing (González *et al*., 2008) and beer (Alba-Lois & Segal-Kischinevzky, 2010). However, very little is known about other traditional fermentative strains from Latin America. Traditional fermentation beverages from Latin America derive from fruits, vegetables and cereal grains, mainly maize. For example, the masato beverage is of great importance to indigenous, natives and Amazon people from Peru and, like other fermentation products, it contributes greatly to their diet. This jungle beer is produced with rods of cassava (*Manihot esculenta* Crantz), which are cooked and ground to obtain dough with starch, this being the main carbohydrate source in their diet. Then in a first fermentation step, dough is chewed and spat out to break down starch to maltose by salivary amylase.

Other interesting traditional fermentation occurs during Mezcal production, which is similar to tequila. Yeast strains ferment the sugar obtained from *Agave salmiana* through acidic hydrolysis together with heat treatment. In this case, the main sugar involved is fructose at a high concentration (~160 mg/mL), but glucose is also present at lower levels (~27 mg/mL) (Michel-Cuello *et al*., 2008). Studying the biotechnological properties of yeast isolated by traditional fermentations from Latin America can be very important because it can lead to new strategies (like the use of different sugar sources) that can improve industrial processes.

In previous studies, we compared the physiological differences, optimal growth temperature (Salvadó *et al*., 2011) and ethanol susceptibility and resistance (Arroyo-López *et al*., 2010) of different *S. cerevisiae* strains and also other species in the *Saccharomyces* genus isolated from natural habitats, wine, sugarcane, agave and masato.

These studies showed that all the strains of *S. cerevisiae* are the best adapted to growth at high temperatures and present the greatest ethanol resistance compared to the other *Saccharomyces* species. Nonetheless, significant differences in ethanol tolerance have been observed among different *S. cerevisiae* strains (Arroyo-López *et al*., 2010). A genomic characterization of these *S. cerevisiae* strains by comparative genome hybridization has revealed also important differences in the copy number of the genes related with amino acid metabolism (Ibáñez *et al*., 2014). A strain isolated form sugarcane juice fermentation, in which arginine was a rare amino acid, has been reported to contain fewer copies of gene *CAR2* (Ibáñez *et al*., 2014). All these results suggest large differences at the physiological and genetic levels, when we compare strains of *S. cerevisiae* from different processes.

Many genome-wide gene expression studies in fermentative strains have been conducted using DNA microarrays to gain a better understanding of winemaking processes (Rossignol *et al*., 2003; Varela *et al*., 2005) or other aspects, such as influence of temperature on growth or aroma production (Beltrán *et al*., 2006; Pizarro *et al*., 2008), the genes involved in aroma production (Rossouw *et al*., 2008), the general or sugar stress response (Marks *et al*., 2008; Erasmus *et al*., 2003), or responses to nitrogen depletion (Backhus *et al*., 2001). Regarding traditional fermentations, only one study that worked with an agave *S. cerevisiae*

strain has been carried out (Ramirez-Córdova *et al*., 2012), and no study has compared the expression profile of strains from different traditional fermentative sources, which could be useful for understanding their differences at the molecular level.

Over the years, many technologies have been used to measure gene expression. Two of these are capable of measuring thousands of genes simultaneously: the older hybridization-based microarray technology and the more recent sequencing type-based RNA sequencing (RNAseq) technology. Although microarrays are a powerful, relatively inexpensive and mature technology, they have several limitations. One of the most important limitations is that the DNA arrays are built on the *S. cerevisiae* laboratory strain S288c genome and the study of yeast expression with differences in their genomic composition could generate partial information on gene expression. New generation sequencing (NGS) techniques are a powerful tool for both genomes and transcriptome analyses, and offer clear advantages over conventional methods, such as their: 1) ability to detect and quantify transcripts deriving from all regions of the genome; 2) wide dynamic range which affords high sensitivity for low-abundance transcripts; 3) single nucleotide resolution (Wang *et al*., 2009).

In this study, we investigated the differential expression in synthetic must fermentation of three *S. cerevisiae* strains isolated from different fermentative environments, using high-throughput sequencing. The selected yeasts were a wine yeast strain (T73), adapted to fermentations with glucose and fructose (50% each) as the main sugars; a masato strain (PE35M), adapted to fermentations with maltose as the main sugar, which presented the greatest ethanol tolerance; an agave strain (Temohaya-26), adapted to fermentation with fructose as the main sugar.

The selection of these strains was doing according to differences physiological properties that could be interesting to improve the wine processes.

The RNAseq platform used was ABI SOLiD given its low error rate and its ability to produce strand-specific sequencing data. Our results determined the transcriptional level for the majority of *S. cerevisiae* annotated genes. The analysis of the most up-regulated genes revealed an induction of genes that encoded mannoproteins (*CCW12, TIP1, SPI1, PIR3* or *SED1*) and comparison of strains revealed genes related with aroma (*ARO10* or *PDC6*) and

also genes implicated in sugar transport, glycerol and alcohol metabolism, which are known to be related with the physiological differences observed between strains and are responsible for activities of potential biotechnological interest. These data will help us to understand the molecular mechanisms and will provide a global insight into strain improvement by metabolic engineering or synthetic biology.

2. Results

2.1. *Studying global gene expression levels by RNA-Seq*

The use of yeast starters is a common procedure in the food industry. However, the market is demanding new properties that centre on aroma and character development in the wine field. Therefore, studying yeasts from traditional fermentations could be interesting to discover the novel strain attributes that could contribute to enhance wine properties. For this reason, we carried out a transcriptomic analysis of three *S. cerevisiae* strains isolated from different sources: wine, masato and mezcal. The gene expression profiles of these strains (T73, PE35M and Temohaya-26) were compared to each other. cDNA libraries were constructed from the poly (A)-enriched mRNA of *S. cerevisiae* synthetic must fermentations at 22ºC and were analyzed by high-throughput sequencing. A paired-end sequencing strategy was adopted in RNA sequencing. RNA samples were prepared from the micro-fermentations of *S. cerevisiae*. Three samples of each fermentation were taken when 50% of sugars had been consumed, which corresponded to the beginning of the stationary phase. Total RNA was isolated. Following mRNA purification rRNA depletion, RNA was sequenced in an AB SOLiD instrument. The depth coverage of coding regions was estimated by mapping sequenced reads to the published *S. cerevisiae* S288c genome. This measure was expressed in fragments per kilobase of exon and million fragments mapped (FPKM). Transcripts were assembled and their relative abundance was calculated with Cufflinks. Only those genes with a fold change in expression above four (positive or negative), compared to *S. cerevisiae* Lalvin T73, were taken into account for further analyses.

One clear advantage of the RNA-Seq technique *vs.* old transcriptomic approaches is that it enables the detection of the RNA from ORFs that is not associated with the reference genome, which was *S. cerevisiae* S288c. The reference genome and annotations for *S. cerevisiae* were downloaded from the Saccharomyces Genome Database (http://downloads.yeastgenome.org/).

In order to find, and to understand, the content of these parts, the reads that did not map onto the reference genome were investigated to see which sequences they derived from. To help us, a database containing genetic sequences from a wide array of different kinds of *Saccharomyces* genera was used.

Our main hypothesis was that some sequences could derive from other *Saccharomyces* species because cross-linked events were present, such as horizontal transfer, introgressions and hybridizations, which have not yet been identified to form part of the *S. cerevisiae* S288c reference genome. After comparing the genetic material that did not fit any known part of the S288c reference genome, no different sequences were found, and the contigs did not produce significant alignments. As this implies that a high level of contigs may be false-positives, it was assumed that most unmapped reads probably represented sequencing artefacts and polymorphisms.

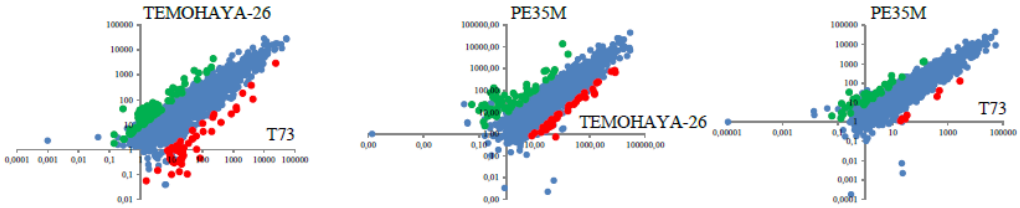
In each strain, 7151 RNAs were found after the alignment to the reference genome. It is highlighted that three of these corresponded to the genes of the mating type locus; 55 were mitochondrial genes, of which three were dubious open reading frames; 27 were unknown; four were genes of the 2-micron plasmid; and 413 were noncoding RNA from 16 chromosomes (Table 3.1). It is noteworthy that no significant differences were observed among the strains at the detected ncRNA levels. Figure 3.1A shows the distribution of the $\log_2$-fold changes observed among the different strains along the genome. PE35M and T73 expressed similar numbers of genes (with the largest overlap level). However, the differences found among the expression patterns of Temohaya-26 *vs.* T73 and PE35M *vs.* Temohaya-26 were bigger. These results generally indicated substantial differences in gene expression among the three strains (Figure 3.1B). There were 162 up- and 47 down-regulated genes in Temohaya-26 with at least 2-fold changes over T73; 54 up- and seven

down-regulated genes in PE35M over T73; 37 up- and 69 down-regulated genes in Temohaya-26 with over PE35M. There were only 23 differentially expressed genes in common with the three comparisons (*FLO1, CHS2, RDS1, AAD3, FMP45, SOR2, YDR261C-D, YER188W, IRC7, YHB1, ARN2, YHR210C, TIR3, YIL169C, SOR1, NFT1, AYT1, PUT1, HRI1, YNL284C-B, FRE4, COS10, FIT2*). The genes overexpressed in T73, compared to Temohaya-26, represented the largest number of differentially expressed genes. To assign putative functional roles to the obtained genes, GO terms were studied according to sequence similarities to known GO-annotated genes (see Figure 3.1B). The genes involved in Cell periphery, Transposition RNA-mediated, Iron chelate transport, Cell wall and Response to stress were the most represented groups. This indicates that yeasts adapt to different environments and conditions, and perform intensive metabolic activities.

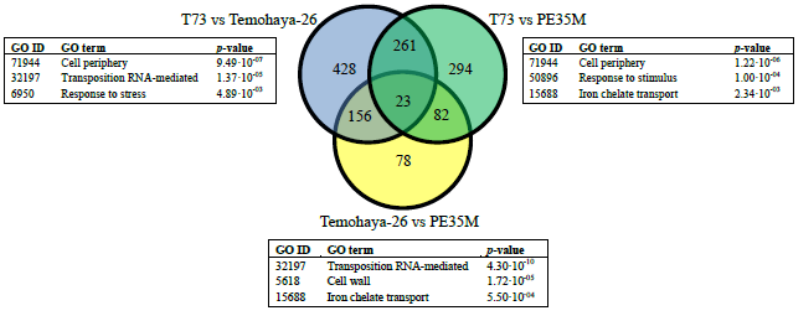
Table 3.1. Summary of all the expressed noncoding RNAs in *S. cerevisiae* strains after 50% sugar consumption in synthetic must.

Noncoding RNA	Number	Location	Names
tRNA	302	Spread across genome	-
snoRNA	75	Spread across genome	-
rRNA	17	Chr XII	15S rRNA, 21S rRNA
snRNA	4	Chr II, Chr V, Chr VII, Chr VII, Chr XIV	*LSR1, SNR14, SNR7-L/SNR7-S, SNR19*
other	15	Spread across genome	*HRA1, TLC1, RPR1, SRG1, SCR1, RUF21, RUF20, RUF22, RUF23, RUF5-1, RUF5-2,*

A



B



GO ID	GO term	*p*-value
71944	Cell periphery	$9.49 \cdot 10^{-07}$
32197	Transposition RNA-mediated	$1.37 \cdot 10^{-05}$
6950	Response to stress	$4.89 \cdot 10^{-03}$

GO ID	GO term	*p*-value
71944	Cell periphery	$1.22 \cdot 10^{-06}$
50896	Response to stimulus	$1.00 \cdot 10^{-04}$
15688	Iron chelate transport	$2.34 \cdot 10^{-03}$

GO ID	GO term	*p*-value
32197	Transposition RNA-mediated	$4.30 \cdot 10^{-10}$
5618	Cell wall	$1.72 \cdot 10^{-05}$
15688	Iron chelate transport	$5.50 \cdot 10^{-04}$

Figure 3.1. Overview of the gene expression analysis after 50% sugar consumption in synthetic must of industrial (T73) and traditional (Temohaya-26, PE35M) fermentative *S. cerevisiae* strains. (A) Scatter plots of the comparison of the expression profiles of the three strains. Each point represents the $\log_2$ (expression in Temohaya-26/expression T73; expression in PE35M/expression Temohaya-26) of the genes. Dots indicate the differentially expressed genes with a false discovery rate of < 0.05. Green dots indicate differentially up-regulated genes, red dots indicate differentially down-regulated genes and blue dots denote unchanged genes. (B) A Venn diagram showing the relations between differentially expressed genes (DEGs) and GO terms. A Venn diagram showing the relations between the DEGs in the comparisons Temohaya-26 versus T73, PE35M versus Temohaya-26 and PE35M versus T73, and the Go terms for DEGs. DEGs were filtered with a threshold of false discovery rate of ≤ 0.05 and an absolute $\log_2$ fold-change of ≥ 2.

2.2. *Transcriptome comparison between industrial wine strains and the strains isolated in traditional fermentations of Latin America*

The objective of this part of the work was to find differences in the gene expression of the *Saccharomyces cerevisiae* strains isolated from traditional fermentations comparing with industrial strain Lalvin T73. Several genes were up-regulated in Temohaya-26 and PE35M compared with the reference strain (Table 3.2). These genes codify the proteins involved in flocculation (*FLO1, FLO5, FLO9*), amino acid permease (*BAP3*), glutathione-related transporters (*OPT2, GEX2*) and members of the seripauperin multigene family, which are active during alcoholic fermentation (*PAU3, PAU10*). The fact that the functions of these genes are related to alcoholic fermentation to some extent suggests that fermentative (wine or Latin America) strains have adapted to their local fermentative environments.

Table 3.2. Up-regulated genes in traditional fermentative strains compared to industrial wine strain Lalvin T73 after 50% sugar consumption in synthetic must [a].

ORF	Gene name	Temohaya/T73	PE35M/T73
Protein involved in flocculation			
YAR050W	*FLO1*	3.67	5.99
YHR211W	*FLO5*	3.08	4.29
YAL063C	*FLO9*	3.22	4.61
Glutathione-related transporter			
YPR194C	*OPT2*	3.96	3.01
YKR106W	*GEX2*	3.12	3.25
Amino acid permease			
YDR046C	*BAP3*	3.16	3.58
Member of the seripauperin multigene family			
YCR104W	*PAU3*	3.21	3.83

[a] Ratio > 3. Values represent Log_2(ratio).

2.3. Gene expression and mannoprotein quantification of *S. cerevisiae strains*

Several mannoproteins (i.e. *CCW12, TIP1, SPI1, PIR3* or *SED1*) were among the most highly expressed during synthetic must fermentation (Table 3.3) and presented significant differences among the different strains, where T73 exhibited the highest expression. In order to phenotypically validate this, the amounts of cell wall mannoproteins and mannoproteins released by T73, Temohaya-26 and PE35M during fermentation al 22ºC were measured. Statistically significant differences were observed in the content of the cell wall mannoproteins in T73 compared to PE35M and Temohaya-26. T73 produced and retained more mannoprotein that the other two strains (Table 3.4). The levels of cell wall mannoproteins were much higher than the released mannoproteins, which is reasonable since most mannoproteins are released at the end of fermentation.

Table 3.3. List of the 20 most highly expressed genes in the different *S. cerevisiae* strains after 50% sugar consumption in synthetic must.

Gene name	Gene function	Gene expression level (FPKM)[a]		
		T73	Temohaya-26	PE35M
YFL031C-A	Unknown molecular function	42732.3	21882.9	7482.74
TDH1	Glyceraldehyde-3-phosphate dehydrogenase (NAD+)	41647.7	23089.2	37025.9
FBA1	Fructose 1,6-bisphosphate aldolase	30050.5	9373.28	20135.7
CCW12	**Cell wall mannoprotein**	**28276.4**	**13741.5**	**17960.1**
SED1	**Structural constituent of the cell wall**	**21173.5**	**14473.4**	**8436.69**
TIP1	**Cell wall mannoprotein**	**18951.2**	**-**	**-**
YDR524W-C	Unknown molecular function	18731	14299.7	8564.11
PDC1	Pyruvate decarboxylase	17722.4	8862.29	12886.3
ENO1	Phosphopyruvate hydratase activity	17418.1	8005.63	20462.7
SPI1	**GPI-anchored cell wall protein involved in acid resistance**	**14472**	**12711.3**	**8202.37**
TEF1	Translational elongation factor EF-1 alpha	10797.2	8352.35	-
PIR3	**O-glycosylated covalently-bound cell wall protein**	**10495.9**	**-**	**-**
HOR7	Unknown molecular function	10375.6	-	10567.3
OLE1	Delta(9) fatty acid desaturase	9950.74	12113.1	10287.8
SIP18	Phospholipid-binding hydrophilin	9512.55	18043.3	
RPL41B	Ribosomal 60S subunit protein L41B	8499.23	13409.1	6593.7
RPL41A	Ribosomal 60S subunit protein L41B	8079.05	12934.5	6556.05
YDR524C-B	Unknown molecular function	8313.9	6865.2	
IZH1	Membrane protein involved in zinc ion homeostasis	8295.94	-	-
HSP150	**O-mannosylated heat shock protein**	**8267.83**	**-**	**-**
YNL144W-A	Unknown Molecular function	-	6402.7	-
YOR302W	Translation regulation	-	23149.2	-
HSP12	Heat Shock Protein	-	9801.97	-
PBI2	Proteinase B Inhibitor	-	7521.25	-
SPS100	Protein required for spore wall maturation	-	6680.62	-
FIT3	Facilitator of Iron Transport	-	-	11335.3
PGK1	3-PhosphoGlycerate Kinase	-	-	9308.28
YKL153W	Unknown molecular function	-	-	8673.92
PRY1	Sterol-binding protein	-	-	7984.31
HSP26	Small heat shock protein (sHSP) with chaperone activity	-	-	7957.13
AHP1	Alkyl HydroPeroxide reductase	-	-	7722.91
ADH1	Alcohol dehydrogenase	-	-	6430

[a]FPKM is a measure of the gene expression level expressed as the number of fragments per 1-Kilo base of transcript and million mapped fragments.

Table 3.4. Mannoprotein content after 50% sugar consumption in synthetic must.

Strain	Cell wall mannoprotein (mg/g dry weight)	Released mannoprotein (mg /g)
T73	187.0 ± 8.8^{a}	11.1 ± 0.6^{a}
PE35M	160.8 ± 5.5^{b}	12.1 ± 0.4^{ab}
Temohaya-26	161.8 ± 3.4^{b}	13.1 ± 0.8^{b}

Values are expressed as mean ± standard deviation.
a,b different superscript letter within the column indicate significant differences (ANOVA and Tukey's test α=0.05)

2.4. *Gene expression and aroma production of S. cerevisiae strains*

The aroma production profile in beverage fermentations is one of the most important characteristics of yeast strain. Thus we focused on the differences in the gene expression of the genes related to aroma production observed among the various strains. We observed 10 of those genes showing significant differences among strains: two (*ADH6, AYT1*) were more expressed in Temohaya-26, three were more expressed in PE35M (*ALD5, PDC5, PDC6*), four (*ARO10, ARO9, BAP3, SFA1*) were more expressed in Temohaya-26 and PE35M than T73, and one (*BAT1*) was more expressed in PE35M and T73 compared to Temohaya-26. We also studied the aroma compounds detected in the three *S. cerevisiae* strains at mid-fermentation (Table 3.5). The results revealed notable differences among the aroma profiles released by yeast from distinct environments, which suggests that inoculated yeasts differ in terms of their fermentation performance. The statistical analysis gave significant differences for Isobutanol, Isoamyl alcohol and 2-Phenylethyl acetate production among the different strains. Temohaya-26 was able to produce more Isobutanol and Isoamyl alcohol, and obtained the highest scores for these three compounds. Yeast T73 produced intermediate levels of Isobutanol and Isoamyl alcohol compared to Temohaya-26 and PE35M, and a similar situation was observed in strains T73 and PE35M for 2-Phenylethyl acetate concentrations. When we calculated the total amount of acetates, alcohols and esters (Table 3.6), a similar result was seen:

Temohya-26 produced the largest amounts of alcohols and esters, and PE35M obtained the lowest ones, while wine strain T73 accomplished intermediate levels. When we compared transcriptomic and aroma compound determination, we found that the increased production of alcohols and esters in Temohaya-26 correlated with the high expression of *ADH6* and *AYT1,* respectively.

Table 3.5. Production of aroma compounds after 50% sugar consumption in synthetic must.

Strains	Ethyl acetate	Isobutyl acetate	Isobutanol	Isoamyl acetate	Isoamyl alcohol	Ethyl caproate	Ethyl caprylate	Ethyl caprate	2-Phenylethyl acetate	2-Phenyl ethanol
PE35M	5.14±0.36a	0.16±0.14a	3.13±0.17a	0.02±0.01a	11.48±1.57a	0.04±0.01a	0.07±0.01a	0.09±0.01a	0.08±0.01a	4.76±1.1a
T73	2.06±2.38a	0.31±0.27a	4.73±1.42ab	0.08±0.07a	16.5±3.76ab	0.07±0.05a	0.07±0.04a	0.05±0.01a	0.11±0.03a	7.1±1.83a
Temohaya-26	2.49±1.06a	0.13±0.03a	6.56±0.93b	0.05±0.04a	24.55±3.38b	0.02±0.01a	0.06±0.02a	0.08±0.03a	0.21±0.04b	8.5±0.57a

Amounts of aroma compounds are expressed in mg l^{-1}. The statistically significant differences in the concentration of different aroma compounds between strains are indicated by superscript letters.
a,b different superscript letter within the column indicate significant differences. Statistically different groups were established with 95% confidence.

Table 3.6. Total aroma compounds from the data in Table 3.5.

Strains	Total Acetates	Total Alcohols	Total Esters
PE35M	5.4±0.5[a]	19.4±2.1[a]	24.8±2.1[a]
T73	2.5±2.4[a]	28.3±5.9[ab]	30.9±5.9[ab]
Temohaya-26	2.9±1.1[a]	39.6±4.9[b]	42.5±4.9[b]

Values are expressed as mean ± standard deviation.
a.b different superscript letter within the column indicate significant differences.
Statistically different groups were established with 95% confidence.

3. Discussion

The most important contribution of the present work has been to underline is the use of a novel approach to study the transcriptome of different *S. cerevisiae* strains during synthetic must fermentation. High-throughput RNA sequencing (RNA-seq) offers many advantages for analyzing a population of RNAs and their composition (Wang *et al*., 2009; Wilhelm *et al*., 2008; Garber *et al*., 2011). Unlike microarrays and reverse transcription polymerase chain reaction (PCR), massive parallel sequencing requires no *a priori* knowledge of content, and enables quantitative transcriptome interrogation (Denoeud *et al*., 2008). This has led to the discovery of many novel transcripts, which in turn, has allowed a larger number of accepted classes of RNAs (Mercer *et al*., 2009; McCutcheon *et al*., 2003). Novel RNAs may be inferred by the computational analysis of RNA-seq results. RNA-seq is also able to make an accurate assessment of the relative levels of individual transcripts, including their related isoforms (Marguerat *et al*., 2010). Examining the transcriptome provides gene expression data, which can be used to detect the expression signatures associated with phenotypes of interest. The sensitivity of experiments to detect differentially expressed genes is determined by their ability to distinguish true changes in gene expression from alternate sources of variance.

Previous studies have revealed a complex nature of the yeast transcriptome, 90% of which has been presented by transcription beyond open reading frames; e.g. the transcriptome of

S. cerevisiae is represented by 10000 unique transcription units, including the transcripts of about 6000 genes, ORFs with upstream transcription starting sites (TSSs), ORFs with internal TSSs, intergenic transcription units, or ORFs with antisense RNAs, snoRNAs and micro-RNAs (Ito et al., 2008). The extensive expression of noncoding RNA is common in eukaryotes and has been demonstrated in yeasts *S. cerevisiae* (David *et al*., 2006; Miura *et al*., 2006; Xu *et al*., 2009) *Schizosaccharomyces pombe* (Ni *et al*., 2010) and *Candida albicans* (Sellam *et al*., 2010). In this experiment, we sequenced our samples profoundly enough to measure the expression of more than 95% of genes. Most of the novel transcripts identified herein were noncoding RNA genes (Table 1). ncRNAs are functional RNA transcripts that are not translated into protein. Previous studies have shown that they perform a wide range of functions in the cell (Eddy *et al*., 2001; Storz, 2002; Mattick *et al*., 2006; Costa, 2007). In *S. cerevisiae*, there is evidence to suggest that only a fraction of ncRNA is known. Gene expression analysis has shown transcription from many locations in the genome, which appear to be unannotated ncRNA genes (Samanta *et al*., 2006; David *et al*., 2006; Miura *et al*., 2006; Kavanaugh *et al*., 2009). In this study we identified several ncRNAs that may play fundamental roles in gene regulation (e.g., regulation of some coding genes) during synthetic must fermentation. Nonetheless, the most intriguing evidence is that no showed significant differences were found among the analyzed strains for ncRNAs, which presented large phenotypical differences. This finding suggests that the function of these molecules must be generally related to basic cell structure, maintenance and housekeeping functions, which are similar among *S. cerevisiae* species.

The transcriptomic comparison between two strains isolated in traditional fermentations and one industrial wine strain revealed large differences, which can be correlated with their physiologic performance as amino acid transport or flocculation. One interesting observation was the differences observed among the distinct strains in the levels of mannoprotein-related genes, and also in the mannoprotein levels themselves. Several studies have demonstrated a positive effect of mannoproteins on sensorial wine properties, and have contributed to the chemical stabilization of wine by preventing crystallization of tartrate salts (Feuillat *et al*., 1998, Gerbaud *et al*., 1996) and protein haze (Dupin *et al*., 2000; Gonzalez-

Ramos *et al*., 2008; Gonzalez-Ramos *et al*., 2006; Gonzalez-Ramos *et al*., 2009; Waters *et al*., 1994). Mannoproteins also stimulate lactic acid bacteria growth in wine environments, and thus the development of malolactic fermentation (Guilloux-Benatier *et al*., 2003), and allow the concentration of some undesirable compounds to lower, such as ochratoxin A (Ringot *et al*., 2005). The analysis of the most up-regulated genes revealed an induction of the genes that encode cell wall mannoproteins. The overexpression of mannoprotein genes correlated with an increased mannoprotein content in the yeast cell wall. The fact that mannoprotein increased in the wine strain compared to other fermentative strains could reflect the genetic domestication observed in this type of strains (Fay *et al*., 2005). The results of the present study suggest that adaptation to wine fermentative environments raises cell wall mannoprotein levels, although the amount of released mannoprotein is lower than in traditional fermentative strains. So the selection of mannoprotein-overproducing yeasts can be an interesting strategy to obtain better quality wines in the wine production process.

Another result is related to the differences observed in the aroma-related genes and compounds among the different strains. It has been claimed that yeasts and fermentation conditions are the most important factors to influence aroma (Lambrechts *et al*., 2000), which is one of the most important wine quality attributes. The most relevant aromatic compounds are higher alcohols, acetate esters and ethyl esters (Mountounet, 1969; Rapp *et al*., 1986). During alcoholic fermentation, several genetically distinct *S. cerevisiae* strains release various aroma compounds, which influence the organoleptic quality of wines. Different yeast strains contribute distinctly to wine quality; therefore, biodiversity studies of wine and natural yeast are necessary to discover strains with new molecular and enological attributes (Lopandic *et al*., 2007). Yeast strain Temohaya-26, isolated from agave juice, produces higher concentrations of alcohols and esters than the yeast isolated from wine T73 and masato PE35M. Esters determine the fruit aroma of wines, which indicates that this yeast from agave may contribute more to the aroma complexity than the other two strains. These results also suggest that using the hybrids strains between Temohaya-26 and a wine strain may favourably influence sensory wine properties. The current winemaking trend

consists in producing wine with different aroma nuances to offer a variety of wines to a developing market. For this reason, the selection and genetic development of yeast starter culture strains with improved aroma profiles would be interesting for the wine market and could help winemakers to meet consumers wine demands.

4. Conclusions

This study is essential to understand how gene expression variations contribute to the fermentation differences of the strains adapted to different fermentative environments. We observed differences in genes that encode mannoproteins, and those involved in aroma, sugar transport, glycerol and alcohol metabolism, which are important under alcoholic fermentation conditions. These differences were subsequently tested and confirmed physiologically. Such knowledge is crucial to better understand yeast mechanisms and overcome current limitations in fermentation processes. This will also support the development of novel tools for the genetic improvement or selection of wine yeasts.

CHAPTER 4

Comparative profiling of gene expression in *Saccharomyces* species using RNA-seq.

1. Introduction

Notwithstanding the fact that *Saccharomyces cerevisiae* remains by far the most widely used industrial yeast species to date, other *Saccharomyces* species, such as *Saccharomyces uvarum* (Naumov *et al*. 2000), *Saccharomyces kudriavzevii* (Naumov *et al*., 2000a) and *Saccharomyces paradoxus* (Redzepović *et al*.,2002), have also claimed their stake as valuable contributors to industrial fermentation processes. *S. kudriavzevii* and *S. uvarum* have interesting oenological properties which lead, for instance, to greater glycerol production or lower ethanol production compared to *S. cerevisiae* (Gamero *et al*., 2013; Oliveira *et al*., 2014). The wines produced by *S. uvarum* strains also have a stronger aromatic intensity than those produced by *S. cerevisiae* (Coloretti *et al*., 2006; Eglinton *et al*., 2000). Moreover, Redzepovic *et al*. (2002) and Orlic et al. (2007) reported the possibility of using *S. paradoxus* strains as starters in fermentation because of their excellent contribution to the aroma of the wines. It is worth noting that *S. paradoxus* strains isolated from fermentative environments exhibit physiological properties of biotechnological interest (Redzepovic *et al*., 2003; Belloch *et al*., 2008).

All these studies have shown the genome plasticity of yeasts revealing common transcriptional features. Studies performed in our laboratory showed that *S. uvarum*, S. *paradoxus* and *S. kudriavzevii* strains, exhibit interesting enological properties, making them an alternative to *S. cerevisiae* as wine starters according to the current winemaking trends (González *et al*., 2007; Orlic *et al*., 2010; Arroyo-López *et al*., 2010; Gamero *et al*., 2013; Gamero *et al*., 2104; Tronchoni J *et al*., 2014; Oliveira *et al*., 2014; Pérez-Torrado *et al*., 2015; Stribny *et al*., 2015). Physiological characterization of *S. kudriavzevii* strains showed up its cryotolerance, growing quite well at low temperatures (10ºC) (Belloch *et al*., 2008). Recently Zhang *et al*. (2014) characterized the sulphite tolerance and cold fermentation capability of strains of *S. uvarum*. They found a high frequency of sulphite-tolerance among the strains, which may be a consequence of their isolation from commercial wine fermentations to which sulphite is commonly added. The *uvarum* strains also showed good fermentation properties and a high degree of cryotolerance.

After the genome sequence of *S. cerevisiae* was reported (Goffeau *et al*., 1996), transcriptomic, proteomic, metabolomic, and phenotypic analysis have been conducted. DNA array is one of the most powerful tools to monitor the expression of genes from a whole genome in one single experiment (Lashkari *et al*., 1997). In the case of *S. cerevisiae*, many studies have been reported on genome-wide expression analysis using DNA microarrays to better understand the winemaking processes (Rossignol *et al*., 2003; Varela *et al*., 2005), temperature influence in growth or in aroma production (Pizarro *et al*., 2008; Beltrán *et al*., 2006), genes involved in aroma production (Rossouw *et al*., 2008), stress response (Marks *et al*., 2008; Erasmus *et al*., 2003) or the response to nitrogen depletion (Backhus *et al*., 2001). Only in few of these studies, genomic expression was correlated with the phenotypical data.

Despite the existence of several genome-wide expression studies in *S. cerevisiae* using DNA microarray technology, very few information was available regarding other species of the genus. Gamero et al., 2014 studied the expression of genes involved in aroma synthesis in three different *Saccharomyces* species (*S. cerevisiae*, *S. bayanus* var. *uvarum* and *S. kudriavzevii*) during winemaking at 12 and 28 °C; Tronchoni *et al*., 2014 compared *S. cerevisiae* and *S. kudriavzevii* transcriptome after yeast adapted to cold shock and their results confirmed that *S. kudriavzevii* is better adapted to grow at low temperatures, which can be relevant for industrial applications. In this study the authors demonstrated that the adaptation to low temperature is based on its enhanced ability to initiate a quick, efficient translation of crucial genes in cold adaptation among other strains, a mechanism that has been suggested for other microorganisms.

As it was explained in the previous chapter, the recent development of next-generation sequencing technologies can generate sequences on an unprecedented scale with a markedly reduced cost compared with traditional technologies (Shendure *et al*., 2008). Next-generation RNA-sequencing (RNA-Seq) has rapidly replaced microarrays as an approach to profile transcriptomes in a high-throughput way (Van Verk *et al*., 2013). It allows detection of transcripts with low abundance, identifies novel transcript units, and reveals their differential expression between different samples (Wang *et al*., 2009; Ozsolak *et al*., 2010). Few studies,

in which expression values were assayed more than once from the same sample, showed that RNA-Seq quantifies relative gene expression accurately (Marioni *et al*., 2008; Nagalakshmi *et al*., 2008).

To date, there have been no reports of using RNA-Seq technology to analyse the differential expression of different *Saccharomyces* species that inhabit different environments. By extending the transcriptome analysis to yeast species belonging to the *Saccharomyces* genus, it is possible to make comparative studies on transcriptome structure, and the results can help to better understand the gene expression regulation and its variability in different *S. cerevisiae* strains and *Saccharomyces* species. The availability of the complete or partially complete genome sequence of *S. uvarum*, *S. kudriavzevii* and *S. paradoxus* has made RNA-seq more informative.

In this study, we used next-generation sequencing technology and bioinformatics tools to analyse the transcriptome of three different *Saccharomyces* yeast species, by the alignment of transcripts to the respectively published *Saccharomyces* genome to analyse differentially expressed genes involved in flavour compound production, cold adaptation, nitrogen metabolism, and transport of solutes such as ions, sugars and metals and discovered specific genes of the new wine species.

2. Results and Discussion

2.1 *Differential gene expression in S. kudriavzevii, S. uvarum and S. paradoxus under fermentation conditions*

We analyzed the expression pattern of the transcripts at their relative abundance state. This comparative transcriptomic study of *S. kudriavzevii*, *S. uvarum* and *S. paradoxus* was carried out during wine micro-fermentations in 500 mL vessels of synthetic must at 22°C. The gene expression of these species was analysed at the middle of the exponential phase by taking samples. After extracting RNA and verifying quality, the samples were submitted to construct the libraries and perform RNA sequencing.

Raw data came with errors and needed to be preprocessed before being fed into downstream analyses like mapping or assembly. Basic tasks such as adapter removal, duplicate quantification and summary statistics on quality score were performed by standard tools like the fastQC toolkit. Otherwise, quality control was not yet formally established for RNA-seq data, and it was largely unclear how raw data trimming and quality filtering affect the end results. First, reads for each strain were mapped to the appropriate reference genome with TopHat. The reference genomes and annotations for *S. kudriavzevii, S. paradoxus* and *S. uvarum* were downloaded from the *Saccharomyces* Genome Database (http://downloads.yeastgenome.org/). After running TopHat, the resulting alignment files were provided to Cufflinks to generate a transcriptome assembly for each yeast strain. The reads were fed to Cuffdiff, which calculates expression levels and tests the statistical significance of observed changes between them. Cuffdiff reported numerous output files containing the results of its differential analysis of the samples. Gene and transcript expression level changes were reported in simple tabular output files that contained familiar statistics such as fold change (in $\log_2$ scale), *P* values (both raw and corrected for multiple testing) and gene/transcript related attributes such as common name and location in the genome. Browsing these files by eye was not especially easy, and working with data across multiple files was quite difficult. Another trouble we had to solve was to work with draft reference genome assemblies missing a substantial fraction of the genes and their common name. Because genome annotations were still incomplete, we had to modify the files with custom scripts written in Perl. We used BLAST-based orthology detection to personalize our reference sequences, and then align them to their *S. cerevisiae* ortholog to get the name of the genes. As long as the technology continues to change, the programs will have to change rapidly to keep up.

At the transcript level, within each library, on the basis of fold change in abundance, approximately 5100 genes were induced by all the three *Saccharomyces* species and a total of 464 genes were found to be differentially expressed in all the libraries taken together. Upon CR85 vs CECT12600, 138 genes were found to be differentially expressed, out of which 63

genes were found to be up regulated and 75 genes were down regulated. Upon CR85 vs 54, 172 genes were differentially expressed out of which 81 genes were up regulated and 91 genes were down regulated. When compared between CECT12600 and 54 libraries, 154 genes were found to be differentially expressed of which 85 genes to be up regulated and 69 genes were down-regulated (Figure 5.1).

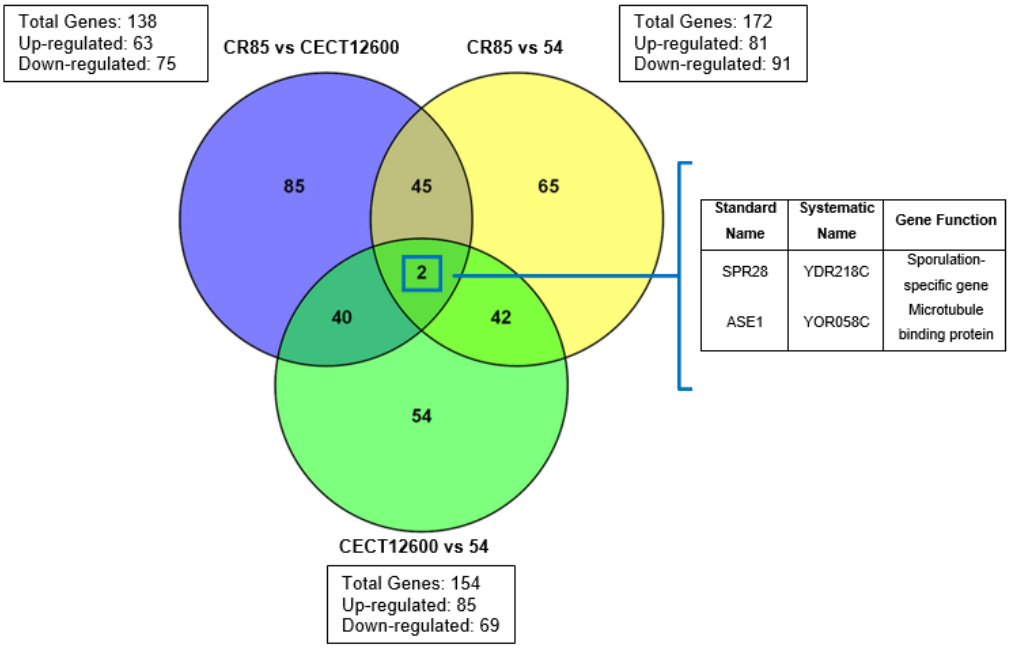


Figure 5.1. Venn diagram of genes identified as significantly differentially expressed (± 2, log2 scale) by RNA-seq data normalized by FPKM. Those genes common between the three *Saccharomyces* species are outlined in accompanying table.

Additionally, we checked the gene expression patterns in signficant GO terms using the FunSpec web server (Robinson *et al*., 2002). The functional details are depicted in Table 5.1.

Table 5.1. Significant gene ontology terms.

	Gene Ontology term	p-value	Genes annotated to the term	Number of Genes
CR85 vs 12600	meiosis [10.03.02]	4,84E-02	RDH54 PCH2 YDL114W TRS85 SPR28 LIF1 PDS5 SPS18 SPO21 SLK19 SRL4 BBP1	12
	spindle pole body [730.05]	4,71E-02	SPC110 CIN8 SFI1 SPC98 SPO21 SLK19 CTF19 BBP1 KAR9	9
CR85 vs 54	development of asco-basidio- or zygospore [43.01.03.09]	9,78E-02	GIP1 SPR28 SPS1 SPR6 LIF1 GPI1 SSP1 IME1 CDA1 SMA2 SPS18 RIM21 MPC54 REC8	14
	nucleus	0.001327	NUP170 REG2 AMN1 YDL063C SCM3 RPA14 UBA2 ECM11 UTP18 OMA1 SIR1 GRC3 CDA1 YLR455W RAD10 TAF13 YMR144W RLP7 NOP8 ASE1 RAD53 KAR9 REC8	23
12600 vs 54	meiosis [GO:0007126]	1,39E-06	PCH2 TRS85 NKP1 SPS1 BNS1 SAE3 REC104 SSP1 IME1 PDS5 TOF1 SPO21 MPC54 SLK19 RMI1 SPO19 REC8 KAR3	18
	cell cycle [GO:0007049]	1.19e-05	MCD1 RGP1 SPR28 NKP1 CIN8 SMC2 SPR3 BNS1 SSP1 AIM20 CLF1 PDS5 BNI5 TOF1 EGT2 SPO21 MPC54 SLK19 RMI1 CLN2 KAR3	21
	prospore membrane [GO:0005628]	1,97E-03	GIP1 SPS1 SPR3 SSP1 SPO21 MPC54 SPO19	7

The most important molecular functions induced include those associated with sporulation and meiosis including also genes involved in cell cycle. One of these differentially expressed genes is a regulator of sporulation encoded by *IME1* (Kassir *et al*., 1988; reviewed in Kupiec *et al*., 1997, Kassir *et al*., 2003). In budding yeast, exit from the cell cycle and entry into meiosis depends on multiple signals, including mating type, absence of glucose, starvation, and stress. These signals are transmitted to the transcriptional activator *IME1*, which serves as the master regulator of meiosis (reviewed in Kupiec *et al*., 1997, Kassir *et al*., 2003). *IME1* transcription is repressed when glucose is provided as unique carbon source; conversely, transcription is induced in the presence of acetate and no other carbon source (Kassir *et al*., 1988). Moreover, nitrogen depletion leads to an induction of *IME1* transcription (Kassir *et al*.,

1988). Thereby, IME1 expression results in the activation of the sporulation program and meiosis. Yeast sporulation, the coupling of meiosis and spore formation in response to nutrient deprivation, provides a well-studied model of a cell fate decision in response to external cues (Kupiec *et al*., 1997). Sporulation efficiency also varies among diverse isolates of yeast (Mortimer, 2000). Thus, these results provide an insight into how the selection pressure due to changes in the environmental conditions of the different species (such as nutrient availability) can drive evolution of a phenotype, such as variation in sporulation efficiency, and raise additional questions regarding signaling specificity.

Much work remains to be done to produce complete finished transcriptomes, but the information that has become available through this effort allows a global perspective on the *Saccharomyces* transcriptomes. Although the details will change as the reference genome sequences are finished, many points are already clear. The transcriptomic landscape shows marked variation in the expression of a number of genes involved in sporulation. Nevertheless, the obtained results need to be supported. This can be performed following two different pathways: confirm the expression variation of the genes which are thought to have a greater influence in the fermentation process or improve the confidence on the pipeline results. These procedures are complementary and, therefore, ideally both should be performed in order to validate the results.

3. Conclusions

In this chapter, we start by presenting background information on the transcriptome analysis of different *Saccharomyces* species and describing the generation, assembly and evaluation of the draft sequences. We then focus on an initial analysis of the differential gene expression. The analysis and comparisons are drawn throughout with the only reference genomes available to date, but will be improved and expanded with data from the genome we are getting in our group. We recognize that it is impossible to provide a comprehensive analysis of this vast dataset, and thus our goal in this work is to illustrate the range of insights

that can be gleaned from the different *Saccharomyces* transcriptomes and thereby to sketch a new biotechnological research for the future.

Our data suggest that sporulation efficiency can be an important target of adaptative evolution when cells face changing environments. Expanding the whole genome sequence and transcriptome sequence datasets of strains from fermentative environments and other anthropic niches will provide a better understanding of the evolutionary history of strains and the frequency of the mechanisms that yeast use to adapt to the fermentative environment, including nucleotide and structural variations, introgressions and horizontal gene transfer.

6. GENERAL DISCUSSION

Due to its compact genome, *S. cerevisiae* has become a useful model for comparative analysis, and in the last few years large-scale sequencing studies have been performed to characterize the strains of the *Saccharomyces* genus used in industrial applications (Pretorius *et al*., 1999; Borneman *et al*., 2008; 2011) and to clarify the complex structure of yeast populations (Liti *et al*., 2009). The hostile conditions of the oenological environment determined by vinification processes (high sugar concentrations, high ethanol levels, low pH, presence of sulphites) have produces numerous selective schemes and have conformed the genomic content of the strains used in industrial applications (Borneman *et al*., 2011). Studies on industrial wine yeast populations have highlighted the process of adaptation of domesticated yeast lineages to the chemical treatments imposed by viticulture and winemaking (Townsed *et al*., 2003; Hodgins-Davis *et al*., 2012). The relationships between gene expression and environmental responses have also been investigated (Hodgins-Davis & Townsend, 2009).

The existing natural diversity of yeast strains provides a rich, yet underexplored source of strains with industrial potential. Recent advances in next-generation sequencing technologies have allowed scientists to chart the diversity to an unprecedented level of detail. This revealed that the genetic diversity of currently employed industrial strains is relatively limited. Therefore, high-throughput screening of (natural) yeast collections or investigation of the phenotypic potential of indigenous strains might already yield yeasts with superior characteristics compared to the currently used strains. Moreover, functional genomic approaches, such as the “array-CGH” (“aCGH”) technique, are powerful tools for the analysis of genomic DNA copy number changes and genome rearrangements (Dunn *et al.*, 2005; Carreto *et al.*, 2008; Ibáñez *et al.*, 2014). This technique has been used to explore the genomic diversity of different *S. cerevisiae* strains (Perez-Ortin *et al*., 2002; Winzeler *et al*., 2003, Infante *et al*., 2003) as well as the genomic architectures (relative to *S. cerevisiae*) of other closely-related *Saccharomyces* species (Bond *et al*., 2004, Edwards-Ingram *et al*., 2004).

In this Doctoral Thesis we focused on diverse strains of the genus *Saccharomyces* with different characteristics, which we sought to decipher through comparative genomics and transcriptome sequencing. Our results contribute to the understanding of the ecology, evolution, and genotype–phenotype relationships of natural yeast strains.

Until recently, studies of the genetic diversity and instability of wine yeasts (reviewed by (Benitez *et al*., 1996; Perez-Ortin *et al*., 2002) were performed using DNA microarrays based on both gene expression patterns (Eisen MB & Brown PO,1999; Holloway AJ *et al*., 2002; Rossignol *et al*., 2003; Varela *et al*., 2005), and also on genomic DNA copy number changes and genome rearrangements through the "array-CGH" ("aCGH") technique (Dunn *et al*., 2005; Carreto *et al*., 2008; Peris *et al*., 2012; Deregowska et al., 2005; Wnuk M *et al*., 2015).

Industrial yeasts, economically important microorganisms, are widely used in diverse biotechnological processes including brewing, winemaking and distilling. In contrast to a well-established genome of wine yeast strains, the comprehensive evaluation of genomic features of natural strains is lacking. It is therefore very interesting from an industrial point of view to know the global mechanism of yeast adaptation to their fermentative lifestyle. This knowledge would allow the selection of novel yeast strains with new properties for use in wine fermentation. Currently there are more than 200 commercially produced wine yeast strains available to winemakers, however, very few of them have been selected based on improving wine aroma, producing high amounts of glycerol, or having good fermentative capacity at low temperatures and high ethanol levels.

At Chapter 1, the genomic differences among several *S. cerevisiae* strains isolated from traditional fermentations around the world were determined by using "microarray karyotyping" (also known as "array-CGH" or "aCGH") technique. Molecular and genetic approaches were used to evaluate the genetic relatedness among isolates of the yeast *S. cerevisiae*. Similar chromosome profiles were observed within the *S. cerevisiae* group. However, some chromosome number variations, that could explain the aneuploidy values obtained by flow

cytometry, as well as chromosomal length variations were evident. Genome-wide array-CGH analysis reveals variations in subtelomeric and intrachromosomal gene families. It is worthwhile to note that the strain relatedness based on array-CGH data was comparable with electrophoretic karyotyping-based similarities among strains. Statistically significant differences in the gene dosage were observed in metabolic functions related to cellular homeostasis, cell-to-cell interactions, and transport of solutes such as ions, sugars and metals. These results may suggest that the strains analyzed may differently respond to changing environments and may have diverse adaptation strategies. Increased and decreased copy number of *AGP3* in the Kyokai nº7 (sake yeast) and GB-FlorC strains was shown, respectively. Decreased copy number of *DAK2* gene encoded a glycerone kinase involved in glycerol catabolism was observed in all the strains. These genes were suggested as the most important genes present in wine yeast with respect to laboratory strains (Dunn *et al*., 2012). The other genes with affected copy number were mainly involved in carbohydrate and amino acid metabolism, and ion transport that may also modulate a biotechnological process. The dosage of the *AYT1* gene, which encodes an acetyltransferase and numerous genes implicated in maltose metabolism was affected. The *MAL* gene family of *Saccharomyces* is comprised of five multigene complexes, *MAL1*, *MAL2*, *MAL3*, *MAL4* and *MAL6*, located at or near the telomere of a different chromosome, any one of which is sufficient for yeast to metabolize the disaccharide maltose.

The traditional fermentative strains also differed in the copy number of *PUT1* (the main player in the assimilation of proline as a nitrogen source), especially highly elevated *PUT1* gene copy number was revealed in strains T73, adapted to grape must fermentation, and PE35M, adapted to cassava (Manihot esculenta). Interestingly, proline is the major amino acid in grape must and is also an important nitrogen source in cassava (Bradbury and Holloway, 1988). This fact might justify the increased copy number of *PUT1* of these two strains. However, the effects observed in the strain Kyokai nº7 were opposite, showing the worst capacity to grow with proline as a nitrogen source. This could be explain because this strain is adapted to sake mash fermentation where proline is absent (Lemura *et al*., 1999). Similarly,

decreased copy number of *CAR2* gene, involved in arginine catabolism, in the CPE7 strain, was observed. This, again, could be due to the absence of arginine in the sugarcane where this strain was isolated (Wiggins and Williams, 1955).

Other interesting results were the differences between four *S. uvarum* strains. In the results of Chapter 2, the genomic diversity was mainly revealed within subtelomeric regions. Among the variable genes, the four strains selected differed in subtelomeric genes involved in zinc metabolism and metabolic genes involved in non-fermentable carbon utilization. For this reason, we determined the ability to growth on non-fermentable carbon sources. Although no growth defect was observed, the strains displayed differences in growth when ethanol, acetate or glycerol was added to the medium. Regarding the study to show zinc requirements, *S. uvarum* strains show increased sensitivity to limiting zinc compared to *S. cerevisiae*, which is in agreement with the smaller copy number of the zinc metabolism genes in this species. The genomic variability found in this study supported other previous studies showing variability of sub-telomeric genes involved in secondary metabolism linked to environmental adaptation (Dunn *et al*., 2005; Carreto *et al*., 2008; Ibáñez *et al*., 2014; Deregowska *et al*., 2015; Pérez-Torrado *et al*., 2015).

The genomic data of both Chapter 1 and 2, may be helpful for better understanding of the fermentative environment-mediated changes in the yeast genome and accompanying phenotypic features. Thus, the knowledge on genetic diversity of nonconventional yeasts strains may be further exploited in economically important biotechnological processes. Unfortunately, we are using spotted DNA microarrays only containing probes for the complete gene set of the S288c reference strain; therefore we are probably missing important differences between the *Saccharomyces* strains because of this technical issue.

Even though, initially microarrays were instrumental in whole transcriptome analysis, currently RNA-seq is becoming a preferred method of choice, since it is considered to effectively surmount the limitations of microarray. To date, one paper has been published on

transcriptomics of three genetically distinct industrial yeast strains belonging to the *S. cerevisiae* species (Aslankoohi *et al*., 2013). The authors performed RNA-seq analysis of RNA samples from *S. cerevisiae* cells embedded in fermenting bread dough, and studied changes in the yeast transcriptome throughout the fermentation process.

The Chapter 3 is the first report on detailed evaluation of transcriptomic features of two *S. cerevisiae* yeast strains isolated from traditional fermentations. For this, we carried out a comparative expression analysis between the representative wine strain of *S. cerevisiae* (T73) and the other two *S. cerevisiae* strains isolated from masato (PE35M) and agave (Temohaya-26). The results obtained from this chapter have shown us that there are substantial differences in gene expression between the three strains but not with the same intensity. PE35M and T73 expressed similar numbers of genes, whereas the differences found among the expression patterns of Temohaya-26 *vs.* T73 and PE35M *vs.* Temohaya-26 were higher. GO analysis showed a significant overrepresentation of genes associated with cell periphery, transposition RNA-mediated, iron chelate transport, cell wall and response to stress. The up-regulated genes obtained in the comparison between the industrial wine strains and the strains isolated in traditional fermentations of *Latin America*, encode proteins involved in flocculation, amino acid metabolism and transporters, as well as members of the seripauperin multigene family, which are active during alcoholic fermentation. This fact suggests that fermentative (wine or *Latin America*) strains have adapted to their local fermentative environments. Furthermore, genes encoding mannoproteins and genes related to aroma production were among the most highly expressed during synthetic must fermentation and have significant differences between the strains. Mannoproteins, have attracted much attention in the winemaking due to their reported contribution to wine quality and chemical stability (Caridi A., 2006; Gonzalez-Ramos & Gonzalez, 2006; Quirós *et al*., 2010; Quirós *et al*., 2012; Perez-Través *et al*., 2015). Previous studies have also suggested the direct involvement of mannoproteins in yeast cold adaptation (López-Malo M. *et al*., 2015). As the transcriptomic analysis revealed this strong up-regulation of several

mannoprotein genes, the amount of extracellular and cellular mannoproteins was determined. T73 released more mannoproteins to the growth medium than PE35M and Temohaya-26 because the synthetic must fermented with this strain presented higher mannoprotein content. However, the levels of cell wall mannoproteins were higher than the released mannoproteins. This fact is justified because mannoproteins are mainly released at the end of fermentation.

To investigate whether the differences at the genomic level have an impact on the oenological performances of the investigated yeast strains we analyzed the aroma compounds in the fermentation products of synthetic must by GC. Significant differences were found among the samples, indicating that *S. cerevisiae* strains show different metabolic activities during the fermentation of the different substrates (grape juice, agave juice and cassava). The analysis of the volatile compounds showed significant heterogeneity among strains in the production of almost all investigated compounds. In general, Temohaya-26 isolated from agave produced higher concentrations of esters and alcohols than the yeast isolate from masato and the wine yeast, but an overall difference in the production of Isobutanol, Isoamyl alcohol and 2-Phenylethyl acetate was also observed. Esters determine the fruity aroma of wines, indicating that the yeast isolates from agave may contribute to the aroma complexity to a larger extent than the yeasts from the other two sources. Thus, the up-regulation of *ADH6* and *AYT1* (enzymes involved in aromatic and branched-chain amino acids synthesis) reveals the better production of alcohols and esters by Temohaya-26.

Apart from all the genes with known function, it is important to mention that a large group of differentially expressed genes encoded putative proteins of unknown functions. This has also been observed by other studies of stress conditions (Gasch *et al*., 2000; Alexandre *et al*., 2001). By integrating results obtained in the transcriptomic analysis performed with physiological data our study provided, for the first time, an integrated view into the adaptive responses of *S. cerevisiae* to the challenging environment of fermentation.

By extending the transcriptome analysis to yeast species belonging to the *Saccharomyces* genus, it was possible to examine the differences in the expression level

of homologous genes and specific genes of new wine species (*S. uvarum*, *S. kudriavzevii* and *S. paradoxus*) that are not found in *S. cerevisiae*. Such knowledge is crucial for improvements in the fermentation processes and defining targets for the genetic improvement or selection of wine yeasts. Previous physiological and oenological works from this laboratory have already indicated the tremendous advantage of *S. kudriavzevii* fermenting at low temperature, and have shown its well-established cryotolerant character (Tronchoni *et al*., 2009; Gamero *et al*., 2013). Moreover, *S. kudriavzevii* and *S. uvarum* show differences in the production of volatile aroma compounds during wine fermentation (Gamero *et al*., 2013, 2014). It is also worth noting that *S. paradoxus* strains isolated from fermentative environments exhibit physiological properties of biotechnological interest (Redzepovic *et al.* 2003; Belloch *et al.* 2008; Orlić *et al*., 2007, 2010).

In Chapter 4, we investigate the potential of a comparative functional transcriptomic approach to correlate these oenologically relevant phenotypes to specific gene expression patterns. We have not seen "key" genes that would explain by itself the better performance of these three species at fermentation conditions, for now. Unfortunately, not all the reference genomes have the quality of *S. cerevisiae* S288c and we had to deal with low-quality genome references for *S. kudriavzevii*, *S. paradoxus* and *S. uvarum*. Errors in annotation impact downstream analyses. The errors that affect the location of annotated features or that result in a missed genomic feature impact the evolutionary studies and biological understanding of the organism, whereas mistakes in functional annotation lead to subsequent problems in the analyses of pathways, systems, and metabolic processes. The *S. kudriavzevii* (IFO1802), *S. bayanus* var. *uvarum* (CBS 7001) and *S. paradoxus* reference genomes are still in the scaffold format and chromosomes still needs some work. If complete genomes are to be efficiently utilized as reference genomes it is essential that they represent the highest quality annotation possible. Continuous reassessment of annotations based on new evidence led to improved annotations on a number of sequences, even though the process is recognized as being time-intensive.

The results obtained have shown that the most important molecular functions induced in these three species include those associated with sporulation and meiosis including also genes involved in cell cycle. As indicated by GO analysis, the regulator of sporulation *IME1* was found to be tightly regulated in all the strains studied, which may explain the activation of the sporulation program and meiosis, since it is well known as the master regulator of meiosis (Kupiec *et al*., 1997, Kassir *et al*., 2003). Previous studies have also provided candidate genes contributing to phenotypic variation in the sporulation efficiency of natural isolates of yeast (Tomar P *et al*., 2013). Thus, these results provide an insight into how selective pressures lead to different genomic outcomes and phenotypic diversity. Of course we know that much work still remains to be done before the final finished transcriptomes. We recognize that gaps remain in this chapter, but now people from our lab are getting on with the final sequence and analysis. Our future work will focus on improving these resources to enable better understanding of metabolic pathways that are active during fermentation in these Saccharomyces species. And the whole-genome sequencing of more *Saccharomyces* strains would complement this preliminary information provided in this doctoral thesis about the genomic bases of adaptive divergence (a process known as transcriptome-induced phenotype remodelling).

Understanding how sequence variation influences phenotypic diversity is a major challenge to address adaptation mechanisms of wine yeasts. In this work we aimed to identify the genetic basis of fermentation traits and gain insight into their relationships with variations in gene expression among yeast strains. The results obtained highlight the usefulness of yeast as a model system to study genomic variability in the context of environmental and evolutionary genomics. However, we are aware that these results are a starting point for new studies to better understanding of the mechanisms underlying adaptation to diverse fermentation conditions. More extensive genotyping and phenotyping of both natural isolates and nonconventional yeasts will help to identify strains and species with novel and/or improved industrially important properties.

7. CONCLUSIONS

- Microarray-based comparative genomic hybridization (array CGH) is a useful tool for analyzing the genome structures of *S. cerevisiae* strains. Gene copy-number changes are very variable among the strains, and the majority of them were observed in subtelomeric and intrachromosomal gene families involved in metabolic functions related to cellular homeostasis, cell-to-cell interactions, and transport of solutes such as ions, sugars and metals. Therefore, array CGH is of great use for studying telomeric genes.

- Nitrogen metabolism has played an important role in the adaptive evolution of *S. cerevisiae.* We observed a strong adaptation to nitrogen sources, illustrated by the GCN changes observed for *PUT1*, the main player in the assimilation of proline as a nitrogen source, and also for *CAR2*, involved in arginine catabolism.

- The genomic comparison of *S. uvarum* strains found the greatest variability in subtelomeric regions and in Ty-element insertion sites, suggesting that this type of genome variability is the main source of genetic diversity in yeast populations. Despite only small variations in gene copy numbers between the different yeast strains, *S. uvarum* exhibited high diversity in the presence, absence and number of genetic material.

- High-throughput sequencing of RNA (RNA-Seq) allows for the first time the simultaneous measurement of sequence and expression of RNAs and the analysis of these data requires novel bioinformatics approaches that have been developed in the present work.

- Yeasts isolated from different fernentations are differentiated by producing mannoproteins. The overexpression of mannoproteins genes is consistent with the increased mannoprotein content in the yeast cell wall. The wine strain showed higher amount of cell wall mannoproteins than the other fermentative strains, so the selection os mannoprotein-overproducing yeasts can be an interesting strategy to obtain better quality wines in the wine production process.

- The yeast strain Temohaya-26, isolated from agave juice, produces higher concentrations of alcohols and esters than the yeasts isolated from wine T73 and masato PE35M. Esters determine the fruit aroma of wines, which indicates that Temohaya-26 may contribute more to the aroma complexity than the other two strains. These results also suggest that using the hybrids strains between Temohaya-26 and a wine strain may favorably influence sensory wine properties.

- Much work remains to be done to classify the genes from *S. kudriavzevii, S. uvarum* and *S. paradoxus* and characterize exhaustively their expression and functions of the subfamilies. Anyway, the preliminary data show how selection pressure due to changes in environmental conditions can lead to the evolution of phenotypes. The main diferences are observed in the expresión of the gene *IME1*, the master regulator of meiosis. The expression of *IME1* is influenced by many factors, including ploidy, cell cycle status, nutritional environment, respiration, and pH. This result suggests that sporulation efficiency could be an important target of adaptative evolution when cells face changing environments.

- These studies represent a thorough overview of transcriptional changes with specific goals in mind utilizing the high-throughput sequencing. Such methods and knowledge provide us a means to characterize sources of phenotypic variation or consequences of physiological conditions to the genetic and genomic level in a whole organism in a single experiment. This ability to gather whole transcriptome information and perform comparative and correlative analyses allows us to extend the scope and capacity to the global scale high-resolution analysis for functional genomics. This opens the door to infinite possibilities including biotechnological applications to wine and other fermentative processes.

8. SPANISH SUMMARY

1. Introducción

Las levaduras del género *Saccharomyces* (principalmente *Saccharomyces cerevisiae*) son las responsables de la fermentación alcohólica (Pretorius, 2000). Aunque *S. cerevisiae* es la especie más frecuente en fermentaciones vínicas, y modelo de estudio (Pretorius, 2000; Serra *et al*., 2005; Barrio *et al*., 2006), también pueden estar presentes durante el proceso especies como *S. uvarum* (Naumov *et al*., 2002; Rementería *et al*., 2003; Demuyter *et al*., 2004), *Saccharomyces paradoxus*, aislada de viñedos croatas (Redžepovic *et al*., 2002) o también híbridos naturales entre especies del género *Saccharomyces* como *S. cerevisiae x S. kudriavzevii* (González *et al*., 2006) *y S. cerevisiae x S. uvarum* (Le Jeune *et al*., 2007).

S. uvarum es una levadura criotolerante y se caracteriza por presentar perfiles de fermentación con mayor velocidad de consumo de fructosa, menor producción de acidez volátil y mayor producción de glicerol, ácido succínico y alcoholes superiores que *S. cerevisiae*, lo que la convierte en una especie biotecnológicamente interesante para la industria vínica. *S. paradoxus* tiene la capacidad de degradar parcialmente el ácido málico y además presenta actividad pectinolítica, dos características que normalmente no se encuentran en las levaduras *S. cerevisiae*. La degradación del ácido málico puede ayudar a la desacidificación biológica del vino, mientras que la actividad pectinolítica puede contribuir a su clarificación y filtrabilidad. *S. kudriavzevii*, que sólo se ha aislado en ambientes naturales (Naumov *et al*., 2000; Sampaio y Gonçalves, 2008; Lopes *et al*., 2010), se ha mostrado como una levadura criotolerante, con una temperatura óptima de crecimiento de 25ºC y un rango de temperatura de crecimiento entre 6 y 32ºC (Arroyo-López *et al*., 2009; Salvadó *et al*., 2011). Aunque nunca ha sido aislada en condiciones de fermentación, sus híbridos (*S. cerevisiae* x *S. kudriavzevii*) se han encontrado en procesos industriales.

Las cepas de interés biotecnológico son organismos altamente especializados que han evolucionado bajo condiciones ambientales rigurosas en diferentes ambientes creados por los seres humanos. Por lo tanto, la variabilidad fisiológica y genética de las levaduras

aisladas de diferentes procesos, está asociada a diferencias en el origen geográfico y en las condiciones de fermentación (temperatura, pH o fuentes de nitrógeno). La composición de azúcares (glucosa, fructosa, maltosa, sacarosa) es extremadamente variable en la naturaleza y tiene consecuencias significativas sobre la adaptación de las levaduras fermentativas (Querol *et al*., 2003; Barrio *et al*., 2006). Se han encontrado diferencias fisiológicas a nivel molecular, lo cual se ha correlacionado con su origen y fuentes de aislamiento (Fay *et al*., 2005; Liti *et al*., 2009).

La diversidad fisiológica y genética han sido bien estudiadas en levaduras asociadas a diferentes procesos industriales como el vino (Querol *et al*., 1994; Querol *et al*., 2003; Alba-Lois *et al*., 2010; Dequin *et al*., 2011; Schuller *et al*., 2012; Franco-Duarte *et al*., 2014) y la cerveza (Alba-Lois *et al*., 2010). Sin embargo, se sabe muy poco acerca de otras levaduras de fermentaciones tradicionales de Latino América. El estudio de las propiedades biotecnológicas de estas levaduras aisladas de fermentaciones tradicionales, puede ser muy importante ya que puede dar lugar a nuevas estrategias para la mejora de los procesos industriales.

Se han llevado a cabo muchos estudios de expresión génica utilizando microarrays de ADN en cepas de *S. cerevisiae* para entender mejor los procesos de elaboración del vino (Rossignol *et al*., 2003; Varela *et al*., 2005), u otros aspectos como la influencia de la temperatura (Beltrán *et al*, 2006; Pizarro *et al*., 2008), el crecimiento o la producción de aroma (Rossouw *et al*., 2008), la respuesta a estrés (Marks *et al*, 2008; Erasmus et al., 2003), o la respuesta al agotamiento de nitrógeno (Backhus *et al*., 2001). Esta técnica también se ha empleado en estudios de expresión en otras levaduras del género *Saccharomyces* como *S. kudriavzevii* y *S. uvarum*, para conocer los mecanismos de adaptación a bajas temperaturas de fermentación (Tronchoni *et al*., 2014), entender las diferencias en cuanto a la síntesis de aromas (Gamero *et al*., 2014, 2015) y estudiar los perfiles de expresión de los híbridos entre estas especies (Combina *et al*., 2012; Gamero et al., 2015). Aunque los microarrays son una tecnología de gran alcance y relativamente barata, tiene varias limitaciones. Una de las más importantes es que los arrays se

construyen con el genoma de la cepa de laboratorio *S. cerevisiae* S288c, y el estudio de la expresión de levaduras con diferencias en su composición genómica podría generar información parcial sobre la expresión génica.

Actualmente, y gracias a los avances en las técnicas de secuenciación del ADN, a través de tecnologías de nueva generación, NGS (del inglés Next Generation Sequencing), se han revolucionado campos como los de la genómica y la transcriptómica. Estas tecnologías están permitiendo no solo generar información con altos rendimientos, sino también abrir nuevos horizontes para el entendimiento detallado y global de procesos de expresión génica. La caracterización completa y el análisis global de la expresión génica, aun sin ninguna información genómica previa, es ahora posible a través de la secuenciación directa del ARN, tecnología conocida como RNA-seq (Wang *et al*., 2009; Garber *et al*., 2011; Egan *et al*., 2012; Ward *et al*., 2012). Esta herramienta transcriptómica cambia la manera de cómo se analizan y comprenden los transcriptomas (Wang *et al*., 2009). La información obtenida es de gran utilidad para vislumbrar procesos metabólicos y mecanismos de adaptación a las condiciones ambientales a las que se exponen las levaduras.

Un aspecto importante de la fisiología de las levaduras está relacionado con el control de la expresión génica. Las diferencias genéticas y los cambios en la estabilidad genómica de las levaduras pueden afectar los procesos fermentativos. Por lo tanto, es importante estudiar los cambios en el número de copias de ADN genómico de las cepas, así como los niveles de expresión de sus genes.

2. Objetivos y Metodología

El objetivo principal de esta tesis ha sido la caracterización genómica y el estudio de la expresión génica de diferentes especies del género *Saccharomyces*. Se han utilizado técnicas moleculares y técnicas de secuenciación masiva para el análisis de la variabilidad genómica y del transcriptoma durante fermentaciones en mosto sintético, con el fin de comprender mejor cómo funciona la adaptación de estas levaduras a los diferentes

ambientes fermentativos, y con ello ayudar a la industria biotecnológica ofreciendo nuevas cepas que puedan generar mejores vinos y otras bebidas fermentadas.

El objetivo general se divide en los siguientes objetivos parciales:

2.1 Caracterización genómica de cepas de *S. cerevisiae*

En el presente estudio, se ha realizado un estudio comparativo de seis cepas de levaduras de *S. cerevisiae* aisladas de diferentes fermentaciones tradicionales (masato, mezcal, aguardiente, sake, vino y vino de Jerez) y una cepa natural.

Se utilizaron chips de Hibridación Genómica Comparada (aCGH) para determinar la variación genética entre las cepas. La técnica consiste en marcar el DNA genómico de la cepa estudio con un fluoróforo, y el de la cepa haploide de referencia *S. cerevisiae* S288c con otro. Los dos DNAs genómicos se hibridan en el mismo chip. Después de la hibridación, la fluorescencia de cada fluoróforo (genoma) se detecta mediante un escáner, y las diferencias en la intensidad de fluorescencia de cada marcaje permiten estimar el número de copias de cada gen de *S. cerevisiae* comparado con la cepa de referencia. La validación de los resultados se realizó por qPCR, para lo cual se diseñaron cebadores específicos para cuatro genes representativos (*HAP3, PUT1, SNF4, VMA5*). Tras este análisis se evaluó la capacidad de crecimiento de las diferentes cepas en un medio con prolina o arginina como fuente de nitrógeno.

Además se estudiaron los patrones cromosómicos de las cepas de *S. cerevisiae* mediante el análisis de cariotipos por electroforesis en campo pulsante, y se analizó su capacidad de esporulación y viabilidad de las esporas de cada una de las cepas, así como su homo/heterotalismo, mediante su crecimiento en un medio con acetato y posterior disección de las ascas utilizando un micromanipulador. También se analizó su ploidía por citometría de flujo, tiñendo el DNA con PI (yoduro de propidio) y comparando contra las señales obtenidas en las cepas *S. cerevisiae* de referencia S288c (haploide) y FY1679 (diploide).

2.2 Caracterización genómica de cepas de *S. uvarum*

Se volvió a utilizar la hibridación genómica comparada (aCGH) para determinar las diferencias a nivel genómico entre cuatro cepas de *S. uvarum* aisladas de mistela, fermentación vínica, exudado de árbol y jugo de grosella, confirmando así su relevancia durante los diferentes procesos fermentativos. Tras el análisis de aCGH, se estudió cómo afecta la dosis génica de algunos genes a nivel fisiológico. Para ello se evaluó la capacidad de crecimiento de las diferentes cepas en diversas fuentes de carbono no fermentables (etanol, acetato y glicerol) y en presencia/ausencia de zinc.

2.3 Análisis transcriptomico de cepas de *S. cerevisiae* mediante RNA-seq

Se realizó un análisis comparativo de la expresión génica de cepas de *S. cerevisiae* aisladas de varias fermentaciones tradicionales, mediante la tecnología RNA-seq. En este objetivo se intenta comprender las diferencias en la regulación de la transcripción entre las diferentes cepas durante la fermentación. Nuestro objetivo es vincular la regulación de la transcripción de la célula cuando está al comienzo de la fase exponencial de crecimiento, con las características enológicas y fisiológicas obtenidas en resultados anteriores de nuestro grupo.

Para ello se llevaron a cabo fermentaciones por triplicado a 22ºC en mosto sintético con 200g/l de azúcares en botellas estériles de 250ml. Se hicieron precultivos en GPY a 28ºC o/n, y se inoculó el mosto sintético con el volumen apropiado de éstos. La población inoculada en cada botella fue de 2 x 10^6 células/ml, para llegar a una población inicial de ufc/ml de 10^6. El pH inicial del medio fue de 3,3±0.1. Las fermentaciones se siguieron con el Monitor de Presión de Gas ANKOM y HPLC (Cromatografía líquida de alto rendimiento), hasta que la cantidad total de azúcar presente fue de 100gl/l, momento en el cual se tomaron las muestras (comienzo de la fase exponencial del crecimiento de las levaduras) para la posterior extracción de RNA. La extracción de RNA total se llevó a cabo por el método Fenol-Cloroformo, y la posterior purificación del RNAm se hizo utilizando el Kit Poly(A) Purist mRNA Purification Kit (Ambion), para eliminar selectivamente el RNA

ribosomal. Se determinó la cantidad de RNAm con el Nanodrop y el Bioanalizador Experion (Bio-Rad) y se llevó a secuenciar. Posteriormente, los datos obtenidos se filtraron para eliminar las lecturas de mala calidad y las lecturas de calidad se mapearon con el genoma de referencia *S. cerevisiae* S288c de la base de datos *Saccharomyces Sensu Stricto*. A continuación se calcularon los niveles de expresión y se llevó a cabo el análisis comparativo entre las diferentes cepas de *S. cerevisiae*. Para ello se utilizaron los programas bioinformáticos BWA (alineamiento y mapeo de las secuencias), Samtools (manipulación de los alineamientos en el formato SAM), Cufflinks y Cuffdiff (programas para ensamblar, estimar la abundancia de los transcritos y ver niveles de expresión), IGV (herramienta para visualizar las secuencias y la cobertura del mapeo) y Artemis (visualización de las características de las secuencias).

2.4 Análisis transcriptomico de diferentes especies del género *Saccharomyces* mediante RNA-seq

En el presente objetivo realizamos un estudio trancriptómico de levaduras con propiedades de interés de las especies *S. uvarum*, *S. paradoxus* y *S. kudriavzevii*, con la finalidad de detectar tanto diferencias de expresión en los genes compartidos (homólogos) como detectar la expresión de genes específicos de estas especies que no se encuentren en *S. cerevisiae*. Se siguió la misma metodología descrita en el objetivo anterior.

En el caso de estas especies, los genomas de referencia de la base de datos no están anotados por completo, queda mejorar el ensamblaje de los contigs, por lo que fue necesario reanotarlos realizando una búsqueda de ortólogos por BLAST, antes de alinearlos con nuestras secuencias.

A partir de la información obtenida, se pretende o bien llevar a cabo una selección de cepas que presenten propiedades de interés, u obtener híbridos artificiales entre cepas de *S. cerevisiae* y cepas de estas especies que contribuyan con nuevas propiedades de interés enológico y permitan resolver algunas de las exigencias actuales de las bodegas.

3. Conclusiones

En base a los resultados obtenidos, se ha llegado a las siguientes conclusiones:

3.1 Caracterización genómica de cepas de *S. cerevisiae*

El número de copias génicas es muy variable entre cepas y la mayoría de las diferencias se observan en genes subteloméricos e intracromosómicos implicados en funciones metabólicas relacionadas con la homeostasis celular, interacciones célula a célula, y el transporte de solutos tales como iones, azúcares y metales. En muchos casos, estos genes no son esenciales, pero pueden desempeñar un papel importante en la adaptación de las levaduras a las nuevas condiciones ambientales.

El metabolismo del nitrógeno tiene un papel importante en la evolución adaptativa de *S. cerevisiae*. Los cambios observados en el número de copias de los genes *PUT1* (uno de los principales genes implicados en la asimilación de prolina como fuente de nitrógeno) y *CAR2* (implicado en el catabolismo de la arginina) muestran una fuerte adaptación de las levaduras a fuentes de nitrógeno.

3.2 Caracterización genómica de cepas de *S. uvarum*

La mayor variabilidad genómica entre las cepas de *S. uvarum* se encuentra en las regiones subteloméricas y en sitios de inserción "Ty-element", al igual que ocurría en las cepas de *S. cerevisiae*, lo que sugiere que este tipo de variabilidad es la principal fuente de diversidad genética en las poblaciones de levadura. Pero además, otros procesos evolutivos que implican otras formas de variación genética, podrían ser responsables de las variantes genéticas que subyacen a la variación fenotípica en las levaduras.

Se encontraron diferencias en el número de copias de genes implicados en el metabolismo del zinc (*Vel1/YOR387C*). El zinc desempeña importantes funciones en el crecimiento y desarrollo de la levadura. Se observó que en ausencia de zinc, las cepas de *S. uvarum* veían disminuido su crecimiento, en comparación con el crecimiento en presencia de zinc, al contrario de lo que ocurría con las cepas de *S. cerevisiae*. Esto indicaría que las cepas

de *S. uvarum* muestran una mayor sensibilidad a la limitación de zinc que las cepas de *S. cerevisiae*.

Otros genes con número de copias variable fueron los relacionados con la utilización de fuentes de carbono no fermentables (*TDH2, TDH3, PDC1, PDC5, HAP3, HOT13*). Aunque las cepas mostraron diferencias en el crecimiento en fuentes de carbono no fermentables (etanol, glicerol y acetato), no se observó defecto de crecimiento en ninguna de ellas.

3.3 Análisis transcriptomico de cepas de *S. cerevisiae*

Los datos indican que la adaptación de las levaduras *S. cerevisiae* a los ambientes fermentativos eleva los niveles de manoproteínas a nivel de la pared celular. Las manoproteínas de la pared celular de *S. cerevisiae*, se han convertido en los últimos años en uno de los productos de mayor interés para la mejora de procesos tecnológicos y de las características sensoriales de los vinos. Además, se comercializan cepas vínicas de *S. cerevisiae* (Lalvin BM45®, comercializada por Lallemand), entre cuyas características destaca una abundante producción de manoproteínas durante la fermentación. A las manoproteínas se les atribuyen diversas propiedades en enología, entre las que destacan su capacidad de evitar o minimizar algunas alteraciones que pueden sufrir los vinos, y que afectan negativamente a su calidad, reduciendo su valor comercial. Por tanto, la selección de nuevas levaduras superproductoras de manoproteínas puede ser una estrategia interesante en el proceso de producción del vino para obtener vinos de mayor calidad.

El análisis de enriquecimiento de términos GO (Gene Ontology) reveló además genes relacionados con la síntesis de aromas. Observamos que la cepa Temohaya-26, aislada de jugo de agave, produce mayores concentraciones de alcoholes superiores y ésteres que la levadura vínica T73 y la levadura aislada de masato PE35M. Estos resultados hacen de este organismo un candidato serio para generar híbridos que tengan una mayor influencia en las propiedades sensoriales del vino, y demuestran que cepas de diferentes nichos ecológicos tienen diferencias en su composición génica.

3.4 Análisis transcriptomico de especies del género *Saccharomyces*

Los resultados obtenidos muestran sobre-expresión de genes relacionados con la esporulación (*IME1*), poniendo de manifiesto que la eficiencia en la esporulación puede ser un objetivo importante en la evolución adaptativa de las levaduras cuando se enfrentan a ambientes en los que las condiciones cambian constantemente.

Por desgracia, no hemos podido profundizar más en el análisis del transcriptoma de estas especies porque los genomas de referencia de las bases de datos no están anotados por completo y contienen errores y genes por duplicado que dificultan el análisis. Por lo tanto es probable que estemos perdiendo importantes diferencias entre ellas, a causa de este problema técnico. La secuenciación completa del genoma de más cepas de *S. kudriavzevii*, *S. uvarum* y *S. paradoxus* por nuestro grupo, y la mejora de los ya secuenciados, complementará la información preliminar proporcionada en esta tesis sobre el transcriptoma de las cepas pertenecientes a estas especies.

9. LIST OF TABLES & FIGURES

List of Figures

INTRODUCTION

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

List of Tables

MATERIAL & METHODS

Additional Files

9. REFERENCES

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

Adams J, Puskas-Rozsa S, Simlar J, Wilke CM. 1992. Adaptation and major chromosomal changes in populations of Saccharomyces cerevisiae. Current Genetics 22, 13-19.

Adams J, Puskas-Rozsa S, Simlar J, Wilke CM. Adaptation and major chromosomal changes in populations of Saccharomyces cerevisiae.. Curr Genet. 1992 Jul;22(1):13-9.

Akao T, Yashiro I, Hosoyama A, Kitagaki H, Horikawa H, Watanabe D, Akada R, Ando Y, Harashima S, Inoue T, Inoue Y, Kajiwara S, Kitamoto K, Kitamoto N, Kobayashi O, Kuhara S, Masubuchi T, Mizoguchi H, Nakao Y, Nakazato A, Namise M, Oba T, Ogata T, Ohta A, Sato M, Shibasaki S, Takatsume Y, Tanimoto S, Tsuboi H, Nishimura A, Yoda K, Ishikawa T, Iwashita K, Fujita N, Shimoi H. 2011. Whole-genome sequencing of sake yeast Saccharomyces cerevisiae Kyokai no. 7. DNA Res. 18(6):423-34.

Alba-Lois, L. & Segal-Kischinevzky, C. Beer & Wine Makers. Nature Education. 2010; 3(9):17.

Alexandre H, Ansanay-Galeote V, Dequin S, Blondin B. 2001. Global gene expression during short-term ethanol stress in *Saccharomyces cerevisiae*. FEBS Lett. 498:98–103.

Almeida P, Gonçalves C, Teixeira S, Libkind D, Bontrager M, Masneuf-Pomarède I, Albertin W, Durrens P, Sherman DJ, Marullo P, Hittinger CT, Gonçalves P, Sampaio JP. 2014. A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*. Nat Commun. Jun 2;5:4044.

Almeida P, Barbosa R, Zalar P, Imanishi Y, Shimizu K, Turchetti B, Legras JL, Serra M, Dequin S, Couloux A, Guy J, Bensasson D, Gonçalves P, Sampaio JP. 2015 . A Population Genomics Insight into the Mediterranean Origins of Wine Yeast Domestication. Mol Ecol. Aug 6.

Antunovics Z, Irinyi L, Sipiczki M. 2005. Combined application of methods to taxonomic identification of *Saccharomyces* strains in fermenting botrytized grape must. J Appl Microbiol. 98(4):971-9.

Argueso JL, Carazzolle MF, Mieczkowski PA, Duarte FM, Netto OV, Missawa SK, Galzerani F, Costa GGTavares FC, Alcarde AR, Dietrich FS, McCusker JH, Petes TD, Pereira GA. 2009. Genome structure of a Saccharomyces cerevisiae strain widely used in bioethanol production. Genome Res. 19(12):2258-70.

Arias A. 2008. Diversidad genética en las especies del complejo *Saccharomyces sensu stricto* de fermentaciones tradicionales. PhD Thesis. University of Valencia.

Arroyo-López FN, Durán Quintana MC, Garrido Fernández A. 2006. Use of the generalized z-value concept to study the effects of temperature, NaCl concentration and pH on *Pichia anomala*, a yeast related to table olive fermentation. International Journal of Food Microbiology 106, 45–51.

Arroyo-López FN, Orlic S, Querol A, Barrio E. 2009. Effects of temperature, pH and sugar concentration on the growth parameters of *Saccharomyces cerevisiae*, *S. kudriavzevii* and their interspecific hybrids. International Journal of Food Microbiology 131, 120-127.

Arroyo-López FN, Salvadó Z, Tronchoni J, Guillamón JM, Barrio E, Querol A., 2010. Susceptibility and resistance to ethanol in *Saccharomyces* strains isolated from wild and fermentative environments. Yeast. Dec; 27(12):1005-15.

Aslankoohi E, Zhu B, Rezaei MN, Voordeckers K, De Maeyer D, Marchal K, Dornez E, Courtin CM, Verstrepen KJ. 2013. Dynamics of the *Saccharomyces cerevisiae* transcriptome during bread dough fermentation. Appl Environ Microbiol. Dec;79(23):7325-33.

Babrzadeh F, Jalili R, Wang C, Shokralla S, Pierce S, Robinson-Mosher A, Nyren P, Shafer RW, Basso LC, de Amorim HV, de Oliveira AJ, Davis RW, Ronaghi M, Gharizadeh B, Stambuk BU. 2012. Whole-genome sequencing of the efficient industrial fuel-ethanol fermentative *Saccharomyces cerevisiae* strain CAT-1. Mol Genet Genomics. 287(6):485-94.

Backhus LE, DeRisi J, Bisson LF. 2001. Functional genomic analysis of a commercial wine strain of *Saccharomyces cerevisiae* under differing nitrogen conditions. FEMS Yeast Res. Jul;1(2):111-25. Erratum in: FEMS Yeast Res. Oct;4(1):123.

Bakalinsky AT & Snow R. 1990. The chromosomal constitution of wine strains of *Saccharomyces cerevisiae*. Yeast. 6(5):367-82.

Baker EC, Wang B, Bellora N, Peris D, Hulfachor A, Koshalek J, Adams M, Libkind D, Hittinger C (*In press*). The genome sequence of *Saccharomyces eubayanus* and the domestication of lager-brewing yeasts. Mol Biol and Evol. 2015 Aug 11. pii: msv168.

Bamforth C. (Ed.) 2005. Food, Fermentation and Microorganisms. Blackwell Publishing, Oxford.

Barre P, Vezinhet F, Dequin S, Blondin B. 1993. Genetic improvement of wine yeasts. Fleet GH. ed. *Wine microbiology and biotechnology*. Chur, Switzerland: Harwood Academic Publishers, pp 265-287.

Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Muertter RN, Edgar R., 2009.

NCBI GEO: archive for high-throughput functional genomic data. Nucleic Acids Research, Vol. 37, Database issue.

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

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

Beltran G, Novo M, Leberre V, Sokol S, Labourdette D, Guillamon JM, Mas A, François J, Rozes N. 2006. Integration of transcriptomic and metabolic analyses for understanding the global responses of low-temperature winemaking fermentations. FEMS Yeast Res. Dec;6(8):1167-83.

Benitez T, Martínez P & Codón AC. 1996. Genetic constitution of industrial yeast. Microbiologia. 12(3):371-84.

Bergström A, Simpson JT, Salinas F, Barré B, Parts L, Zia A, Nguyen Ba AN, Moses AM, Louis EJ, Mustonen V, Warringer J, Durbin R, Liti G. 2014. A high-definition view of functional genetic variation from natural yeast genomes. Mol Biol Evol. Apr;31(4):872-88.

Bertolini L, Zambonelli C, Giudici P & Castellari L. 1996. Higher alcohol production by cryotolerant *Saccharomyces* strains. Am J Enol Vitic 47: 343–345.

Bidenne C, Blondin B, Dequin S, Vezinhet F. 1992. Analysis of the chromosomal DNA polymorphism of wine strains of *Saccharomyces cerevisiae*. Curr Genet. Jul;22(1):1-7.

Bing J, Han PJ, Liu WQ, Wang QM, Bai FY. 2014. Evidence for a Far East Asian origin of lager beer yeast. Curr Biol. May 19;24(10):R380-1.

Bisson LF. 2004. Biotechnology of wine yeast. Food Biotechnol. 18, 63-96.

Blondin B, Dequin S, Querol A, Legras JL. 2009. Genome of *Saccharomyces cerevisiae* and related yeasts. In: König, H., Unden, G., Fröhlich, J. (Eds.), Biology of microorganisms on grapes, in must and in wine. Springer Verlag, Heidelberg, Germany, pp.361-378.

Bond U, Neal C, Donnelly D, James TC. 2004. Aneuploidy and copy number breakpoints in the genome of lager yeasts mapped by microarray hybridisation. Curr Genet. 45(6):360-70.

Borneman AR, Forgan AH, Pretorius IS, Chambers PJ. 2008. Comparative genome analysis of a *Saccharomyces cerevisiae* wine strain. FEMS Yeast Res. 8(7):1185-95.

Borneman AR, Desany BA, Riches D, Affourtit JP, Forgan AH, Pretorius IS, Egholm M, Chambers PJ. 2011. Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. PLoS Genet. Feb 3;7(2):e1001287.

Borneman AR, Desany BA, Riches D, Affourtit JP, Forgan AH, Pretorius IS, Egholm M, Chambers PJ. 2012. The genome sequence of the wine yeast VIN7 reveals an allotriploid hybrid genome with *Saccharomyces cerevisiae* and *Saccharomyces kudriavzevii* origins. FEMS Yeast Res. Feb;12(1):88-96.

Boynton PJ & Greig D. 2014. The ecology and evolution of non-domesticated *Saccharomyces* species. Yeast. Dec;31(12):449-62.

Bradbury JH & Holloway WD. 1988. Chemistry of tropical root crops: significance for nutrition and agriculture in the Pacific. Inkata Press Pty Ltd, Canberra, Australia, p. 201.

Byrne KP, Wolfe KH. 2005. The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. Genome Res. Oct;15(10):1456-61.

Caridi A & Corte V. 1997. Inhibition of malolactic fermentation by cryotolerant yeast. Biotechnol Lett 19:723-726.

Caridi A. 2006. Enological functions of parietal yeast mannoproteins. Antonie Van Leeuwenhoek. Apr-May;89(3-4):417-22. Epub 2006 Apr 19. Review.

Carle GF & Olson MV. 1985. An electrophoretic karyotype for yeast. Proceedings of the National Academy of Sciences of the United States of America 82, 3756-3760.

Carreto L, Eiriz MF, Gomes AC, Pereira PM, Schuller D, Santos MA. 2008. Comparative genomics of wild type yeast strains unveils important genome diversity. BMC Genomics. 4;9:524.

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

Castellari L, Ferruzzi M, Magrini A, Giudici P, Passarelli P & Zambonelli C. 1994. Unbalanced wine fermentation by cryotolerant vs. non-cryotolerant *Saccharomyces* strains. Vitis 33: 49–52.

Charoenchai C, Fleet G, Henschke PA. 1998. Effects of temperature, pH and sugar concentration on the growth rates and cell biomass of wine yeasts. American Journal of Enology and Viticulture 49, 283–288.

Charron MJ, Dubin RA, Michels CA. 1986. Structural and functional analysis of the *MAL1* locus of *Saccharomyces cerevisiae*. Molecular and Cellular Biology 6, 3891-3899.

Coloretti F, Zambonelli C, Tini V. Characterization of flocculent *Saccharomyces* interspecific hybrids for the production of sparkling wines. Food Microbiol. 2006. Oct;23(7):672-6.

Combina M, Pérez-Torrado R, Tronchoni J, Belloch C, Querol A. 2012. Genome-wide gene expression of a natural hybrid between *Saccharomyces cerevisiae* and *S. kudriavzevii* under enological conditions. Int J Food Microbiol. Jul 16;157(3):340-5.

Costa FF. Non-coding RNAs: lost in translation?. Gene. 2007 Jan 15;386(1-2):1-10. Epub 2006 Oct 10.

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

Cubillos FA, Billi E, Zörgö E, Parts L, Fargier P, Omholt S, Blomberg A, Warringer J, Louis EJ, Liti G. 2011. Assessing the complex architecture of polygenic traits in diverged yeast populations. Molecular Ecology 20, 1401-1413.

David L, Huber W, Granovskaia M, Toedling J, Palm CJ, et al. A high-resolution map of transcription in the yeast genome. Proc Natl Acad Sci USA. 2006; 103: 5320–5325.

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

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

Denoeud F, Aury JM, Da Silva C, Noel B, Rogier O, Delledonne M, Morgante M, Valle G, Wincker P, Scarpelli C, Jaillon O, Artiguenave F. Annotating genomes with massive-scale RNA sequencing. Genome Biol. 2008; 9(12):R175.

Dequin S. 2001. The potential of genetic engineering for improving brewing, wine-making and baking yeasts. Appl Microbiol Biotechnol. 56(5-6):577-88.

Dequin S & Casaregola S. 2011. The genomes of fermentative *Saccharomyces*. C R Biol. Aug-Sep; 334(8-9):687-93.

Deregowska A, Skoneczny M, Adamczyk J, Kwiatkowska A, Rawska E, Skoneczna A, Lewinska A, Wnuk M. 2015. Genome-wide array-CGH analysis reveals YRF1 gene copy

number variation that modulates genetic stability in distillery yeasts. Oncotarget. Oct 13;6(31):30650-63.

Dietrich FS, Voegeli S, Brachat S, Lerch A, Gates K, Steiner S, Mohr C, Pöhlmann R, Luedi P, Choi S, Wing RA, Flavier A, Gaffney TD, Philippsen P. 2004. The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. Science. Apr 9;304(5668):304-7.

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

Dunham MJ, Badrane H, Ferea T, Adams J, Brown PO, Rosenzweig F, Botstein D. 2002. Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A. 99(25):16144-9.

Dunn B, Levine RP, Sherlock G. 2005. Microarray karyotyping of commercial wine yeast strains reveals shared, as well as unique, genomic signatures. BMC Genomics. 16;6:53.

Dunn B, Richter C, Kvitek DJ, Pugh T, Sherlock G. 2012. Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. Genome Research 22, 908-924.

Dupin IVS, McKinnon BM, Ryan C, Boulay M, Markides AJ, Jones GP, Williams PJ, & Waters EJ. *Saccharomyces cerevisiae* mannoproteins that protect wine from protein haze: Their release during fermentation and lees contact and a proposal for their mechanism of action. Journal of Agricultural and Food Chemistry. 2000; 48, 3098–3105.

Eddy SR. Non-coding RNA genes and the modern RNA world. Nat Rev Genet. 2001 Dec;2(12):919-29. Review.

Edwards-Ingram, L.C., Gent, M.E., Hoyle, D.C., Hayes, A., Stateva, L.I., Oliver, S.G., 2004. Comparative Genomic Hybridization Provides New Insights Into the Molecular Taxonomy of the *Saccharomyces Sensu Stricto* Complex. Genome Research 14, 1043-1051.

Eglinton JM, McWilliam SJ, Fogarty MW, Francis IL, Kwiatkowski MJ, Høj PB, Henschke PA 2000. The effect of *Saccharomyces bayanus*-mediated fermentation on the chemical composition and aroma profile of Chardonnay wine. Aust. J. Grape Wine Res. 6: 190–196.

Eisen MB, Brown PO. 1999. DNA arrays for analysis of gene expression. Methods Enzymol. 303:179-205.

Engel SR, Dietrich FS, Fisk DG, Binkley G, Balakrishnan R, Costanzo MC, Dwight SS, Hitz BC, Karra K, Nash RS, Weng S, Wong ED, Lloyd P, Skrzypek MS, Miyasato SR, Simison

M, Cherry JM. 2014. The reference genome sequence of *Saccharomyces cerevisiae*: then and now. G3 (Bethesda). Mar 20;4(3):389-98.

Erasmus DJ, van der Merwe GK, van Vuuren HJ., 2003. Genome-wide expression analyses: Metabolic adaptation of *Saccharomyces cerevisiae* to high sugar stress. FEMS Yeast Res. 3:375-99.

Fay JC & Benavides JA. 2005. Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. PLOS Genetics 1, 66-71.

Feuillat M, Charpentier C, Nguyen T, van Long. 1998. Yeast's mannoproteins: a new possible enological adjuvant. Bulletin de l'OIV, 71; pp. 944–967.

Franco-Duarte R, Mendes I, Umek L, Drumonde-Neves J, Zupan B, Schuller D. 2014. Computational models reveal genotype-phenotype associations in *Saccharomyces cerevisiae*. Yeast. Jul; 31(7):265-77.

Gamero A, Hernández-Orte P, Querol A, Ferreira V. 2011. Effect of aromatic precursor addition to wine fermentations carried out with different *Saccharomyces* species and their hybrids. Int J Food Microbiol. May 14;147(1):33-44.

Gamero A, Tronchoni J, Querol A, Belloch C. 2013. Production of aroma compounds by cryotolerant *Saccharomyces* species and hybrids at low and moderate fermentation temperatures. J Appl Microbiol. May;114(5):1405-14.

Gamero A, Belloch C, Ibáñez C, Querol A. 2014. Molecular analysis of the genes involved in aroma synthesis in the species *S. cerevisiae*, *S. kudriavzevii* and *S. bayanus* var. *uvarum* in winemaking conditions. PLoS One. May 22;9(5):e97626.

Gangl H, Batusic M, Tscheik G, Tiefenbrunner W, Hack C, Lopandic K. 2009. Exceptional fermentation characteristics of natural hybrids from *Saccharomyces cerevisiae* and *S. kudriavzevii*. N Biotechnol. Apr;25(4):244-51.

Garber M, Grabherr MG, Guttman M, Trapnell C. 2011. Computational methods for transcriptome annotation and quantification using RNA-seq. Nat Methods. Jun;8(6):469-77.

Gerbaud V, Gabas N, Laguerie C, Blouin J, Vidal S, Moutounet M, & Pellerin P. 1996. Effect of wine polysaccharides on the nucleation of potassium hydrogen tartrate in model solutions. Chemical Engineering Research & Design. 74(A7), 782–790.

Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G,
Botstein D, Brown PO. 2000. Genomic expression programs in the response of yeast cells to environmental changes. Mol. Biol. Cell 11:4241.

Gibson B, Liti G. 2015. *Saccharomyces pastorianus*: genomic insights inspiring innovation for industry. Yeast. Jan;32(1):17-27.

Gitan RS, Luo H, Rodgers J, Broderius M, Eide D. 1998. Zinc-induced inactivation of the yeast ZRT1 zinc transporter occurs through endocytosis and vacuolar degradation. J Biol Chem. Oct 30;273(44):28617-24.

Giudici P, Zambonelli C, Passarelli P & Castellari L. 1995. Improvement of wine composition with cryotolerant *Saccharomyces* strains. *American Journal of Enology and Viticulture*, *46,* 143-147.

Glushakova AM, Ivannikova Iu.V, Naumova ES, Chernov I.Iu. & Naumov GI. 2007. Massive isolation and identification of *Saccharomyces paradoxus* yeasts from plant phyllosphere. Mikrobiologiia 76: 236-242. In Russian.

Goddard MR. 2008. Quantifying the complexities of *Saccharomyces cerevisiae*'s ecosystem engineering via fermentation. Ecology. Aug;89(8):2077-82.

Goff L & Trapnell C. 2011. cummeRbund: analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data. R package version 1.2.0.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG. 1996. Life with 6000 genes. Science. Oct 25;274(5287):546, 563-7. Review.

Goldemberg, J., 2007. Ethanol for a sustainable energy future. Science 315:808-810.

Gómez-Pastor R, Pérez-Torrado R, Cabiscol E, Ros J, Matallana E. 2010. Reduction of oxidative cellular damage by overexpression of the thioredoxin *TRX2* gene improves yield and quality of wine yeast dry active biomass. Microbial Cell Factories 9, 9.

Gonzalez-Ramos, D., & Gonzalez, R. 2006. Genetic determinants of the release of mannoproteins of enological interest by *Saccharomyces cerevisiae*. Journal of Agricultural and Food Chemistry, 54(25), 9411–9416.

Gonzalez-Ramos, D., Cebollero, E., & Gonzalez, R. 2008. A recombinant *Saccharomyces cerevisiae* strain overproducing mannoproteins stabilizes wine against protein haze. Applied and Environmental Microbiology, 74(17), 5533–5540.

Gonzalez-Ramos, D., Quirós, M., & Gonzalez, R. 2009. Three different targets for the genetic modification of wine yeast strains resulting in improved effectiveness of bentonite fining. Journal of Agricultural and Food Chemistry, 57(18), 8373–8378.

González SS, Barrio E, Gafner J & Querol A. 2006. Natural hybrids from *Saccharomyces cerevisiae*, *Saccharomyces bayanus* and *Saccharomyces kudriavzevii* in wine fermentations. FEMS Yeast Res. 6(8):1221-34.

González SS, Gallo L, Climent MA, Barrio E, Querol A. Enological characterization of natural hybrids from *Saccharomyces cerevisiae* and *S. kudriavzevii*. 2007. Int J Food Microbiol. May 1;116(1):11-8.

González SS, Barrio E & Querol A. 2008. Molecular characterization of new natural hybrids between *S. cerevisiae* and *S. kudriavzevii* from brewing. Appl Environ Microb 74: 2314-2320.

González SS, Gallo L, Climent MA, Barrio E & Querol A. 2007. Enological characterization of natural hybrids from *Saccharomyces cerevisiae* and *S. kudriavzevii*. Int J Food Microbiol. 116(1):11-8.

Guilloux-Benatier M, & Chassagne D. 2003. Comparison of components released by fermented or active dried yeasts after aging on lees in a model wine. Journal of Agricultural and Food Chemistry, 51(3), 746–751.

Hauser NC, Fellenberg K, Gil R, Bastuck S, Hoheisel JD, Pérez-Ortín JE. 2001. Whole genome analysis of wine yeast strain. Comp.Funct.Genom. 2, 69-79.

Hittinger CT, Gonçalves P, Sampaio JP, Dover J, Johnston M, Rokas A. 2010. Remarkably ancient balanced polymorphisms in a multi-locus gene network. Nature. Mar 4;464(7285):54-8.

Hittinger CT. 2013. *Saccharomyces* diversity and evolution: a budding model genus. Trends in Genetics 29, 309-317.

Hodgins-Davis A, Townsend JP. 2009. Evolving gene expression: from G to E to GxE. Trends Ecol Evol. Dec;24(12):649-58.

Hodgins-Davis A, Adomas AB, Warringer J, Townsend JP. 2012. Abundant gene-by environment interactions in gene expression reaction norms to copper within *Saccharomyce scerevisiae*. Genome Biol Evol. 4(11):1061-79.

Holloway AJ, van Laar RK, Tothill RW, Bowtell DD. 2002. Options available--from start to finish--for obtaining data from DNA microarrays II. Nat Genet. Dec;32 Suppl:481-9. Review.

Horinouchi T, Yoshikawa K, Kawaide R, Furusawa C, Nakao Y, Hirasawa T, Shimizu H. 2010. Genome-wide expression analysis of *Saccharomyces pastorianus* orthologous genes using oligonucleotide microarrays. J Biosci Bioeng. Nov;110(5):602-7.

Hui Y H, Meunier-Goddik L, Solvejg Hansden A, Josephson J, Nip W-K, Stanfield PS, & Toldrá F (Eds). 2004. Handbook of Food Fermentation and Beverage Fermentation Technology. Marcel-Dekker, New York.
Hutkins RW. 2006. Microbiology and Technology of Fermented Foods. Blackwell Publishing, Oxford.

Huxley C, Green ED, Dunham I. 1990. Rapid assessment of *S. cerevisiae* mating type by PCR. Trends in Genetics 6, 236.

Ibáñez C, Pérez-Torrado R, Chiva R, Guillamón JM, Barrio E, Querol A. 2014. Comparative genomic analysis of *Saccharomyces cerevisiae* yeasts isolated from fermentaions of traditional beverages unveils different adaptive strategies. Int J Food Microbiol. Feb 3;171:129-35.

Kellis M, Birren BW, Lander ES. . 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature. Apr 8;428(6983):617-24.

Langkjaer RB, Cliften PF, Johnston M, Piskur J. 2003. Yeast genome duplication was followed by asynchronous differentiation of duplicated genes. Nature. Feb 20;421(6925):848-52.

Lemura Y, Yamada T, Takahashi T, Furukawa K, Hara S. 1999. Influence of amino acid content in seed mash on peptide uptake by yeast cells in main mash in sake brewing process. Journal of Bioscience and Bioengineering 88, 679-681.

Lopes CA, Barrio E, Querol A. 2010. Natural hybrids of *S. cerevisiae* x *S. kudriavzevii* share alleles with European wild populations of *Saccharomyces kudriavzevii*. FEMS Yeast Res. Jun;10(4):412-21.

Infante JJ, Dombek KM, Rebordinos L, Cantoral JM, Young ET. 2003. Genome-wide amplifications caused by chromosomal rearrangements play a major role in the adaptive evolution of natural yeast. Genetics 165, 1745-1759.

Ito T, Miura F, Onda M. Unexpected complexity of the budding yeast transcriptome. Iubmb Life. 2008; 60: 775–781.

Jackson CL & Hartwell LH. 1990. Courtship in *S. cerevisiae*: both cell types choose mating partners by responding to the strongest pheromone signal. Cell. 63(5):1039-51.

Kavanaugh LA & Dietrich FS. Non-coding RNA prediction and verification in *Saccharomyces cerevisiae*. PLoS Genet. 2009 Jan;5(1):e1000321.

Kassir Y, Granot D, Simchen G. 1988. IME1, a positive regulator gene of meiosis in *S. cerevisiae*. Cell. Mar 25;52(6):853-62.

Kassir Y, Adir N, Boger-Nadjar E, Raviv NG, Rubin-Bejerano I, Sagee S, Shenhar G. 2003 Transcriptional regulation of meiosis in budding yeast. Int Rev Cytol 224: 111–171.

Kishimoto M & Goto S. 1995. Growth temperatures and electrophoretic karyotyping as tools for practical discrimination of *Saccharomyces bayanus* and *Saccharomyces cerevisiae*. J. Gen.Appl.Microbiol.41, 239-247.

Kupiec M, Byers B, Esposito RE. 1997. Meiosis and sporulation in *Saccharomyces cerevisiae*. In The molecular and cell biology of the yeast *Saccharomyces*. Pringle JR, Broach

JR, Jones EW eds, pp 889–1036. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Kurtzman CP & Fell JW. 1998. *The Yeasts*, A taxonomic study. 4th ed. Elsevier Science, Amsterdam.

Kurtzman CP, Fell JW & Boekhout T. 2011. *The yeast: A Taxonomic study*. 5th ed. Elsevier, Amsterdam.

Lachance MA, Bowles JM, Starmer WT., 2003. Geography and niche occupancy as determinants of yeast biodiversity: the yeast-insect-morning glory ecosystem of *Kīpuka Puaulu*, Hawaii. FEMS Yeast Res. Oct;4(1):105-11.

Lambrechts MG, Pretorius IS. 2000. Yeast and its importance to wine aroma. South African Journal of Enology and Viticulture. 21,97-129.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10:R25.

Langmead B. 2010. Aligning short sequencing reads with Bowtie. Curr Protoc Bioinformatics. Dec;Chapter 11:Unit 11.7.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods. Mar 4;9(4):357-9.

Lashkari DA, McCusker JH, Davis RW. 1997. Whole genome analysis: experimental access to all genome sequenced segments through larger-scale efficient oligonucleotide synthesis and PCR. Proc Natl Acad Sci U S A. Aug 19;94(17):8945-7.

Lassmann T, Hayashizaki Y, Daub CO. 2011. SAMStat: monitoring biases in next generation sequencing data. Bioinformatics. Jan 1; 27(1):130-1.

Legras JL, Merdinoglu D, Cornuet JM, Karst F. 2007. Bread, Beer and Wine: *Saccharomyces cerevisiae* diversity reflects human history. Molecular Ecology 16, 2091-2102.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R & 1000 Genome Project Data Processing Subgroup. 2009. The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics, 25, 2078-9.

Libkind D, Hittinger CT, Valérico E, Gonçalves C, Dover J, *et al*., 2011. Microbe domestication and the identification of the wild genetic stock of larger-brewing yeast. PNAS 108: 14539-14544.

Liti G, Barton DB, Louis EJ. 2006. Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. Genetics. Oct;174(2):839-50.

Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V, Tsai IJ, Bergman CM, Bensasson D, O'Kelly MJT, van Oudenaarden A, Barton DBH, Bailes E, Nguyen AN, Jones M, Quail MA, Goodhead I, Sims S, Smith F, Blomberg A, Durbin R, Louis EJ. 2009. Population genomics of domestic and wild yeasts. Nature 458, 337-341.

Liti G. & Louis E.J., 2012. Advances in quantitative trait analysis in yeast. PLoS Genet 8, e1002912.

Liti G, Nguyen Ba AN, Blythe M, Müller CA, Bergström A, Cubillos FA, Dafhnis-Calas F, Khoshraftar S, Malla S, Mehta N, Siow CC, Warringer J, Moses AM, Louis EJ, Nieduszynski CA. 2013. High quality de novo sequencing and assembly of the *Saccharomyces arboricolus* genome. BMC Genomics. Jan 31;14:69.

Le Jeune C, Lollier M, Demuyter CC, Erny Legras J-L, Aigle M & Masneuf-Pomarède I. 2007. Characterization of natural hybrids of *Saccharomyces cerevisiae* and *Saccharomyces bayanus* var. *uvarum*. FEMS Yeast Res. 7:540-549.

Lopandic K, Gangl H, Wallner E, Tscheik G, Leitner G, Querol A, Borth N, Breitenbach M, Prillinger H & Tiefenbrunner W. 2007. Genetically different wine yeasts isolated from Austrian vine-growing regions influence wine aroma differently and contain putative hybrids between *Saccharomyces cerevisiae* and *Saccharomyces kudriavzevii*. FEMSYeaast Res 7:953-965.

Lopes C A, Barrio E & Querol A. 2010. Natural hybrids of *S. cerevisiae* x *S. kudriavzevii* share alleles with European wild populations of *Saccharomyces kudriavzevii*. FEMS Yeast Research 4:412-421.

López S, Prieto M, Dijkstra J, Dhanoa MS, France J. 2004. Statistical evaluation of mathematical models for microbial growth. Int J Food Microbiol 96:289-300.

López-Malo M, García-Rios E, Melgar B, Sanchez MR, Dunham MJ, Guillamón JM. 2015. Evolutionary engineering of a wine yeast strain revealed a key role of inositol and mannoprotein metabolism during low-temperature fermentation. BMC Genomics. Jul 22;16:537.

Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K, Kulkarni S, Hartl DL, Thomas WK. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. Proceedings of the National Academy of Sciences 105, 9272-9277.

Llauradó JM, Rozès N, Bobet R, Mas A, Constantí M. 2002. Low temperatura alcoholic fermtentation in high sugar concentration grape must. Journal of Food Science, 67, 268-273.

Llauradó JM, Rozès N, Constantí M, Mas A. 2005. Study of some *Saccharomyces cerevisiae* strains for winemaking after preadaptation at low temperatures. Journal of Agricultural and Food Chemistry, 4, 1003-1011.

Llopis S, Querol A, Heyken A, Hube B, Jespersen L, Fernández-Espinar MT, Pérez-Torrado R. 2012. Transcriptomics in human blood incubation reveals the importance of oxidative stress response in *Saccharomyces cerevisiae* clinical strains. BMC Genomics. Aug 23;13:419.

Marguerat S & Bähler J. 2010. RNA-seq: from technology to biology. Cell Mol Life Sci. Feb;67(4):569-79.

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. Sep;18(9):1509-17.

Marks VD, Ho Sui SJ, Erasmus DJ, van der Merwe GK, Brumm J, Wasserman WW, Bryan J, van Vuuren HJ. 2008. Dynamics of the yeast transcriptome during wine fermentation reveals a novel fermentation stress response. FEMS Yeast Res. Feb;8(1):35-52.

Masneuf I, Hansen J, Groth C, Piskur J, Dubourdieu D. 1998. New hybrids between *Saccharomyces* sensu stricto yeast species found among wine and cider production strains. Appl Environ Microbiol. Oct;64(10):3887-92.

Masneuf I, Aigle M & Dubourdieu D. 1996. Development of a polymerase chain reaction/restriction fragment length polymorphism method for *Saccharomyces cerevisiae* and *Saccharomyces bayanus* identification in enology. FEMS Microbiology Letters 138, 239–244.

Mattick JS, Makunin IV. Non-coding RNA. Hum Mol Genet. 2006; Apr 15;15 Spec No 1:R17-29. Review.

McCutcheon JP & Eddy SR., 2003. Computational identification of non-coding RNAs in *Saccharomyces cerevisiae* by comparative genomics. Nucleic Acids Res. Jul 15;31(14):4119-28.

Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. Nat Rev Genet. 2009 Mar;10(3):155-9.

Michel-Cuello C, Juárez-Flores BI, Aguirre-Rivera JR, Pinos-Rodríguez JM. Quantitative characterization of nonstructural carbohydrates of mezcal Agave (Agave salmiana Otto ex Salm-Dick). J Agric Food Chem. 2008 Jul 23;56(14):5753-7.

Miura F, Kawaguchi N, Sese J, Toyoda A, Hattori M, et al. A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. Proc Natl Acad Sci USA 2006, 103: 17846–17851.

Moorthie S, Mattocks CJ, Wright CF. 2011. Review of massively parallel DNA sequencing technologies. Hugo J. Dec;5(1-4):1-12.

Morozova O, Hirst M, Marra MA. 2009. Applications of new sequencing technologies for transcriptome analysis. Annu Rev Genomics Hum Genet. 10:135-51.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. Jul;5(7):621-8.

Mortimer RK, Romano P, Suzzi G, Polsinelli M. 1994. Genome renewal: a new phenomenon revealed from a genetic study of 43 strains of *Saccharomyces cerevisiae* derived from natural fermentation of grape musts. Yeast. Dec;10(12):1543-52.

Mortimer R, Polsinelli M. 1999. On the origins of wine yeast. Res Microbiol. Apr;150(3):199-204.

Mortimer, R. K., 2000. Evolution and variation of the yeast (*Saccharomyces*) genome. Genome Res. 10(4): 403–409.

Mountounet M. 1969. Biosynthèse des alcools supérieurs des boissons fermentées. Ann. Technol. Agric. 18, 249-261.

Mustacchi R, Hohmann S & Nielsen J. 2006. Yeast systems biology to unravel the network of life. Yeast 23:227-238.

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science. Jun 6;320(5881):1344-9.

Naumov GI. 1987. Genetic basis for classification and identification of the ascomycetus yeasts. Stud Mycol 30:469-475.

Naumov GI & Nikonenko TA. 1989. Occurrence and physiological characteristics of biological species *Saccharomyces bayanus* from hybridological analysis. Microbiology (Engl. Transl.), 57: 526–530.

Naumov GI. 1996. Genetic identification of biological species in the *Saccharomyces* sensu stricto complex. J. Ind. Microbiol. 17, 295–302.

Naumov GI, Naumova ES & Sniegowski PD. 1997. Differentiation of European an Far East Asian populations of *Saccharomyces paradoxus* by allozyme analysis. Int. J. Syst. Bacteriol. 47: 341-344.

Naumov GI, Masneuf I, Naumova ES, Aigle M, & Dubourdieu D. 2000. Association of *Saccharomyces bayanus var. uvarum* with some French wines: genetic analysis of yeast populations. *Research in Microbiol*ogy*, 151,* 683-691.

Naumov GI, James SA, Naumova ES, Louis EJ & Roberts IN. 2000a. Three new species in the *Saccharomyces sensu stricto* complex: *Saccharomyces cariocanus*, *Saccharomyces kudriavzevii* and *Saccharomyces mikatae*. Int J Syst Evol Micr 50:1931-1942.

Naumov GI, Nguyen HV, Naumova ES, Michel A, Aigle M, Gaillardin C. 2001. Genetic identification of *Saccharomyces bayanus* var. *uvarum*, a cider-fermenting yeast.. Int J Food Microbiol. May 10;65(3):163-71.

Naumov GI, Naumova ES, Antunovics Z, Sipiczki M. 2002. *Saccharomyces bayanus* var. *uvarum* in Tokaj wine-making of Slovakia and Hungary. Appl Microbiol Biotechnol. Sep;59(6):727-30.

Naumov GI, Naumova ES & Masneuf-Pomarède I. 2010. Genetic identification of new biological species *Saccharomyces arboricolus*. Wang et Bai. A van Leeuw 98:1-7.

Naumova ES, Naumov GI, Barrio E, Querol A. 2010. Mitochondrial DNA polymorphism of the yeast *Saccharomyces bayanus* var. *uvarum*. Mikrobiologiia. Jul Aug;79(4):543-50.

Nguyen H-V & Gaillardin C. 2005. Evolutionary relationships between the former species *Saccharomyces uvarum* and the hybrids *Saccharomyces bayanus* and *Saccharomyces pastorianus*; reinstatement of *Saccharomyces uvarum* (Beijerinck) as a distint species. FEMS Yeast Res. 5: 471-483.

Nguyen H-V, Lepingle A & Gaillardin CA. 2000. Molecular typing demonstrates homogeneity of *Saccharomyces uvarum* strains and reveals the existence of hybrids between *S. uvarum* and *S. cerevisiae*, including the *S. bayanus* type strin CBS 380. Syst. Appl. Microbiol. 23:71-85

Ni T, Tu K, Wang Z, Song S, Wu H, Xie B, Scott KC, Grewal SI, Gao Y, Zhu J. 2010. The prevalence and regulation of antisense transcripts in *Schizosaccharomyces pombe*. PLoS One. Dec 20;5(12):e15271.

Nookaew I, Papini M, Pornputtapong N, Scalcinati G, Fagerberg L, Uhlén M, Nielsen J. 2012. A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in *Saccharomyces cerevisiae*. Nucleic Acids Res. Nov 1;40(20):10084-97.

Novo MT, Beltran G, Torija MJ, Poblet M, Rozès N, Guillamón JM, Mas A. 2003. Changes in wine yeast storage carbohydrate levels during preadaptation, rehydration and low temperature fermentations. Int J Food Microbiol. Sep 1;86(1-2):153-61.

Novo M, Bigey F, Beyne E, Galeote V, Gavory F, Mallet S, Cambon B, Legras JL, Wincker P, Casaregola S, Dequin S. 2009. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. Proc Natl Acad Sci U S A. Sep 22;106(38):16333-8.

Oliveira BM, Barrio E, Querol A, Pérez-Torrado R. 2014. Enhanced enzymatic activity of glycerol-3-phosphate dehydrogenase from the cryophilic *Saccharomyces kudriavzevii*. PLoS One. Jan 30;9(1):e87290.

Oliver SG. 2007. Introduction to functional analysis in yeast. Methods Microbiol. 36: 1-21.

Orlic S, Redzepovic S, Jeromel A, Herjavec, S and Iacumin L. 2007. Influence of indigenous *Saccharomyces paradoxus* strains on Chardonnay wine fermentation aroma. Int J Food Sci Technol 42, 95–101.

Orlić S, Arroyo-López FN, Huić-Babić K, Lucilla I, Querol A, Barrio E. 2010. A comparative study of the wine fermentation performance of *Saccharomyces paradoxus* under different nitrogen concentrations and glucose/fructose ratios. J Appl Microbiol. Jan;108(1):73-80.

Ozsolak F, Milos PM. 2011. RNA sequencing: advances, challenges and opportunities. Nat Rev Genet. Feb;12(2):87-98.

Phaff HJ, Miller MW & Shifrine M. 1956. The taxonomy of yeasts isolated from *Drosophila* in the Yosemite region of California. Antonie van Leewenhoek 22: 145-161.

Pérez-Ortín JE, García-Martínez J, Alberola TM. 2002a. DNA chips for yeast biotechnology. The case of wine yeasts. Journal of Biotechnology 98, 227-241.

Pérez-Ortín JE, Querol A, Puig S, Barrio E. 2002b. Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains. Genome Research 12, 1533-1539.

Pérez-Torrado R, Llopis S, Perrone B, Gómez-Pastor R, Hube B, Querol A. 2015. Comparative genomic analysis reveals a critical role of de novo nucleotide biosynthesis for *Saccharomyces cerevisiae* virulence. PLoS One. Mar 27;10(3):e0122382

Pérez-Través L, Lopes CA, Querol A, Barrio E. 2014. On the complexity of the *Saccharomyces bayanus* taxon: hybridization and potential hybrid speciation. PLoS One. Apr 4;9(4):e93729.

Pérez-Través L, Lopes CA, González R, Barrio E, Querol A. 2015. Physiological and genomic characterisation of *Saccharomyces cerevisiae* hybrids with improved fermentation performance and mannoprotein release capacity. Int J Food Microbiol. Jul 16;205:30-40.

Peris D, Belloch C, Lopandic K, Álvarez-Pérez JM, Querol A & Barrio E. 2011. The molecular characterization of new types of *S. cerevisiae* x *S. kudriavzevii* hybrid yeasts unveils a high genetic diversity. Yeast 29:81-91.

Peris D, Lopes CA, Arias A & Barrio E. 2012a. Reconstruction of the evolutionary history of *Saccharomyces cerevisiae* x *S. kudriavzevii* hybrids based on multilocus sequence analysis. PLoS One. 7(9):e45527.

Peris D, Lopes CA, Belloch C, Querol A & Barrio E. 2012b. Comparative genomics among *Saccharomyces cerevisiae* x *Saccharomyces kudriavzevii* natural hybrid strains isolated from wine and beer reveals different origins. BMC Genomics 20;13:407.

Peris D, Sylvester K, Libkind D, Gonçalves P, Sampaio JP, Alexander WG, Hittinger CT. 2014. Population structure and reticulate evolution of *Saccharomyces eubayanus* and its lager-brewing hybrids. Mol Ecol. Apr;23(8):2031-45.

Phaff HJ & knapp EP. 1956. The taxonomy of yeasts found in exudates of certain trees and other natural breeding sites of some species of *Drosophila*. Antonie Van Leeuwenhoek. 22(2):117-30.

Phaff HJ, Miller MW & Mrak EM. 1978. *The life of Yeast*. 2nd ed. (Press HU, ed.). Harvard University Press, Cambridge.

Piskur J, Langkjaer RB. Yeast genome sequencing: the power of comparative genomics. Mol Microbiol. 2004 Jul;53(2):381-9. Review.

Pizarro FJ, Jewett MC, Nielsen J, Agosin E. 2008. Growth temperature exerts differential physiological and transcriptional responses in laboratory and wine strains of *Saccharomyces cerevisiae*. Appl Environ Microbiol. Oct;74(20):6358-68.

Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO. 1999. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nature Genetics 23, 41-46.

Porro D, Sauer M, Branduardi P, Mattanovich D. 2005. Recombinant protein production in yeasts. Mol Biotechnol. Nov;31(3):245-59. Review.

Prescott SC & Dunn CG. 1940. Industrial Microbiology, 1st edn. McGraw Hill, New York.

Puig S, Querol A, Barrio E, Pérez-Ortín JE. 2000. Mitotic recombination and genetic changes in *Saccharomyces cerevisiae* during wine fermentation. Appl Environ Microbiol. May;66(5):2057-61.

Pulvirenti A, Nguyen H, Caggia C, Giudici P, Rainieri S, Zambonelli C. 2000. *Saccharomyces uvarum*, a proper species within *Saccharomyces sensu stricto*. FEMS Microbiol Lett. 192(2):191-6.

Presser A, Elowitz MB, Kellis M, Kishony R. 2008. The evolutionary dynamics of the *Saccharomyces cerevisiae* protein interaction network after duplication. Proc Natl Acad Sci U S A. Jan 22;105(3):950-4.

Pretorius IS, Van Der Westhuizen TI & Augustyn OPH. 1999. Yeast Biodiversity In Vineyards And Wineries And Its Importance To The South African Wine Industry. S. Afi: 1. Enol. Vitic. 20, 61-74.

Pretorius IS. 2000. Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking. Yeast 16, 675-729.

Pulvirenti A, Nguyen HV, Caggia C, Giudici P, Rainieri S & Zambonelli C. 2000. *Saccharomyces uvarum*, a proper species within *Saccharomyces sensu stricto*. *FEMS Microbiology Letters, 192,* 191-196.

Querol A, Barrio E, Ramón D. 1992a. A comparative study of different methods of yeast-strain characterization. Systematic and Applied Microbiology 15, 439-446.

Querol A, Huerta T, Barrio E, Ramón D. 1992b. Dry yeast-strain for use in fermentation of Alicante wines - Selection and DNA patterns. Journal of Food Science 57, 183-&.

Querol A, Barrio E, Huerta T, Ramón D. Molecular monitoring of wine fermentations conducted by active dry yeast strains. Appl Environ Microbiol. 1992 Sep;58(9):2948-53.

Querol A, Barrio E, Ramón D. Population dynamics of natural *Saccharomyces* strains during wine fermentation. Int J Food Microbiol. 1994 Mar; 21(4):315-23.

Querol A, Belloch C, Fernández-Espinar MT, Barrio E. Molecular evolution in yeast of biotechnological interest. Int Microbiol. 2003 Sep; 6(3):201-5.

Querol A & Fleet GH. 2006. Yeast in Food and Beverages. Springer-Verlag, Berlin.

Quirós M, Gonzalez-Ramos D, Tabera L, Gonzalez R. 2010. A new methodology to obtain wine yeast strains overproducing mannoproteins. Int J Food Microbiol. Apr 30;139(1-2):9-14.

Quirós M, Gonzalez R, Morales P. 2012. A simple method for total quantification of mannoprotein content in real wine samples. Food Chem. Sep 15;134(2):1205-10.

Rachidi N, Barre P, Blondin B. 1999. Multiple Ty-mediated chromosomal translocations lead to karyotype changes in a wine strain of *Saccharomyces cerevisiae*. Mol Gen Genet. Jun;261(4-5):841-50.

Rainieri S, Zambonelli C, Tini V, Castellari L & Giudici P. 1998. The enological traits of thermotolerant *Saccharomyces* strains. *American of Enology and Viticulture, 49,* 319-324.

Rainieri S, Zambonelli C, Hallsworth JE, Pulvirenti A, Giudici P., 1999. *Saccharomyces uvarum*, a distinct group within *Saccharomyces* sensu stricto. FEMS Microbiol Lett. Aug 1;177(1):177-85.

Rainieri S, Kodama Y, Kaneko Y, Mikata K, Nakao Y & Ashikari T. 2006. Pure and mixed genetic lines of *Saccharomyces bayanus* and *Saccharomyces pastorianus* and their contribution to the larger brewing strain genome. Appl Environ Microb 72: 3968-3974.

Ramirez-Córdova J, Drnevich J, Madrigal-Pulido JA, Arrizon J, Allen K, Martínez-Velázquez M, Alvarez-Maya I. 2012. Transcriptome analysis identifies genes involved in ethanol response of *Saccharomyces cerevisiae* in Agave tequilana juice. Antonie Van Leeuwenhoek. Aug;102(2):247-55.

Rapp A, Mandery H. Wine aroma. Experientia. 1986; 42, 873-884.

Redzepovic S, Orlic S, Sikora A, Majdak A & Pretorius IS. 2002. Identification and characterization of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* strains isolated from Croatian vineyards. Lett. Appl. Microbiol 35:305-310.

Redzepovic S, Orlic S, Majdak A, Kozina B, Volschenk H, Viljoen-Bloom M. 2003. Differential malic acid degradation by selected strains of *Saccharomyces* during alcoholic fermentation. Int J Food Microbiol. May 25;83(1):49-61.

Rementeria A, Rodriguez JA, Cadaval A, Amenabar R, Muguruza JR, Hernando FL, Sevilla MJ. 2003. Yeast associated with spontaneous fermentations of white wines from the "Txakoli de Bizkaia" region (Basque Country, North Spain). Int J Food Microbiol. Sep 1;86(1-2):201-7.

Ringot D, Lerzy B, Bonhoure J P, Auclair E, Oriol E, Larondelle Y. Effect of temperature on in vitro ochratoxin A biosorption onto yeast cell wall derivatives. Process Biochemistry. 2005; 40(9), 3008–3016.

Robinson MD, Grigull J, Mohammad N, Hughes TR. 2002. FunSpec: a web-based cluster interpreter for yeast. BMC Bioinformatics. Nov 13;3:35. Epub 2002 Nov 13.

Rokas A, Williams BL, King N & Carroll SB. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425:798-804.

Rose AH & Harrison JS. 1993. The Yeast, Yeast Technology, Vol. 5. Academic Press, London.

Rossignol T, Dulau L, Julien A, Blondin B. 2003. Genome-wide monitoring of wine yeast gene expression during alcoholic fermentation. Yeast. Dec;20(16):1369-85.

Rossouw D, Næs T, Bauer FF. 2008. Linking gene regulation and the exometabolome: A comparative transcriptomics approach to identify genes that impact on the production of volatile aroma compounds in yeast. BMC Genomics. 9 (1), 530-547.

Salinas F, Cubillos FA, Soto D, García V, Bergström A, Warringer J, Ganga MA, Louis EJJ, Liti G, Martínez C. 2012. The genetic basis of natural variation in oenological traits in *Saccharomyces cerevisiae*. PLOS One 7, e49640.

Salmon JM & Barre P. 1998. Improvement of Nitrogen Assimilation and FermentationKinetics under Enological Conditions by Depression of Alternative Nitrogen-Assimilatory Pathways in an Industrial *Saccharomyces cerevisiae* Strain. Applied and Environmental Microbiology 64, 3831-3837.

Salvadó Z, Arroyo FN, Barrio E, Querol A & Guillamón JM. 2011. Quantifiying the individual effects of etanol and temperatura on the fitness advantage of *Saccharomyces cerevisiae*. Food Microbiol 28:1155-1161.

Salvadó Z, Arroyo-López FN, Guillamón JM, Salazar G, Querol A, Barrio E. Temperature adaptation markedly determines evolution within the genus *Saccharomyces*. Appl Environ Microbiol. 2011 Apr; 77(7):2292-302.

Salvadó Z, Chiva R, Rozès N, Cordero-Otero R, Guillamón JM. 2012. Functional analysis to identify genes in wine yeast adaptation to low-temperature fermentation. J Appl Microbiol. Jul;113(1):76-88.

Samanta MP, Tongprasit W, Sethi H, Chin CS, Stolc V. Global identification of noncoding RNAs in *Saccharomyces cerevisiae* by modulating an essential RNA processing pathway. Proc Natl Acad Sci U S A. 2006 Mar 14;103(11):4192-7. Epub 2006 Mar 6.

Sampaio JP & Gonçalves P. 2008. Natural populations of *Saccharomyces kudriavzevii* in Portugal are associated with oak bark and are sympatric with *S. cerevisiae* and *S. paradoxus*. Appl. Environ. Microbiol. 74: 2144-2152.

Scannell DR, Zill OA, Rokas A, Payen C, Dunham MJ, Eisen MB, Rine J, Johnston M, Hittinger CT. 2011. The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. G3: Genes, Genomes, Genetics 1, 11-25.

Schacherer J, Shapiro JA, Ruderfer DM, Kruglyak L. 2009. Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. Nature. Mar 19;458(7236):342-5.

Shendure J, Ji H. 2008. Next-generation DNA sequencing. Nat Biotechnol. Oct;26(10):1135-45.

Schuller D, Cardoso F, Sousa S, Gomes P, Gomes AC, Santos MA, Casal M. Genetic diversity and population structure of *Saccharomyces cerevisiae* strains isolated from different grape varieties and winemaking regions. PLoS One. 2012; 7(2):e32507.

Sellam A, Hogues H, Askew C, Tebbji F, van het Hoog M, et al. Experimental annotation of the human pathogen *Candida albicans* coding and noncoding transcribed regions using high-resolution tiling arrays. Genome Biol 2010, 11. R71.

Serra A, Strehaiano P, Taillandier P. 2005. Influence of temperature and pH on *Saccharomyces bayanus* var. *uvarum* growth; impact of a wine yeast interspecific hybridization on these parameters. Int J Food Microbiol. Oct 25;104(3):257-65.

Sipiczki M, Romano P, Lipani G, Mikos I, Antunovics Z. 2001. Analysis of yeasts derived from natural fermentation in a Tokaj winery. Antonie Van Leeuwenhoeck 79, 97–105.

Sipiczki M. 2002. Taxonomic and physiological diversity of *Saccharomyces bayanus*. In M. Ciani (Ed.), *Biodiversity and biotechnology of wine yeasts* (pp. 53–69). Kerala: Research Signpost.

Sniegowski PD, Dombrowski PG, Fingerman E. 2002. *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. FEM Yeast Res 1: 299–306.

Sørensen BB, Jakobsen M. 1997. The combined effects of temperature, pH and NaCl on growth of *Debaryomyces hansenii* analyzed by flow cytometry and predictive microbiology. Int J Food Microbiol. 1997 Mar 3;34(3):209-20. Erratum in: Int J Food Microbiol Apr 29;36(1):83-4.

Stambuk BU, Dunn B, Alves SL, Duval EH, Sherlock G. 2009. Industrial fuel ethanol yeasts contain adaptive copy number changes in genes involved in vitamin B1 and B6 biosynthesis. Genome Research 19, 2271-2278.

Storz G. 2002. An expanding universe of noncoding RNAs. Science. May 17;296(5571):1260-3.

Stribny J, Gamero A, Pérez-Torrado R, Querol A. 2015. *Saccharomyces kudriavzevii* and *Saccharomyces uvarum* differ from *Saccharomyces cerevisiae* during the production of aroma-active higher alcohols and acetate esters using their amino acidic precursors. Int J Food Microbiol. Jul 16;205:41-6.

Suárez Valles B, Pando Bedriñana R, González García A & Querol Simón A. 2007. A molecular genetic study of natural strains of *Saccharomyces* isolated from Asturian cider fermentations. J.Appl. Microbiol. 103:778-786.

Suh SO, Blackwell M, Kurtzman CP & Lachance MA. 2006. Phylogenetics of *Saccharomycetales*, the ascomycete yeasts. Mycologia 98: 1006-1017.

Schuller D, Pereira L, Alves H, Cambon B, Dequin S, Casal M. 2007. Genetic characterization of commercial *Saccharomyces cerevisiae* isolates recovered from vineyard environments. Yeast. Aug;24(8):625-36.

Sweeney JY, Kuehne HA, Sniegowski PD. 2004. Sympatric natural *Saccharomyces cerevisiae* and *S. paradoxus* populations have different thermal growth profiles. FEMS Yeast Res. Jan;4(4-5):521-5.

Taxis C, Keller P, Kavagiou Z, Jensen LJ, Colombelli J, *et al*., 2005. Spore number control and breeding in *Saccharomyces cerevisiae*. J Cell Biol 171:627-640.

Tomar P, Bhatia A, Ramdas S, Diao L, Bhanot G, Sinha H. 2013. Sporulation genes associated with sporulation efficiency in natural isolates of yeast. PLoS One. Jul 17;8(7):e69765.

Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. Mar;14(2):178-92.

Torija MJ, Beltran G, Novo M, Poblet M, Guillamón JM, Mas A, Rozès N. Effects of fermentation temperature and *Saccharomyces* species on the cell fatty acid composition and presence of volatile compounds in wine. Int J Food Microbiol. 2003 Aug 15;85(1-2):127-36.

Torriani S, Zapparoli G, Suzzi G. 1999. Genetic and phenotypic diversity of *Saccharomyces* sensu stricto strains isolated from Amarone wine. Diversity of *Saccharomyces* strains from Amarone wine. Antonie Van Leeuwenhoek. Apr;75(3):207-15.

Townsend JP, Cavalieri D, Hartl DL. 2003. Population genetic variation in genome-wide gene expression. Mol Biol Evol. Jun;20(6):955-63..

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105-1111.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010, 28:511-5.

Tronchoni J, Gamero A, Arroyo-López FN, Barrio E, Querol A. 2009 .Differences in the glucose and fructose consumption profiles in diverse *Saccharomyces* wine species and their hybrids during grape juice fermentation. Int J Food Microbiol. Sep 15;134(3):237-43.

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

Van Verk MC, Hickman R, Pieterse CM, Van Wees SC. 2013. RNA-Seq: revelation of the messengers. Trends Plant Sci. Apr;18(4):175-9.

Varela C, Cárdenas J, Melo F, Agosin E. 2005. Quantitative analysis of wine yeast gene expression profiles under winemaking conditions. Yeast. Apr 15;22(5):369-83.

Vaughan-Martini A, Martini A. 1998. *Saccharomyces* Meyen ex Reess. In The Yeasts: A Taxonomic Study, 4th edn, Kurztman CP, Fell JW (eds). Elsevier: Amsterdam; 358–371.

Vinogradov AE, Anatskaya OV. 2009. Loss of protein interactions and regulatory divergence in yeast whole-genome duplicates. Genomics. Jun;93(6):534-42.

Wang SA & Bai FY. 2008. *Saccharomyces arboricolus* sp. nov., a yeast species from the tree bark. Int J Syst Evol Microbiol. Feb;58(Pt 2):510-4.

Wang Z, Gerstein M, Snyder M .2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10: 57-63.

Warringer J, Zörgö E, Cubillos FA, Zia A, Gjuvsland A, Simpson JT, Forsmark A, Durbin R, Omholt SW, Louis EJ, Liti G, Moses A, Blomberg A. 2011. Trait variation in yeast is defined by population history. PLoS Genet. Jun;7(6):e1002111.

Waters EJ, Pellerin P & Brillouet JM. 1994. A *Saccharomyces* mannoprotein that protects wine from protein haze. Carbohydrate Polymers, 23(3), 185–191.

Wiggins LF & Williams JH. 1955. Nutritional value of crops, amino acid content of West Indies sugar cane. Journal of Agricultural and Food Chemistry 3, 341-345.

Wilhelm J, Pingoud A, Hahn M. 2003. Real-time PCR-based method for the estimation of genome sizes. Nucleic Acids Research 31, e56.

Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bähler J. 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature. Jun 26;453(7199):1239-43.

Wilhelm BT & Landry JR. 2009. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. Methods. Jul;48(3):249-57

Winzeler EA, Castillo-Davis CI, Oshiro G, Liang D, Richards DR, Zhou Y, Hartl DL. 2003. Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays. Genetics 163, 79-89.

Wolfe KH, Shields DC. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. Nature. Jun 12;387(6634):708-13.

Wnuk M, Panek A, Golec E, Magda M, Deregowska A, Adamczyk J, Lewinska A. 2015. Genetic profiling of yeast industrial strains using in situ comparative genomic hybridization (CGH). J Biotechnol. Sep 20;210:52-6.

Wu CY, Bird AJ, Chung LM, Newton MA, Winge DR, Eide DJ. 2008. Differential control of Zap1-regulated genes in response to zinc deficiency in *Saccharomyces cerevisiae*. BMC Genomics. Aug 1;9:370.

Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM. Bidirectional promoters generate pervasive transcription in yeast. Nature. 2009 Feb 19;457(7232):1033-7.

Zambonelli C, Passarelli P, Rainieri S, Bertolini L, Giudici P & Castellarl I. 1997. technological properties and temperature response of interspecific *Saccharomyces* hybrids. J. Sci. Food Agric. 74,7-12.

Zara S, Bakalinsky AT, Zara G, Pirino G, Demontis MA, Budroni M. 2005. *FLO11*-based model for air-liquid interfacial biofilm formation by *Saccharomyces cerevisiae*. Applied and Environmental Microbiology 71, 2934-2939.

Zerbino DR & Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008, May;18(5):821-9.

Zhang H, Richards KD, Wilson S, Lee SA, Sheehan H, Roncoroni M, Gardner RC. 2015. Genetic characterization of strains of *Saccharomyces uvarum* from New Zealand wineries. Food Microbiol. Apr;46:92-9.

11. ANNEX I

Suplementary data of Chapter 1

Table S1.1. Genes showing differences in the copy number (GCN) at least in one strain. These differences are highlighted in red and green, meaning higher or lower copy number than the reference strain S288c respectively.

ID	Gene ID	T73	CECT 10131	CPE7	GB-FlorC	Kyokai	PE35M	Temohaya-26	Function	Process
YAL004W		3,10	-1,91	0,68	-0,07	0,30	-0,15	0,43	molecular_function unknown	biological_process unknown
YAL007C	ERP2	-0,32	-0,13	0,22	-0,36	1,32	-0,61	-0,53	homology to YOR016c	ER to Golgi transport
YAL009W	SPO7	-0,54	-0,42	0,39	0,22	1,34	-1,07	-0,78	meiotic protein	sporulation (sensu Saccharomyces); nuclear organization and biogenesis; meiosis
YAL011W	SWC1	-0,53	-0,30	0,61	0,24	1,30	-0,79	-0,37	hypothetical protein	biological_process unknown
YAL012W	CYS3	-0,43	-0,16	0,70	-0,05	1,31	-0,53	-0,16	cystathionine gamma-lyase	cysteine metabolism; sulfur amino acid metabolism; transsulfuration
YAL013W	DEP1	-0,44	-0,25	1,61	-0,41	0,72	-0,64	-0,28	regulator of phospholipid metabolism	regulation of transcription from Pol II promoter; phospholipid metabolism
YAL015C	NTG1	-0,13	0,05	0,35	0,09	1,26	-0,29	-0,39	similarity to UV endonuclease	DNA repair; base-excision repair, AP site formation; base-excision repair
YAL020C	ATS1	-0,64	-0,24	1,33	-0,67	0,03	-0,40	-0,14	alpha-tubulin supressor	microtubule-based process; bud growth; microtubule cytoskeleton organization and biogenesis
YAL025C	MAK16	-0,63	-0,38	0,37	0,06	1,18	-0,79	-0,54	nuclear viral propagation protein	ribosomal large subunit biogenesis; host-pathogen interaction
YAL026C	DRS2	-1,46	-0,46	0,15	-0,17	-0,19	-0,22	-0,14	membrane-spanning P-type Ca-ATPase	post-Golgi transport; intracellular protein transport; processing of 20S pre-rRNA
YAL027W	YAL027W	-1,07	-0,86	0,92	0,03	0,84	-0,95	-0,44	hypothetical protein	biological_process unknown
YAL030W	SNC1	-1,06	-0,66	0,51	0,04	0,96	-0,90	-0,58	homology to synaptic vesicle-associated	vesicle fusion; endocytosis; Golgi to plasma membrane transport
YAL034C-B		-1,22	-0,77	0,23	0,08	0,92	-0,97	-0,52		
YAL034W-A	MTW1	-0,34	-0,39	0,13	1,35	0,78	-0,50	-0,31		chromosome segregation
YAL037W	YAL037W	-0,26	-0,41	0,24	-0,38	1,13	-0,74	-0,13	hypothetical protein	biological_process unknown
YAL049C	YAL049C	-1,09	-0,71	0,13	-0,34	1,02	-1,22	-0,50	hypothetical protein	biological_process unknown
YAL053W	YAL053W	-0,34	0,00	-0,06	-0,29	0,05	1,20	-0,18	homology to YOR365c,YGL139w,YPL221w	biological_process unknown
YAL054C	ACS1	-0,14	-0,40	0,26	-0,03	1,34	-0,72	-0,22	acetyl-CoA synthetase	acetate fermentation; acetyl-CoA biosynthesis
YAL059W	ECM1	-0,89	-0,77	0,12	-0,07	1,42	-1,32	-0,19	hypothetical protein	cell wall organization and biogenesis; ribosomal large subunit-nucleus export
YAL060W	BDH1	-0,93	-1,12	0,46	-0,04	1,95	-1,60	-0,50	similarity to alcohol/sorbitol dehydroge	butanediol fermentation
YAL061W	YAL061W	-0,61	-0,61	0,43	-0,26	1,13	-0,79	-0,21	similarity to alcohol/sorbitol dehydroge	biological_process unknown

YAL062W	**GDH3**	-0,71	-1,10	0,37	-0,33	0,86	-1,36	-0,34	NADP-glutamate dehydrogenase	glutamate biosynthesis
YAL063C	**FLO9**	-1,24	-0,42	-0,12	-0,48	0,02	-0,91	-0,49	homology to Flo1p	biological_process unknown
YAL064W	**YAL064W**	-3,20	-1,68	-1,48	-5,37	-4,53	-3,29	-2,78	putative cell wall protein involved in	biological_process unknown
YAL065C		-2,40	-0,83	-0,27	-0,73	0,27	-0,81	-0,91	homology to Flo1p/putative pseudogene	biological_process unknown
YAL065C-A	**YAL065C-A**	-2,05	-2,76	-0,68	-2,19	-2,09	-4,01	-1,83		
YAL066W		-1,84	-1,69	-0,15	0,73	1,75	-0,83	-0,67	hypothetical protein	biological_process unknown
YAL067C	**SEO1**	-1,60	-1,31	-0,01	0,82	1,55	-1,01	-0,57	similarity to allantoate permease Dal5p	transport
YAR007C	**RFA1**	-0,74	-0,52	0,66	0,19	1,67	-0,74	-0,22	DNA replication factor-A protein 1	DNA recombination; double-strand break repair; postreplication repair; nucleotide-excision repair; DNA strand elongation; DNA replication, priming; DNA unwinding
YAR009C	**YAR009C**	-1,27	-0,40	0,39	-1,23	-1,06	-1,25	-0,60	Ty1B protein	Ty element transposition
YAR010C	**YAR010C**	-1,36	-0,69	0,32	-1,84	-1,65	-1,67	-0,72	TY1A protein	Ty element transposition
YAR018C	**KIN3**	-0,45	-0,17	0,49	0,14	1,14	-0,75	-0,30	ser/thr protein kinase	chromosome segregation
YAR023C	**YAR023C**	-1,00	-0,20	0,39	0,20	1,19	-0,70	-0,67	member of the YBR302p/YCR007p/YHL048p/YK	biological_process unknown
YAR027W	**UIP3**	-0,95	-0,77	0,79	0,30	-2,27	-0,40	-1,91	member of the YBR302p/YCR007p/YHL048p/YK	biological_process unknown
YAR028W	**YAR028W**	-0,56	0,35	-1,93	-0,04	0,70	-0,31	-2,60	member of the YBR302p/YCR007p/YHL048p/YK	biological_process unknown
YAR029W	**YAR029W**	-0,37	0,48	-3,17	0,35	0,58	-0,18	-2,27	member of the YBR302p/YCR007p/YHL048p/YK	biological_process unknown
YAR030C		-0,36	0,27	-4,19	0,32	0,08	-0,11	-2,76	hypothetical protein	biological_process unknown
YAR031W	**PRM9**	-1,27	0,10	-2,11	-0,90	-0,17	-1,47	-2,40	member of the YBR302p/YCR007p/YHL048p/YK	conjugation with cellular fusion
YAR033W	**MST28**	-1,18	-0,16	-2,82	-0,60	-0,08	-0,46	-2,22	member of the YBR302p/YCR007p/YHL048p/YK	vesicle organization and biogenesis
YAR040C		0,75	0,29	1,95	-0,82	0,53	-0,06	0,82	hypothetical protein	
YAR047C	**YAR047C**	-3,30	-4,40	-4,14	-2,46	-4,73	-4,01	-0,69	hypothetical protein	biological_process unknown
YAR050W	**FLO1**	-0,76	-0,72	-1,03	-1,26	-0,70	-1,38	-0,92	putative lectin-like cell wall protein	flocculation
YAR052C		-0,67	-0,10	-0,42	-1,20	-0,62	-0,71	-0,68	hypothetical protein	
YAR053W	**YAR053W**	-1,29	-0,86	-1,11	-6,94	-5,67	-2,81	-3,27	hypothetical protein	biological_process unknown
YAR060C	**YAR060C**	-4,14	-1,16	-0,93	-6,58	-4,09	-3,23	-2,22	homology to hypothetical protein YHR212c	biological_process unknown
YAR061W	**YAR061W**	-4,21	-0,83	-1,80	-6,23	-5,77	-3,12	-1,78	similarity to Flo1p/putative pseudogene	biological_process unknown
YAR062W	**YAR062W**	-1,21	-1,05	0,02	-1,43	-0,64	-2,24	-1,59	homology to Flo1p/putative pseudogene	biological_process unknown

YAR064W	**YAR064W**	-0,45	-0,17	0,65	0,18	-0,28	-0,81	-1,20	hypothetical protein	biological_process unknown
YAR066W	**YAR066W**	-0,04	0,15	0,56	0,05	0,42	-0,67	-0,99		biological_process unknown
YAR068W	**YAR068W**	-0,36	0,18	1,06	0,37	1,61	-0,42	-0,93	homology to hypothetical protein YHR214w	biological_process unknown
YAR069C		-0,95	-0,41	-0,16	-0,34	0,38	-0,91	-0,57	hypothetical protein	biological_process unknown
YAR070C		0,16	0,72	0,16	0,53	-1,55	0,28	-1,76	hypothetical protein	biological_process unknown
YAR071W	**PHO11**	-1,18	-0,40	-0,13	-0,46	-0,42	-1,24	-1,20	secreted acid phosphatase,56 kDa isozym	phosphate metabolism
YBL002W	**HTB2**	-0,71	-0,08	0,43	0,37	-9,17	-0,08	-0,13	histone H2B,2	chromatin assembly/disassembly
YBL005W	**YBL005W**	-0,14	-1,78	0,01	-0,41	-1,14	-0,24	0,00	DNA binding; transcriptional activator activity	response to drug; regulation of transcription from Pol II promoter
YBL005W-A	**YBL005W-A**	-1,41	-0,71	0,17	-3,70	1,67	-2,06	-0,88	TY1A protein	Ty element transposition
YBL005W-B	**YBL005W-B**	-0,73	0,00	0,02	-0,45	-0,32	-0,99	-0,41	protein binding; RNA binding; ribonuclease activity; peptidase activity; RNA-directed DNA polymerase activity; DNA-directed DNA polymerase activity	Ty element transposition
YBL006C	**LDB7**	-0,45	0,10	0,25	0,18	2,42	0,08	-0,11	hypothetical protein	cell wall mannoprotein biosynthesis
YBL013W	**FMT1**	-0,16	0,38	1,17	0,23	-0,03	0,45	-0,07	homology to methionyl-tRNA formyltransfe	methionyl-tRNA aminoacylation; translational initiation
YBL021C	**HAP3**	1,04	1,31	0,29	-0,12	-0,45	1,18	0,31	transcriptional activator	regulation of carbohydrate metabolism; transcription
YBL024W	**NCL1**	0,94	0,50	-0,13	-1,45	-0,18	0,72	-0,05	similarity to nucleolar Nop2p	tRNA methylation
YBL025W	**RRN10**	-0,66	0,30	0,44	0,26	4,94	0,25	-0,47	RNA polymerase I-specific transcription	transcription from Pol I promoter
YBL026W	**LSM2**	0,68	0,44	0,00	1,78	0,21	0,20	-0,20	snRNP-related protein	nuclear mRNA splicing, via spliceosome; rRNA processing
YBL027W	**RPL19B**	-0,32	0,18	0,30	-0,30	1,20	0,14	-0,26	ribosomal protein L19,e	protein biosynthesis
YBL028C	**YBL028C**	1,30	0,70	-0,06	0,98	0,18	0,84	-0,24	hypothetical protein	biological_process unknown
YBL029W	**YBL029W**	0,12	0,69	0,13	1,35	-0,57	0,50	-0,13	hypothetical protein	biological_process unknown
YBL030C	**PET9**	1,02	0,80	0,13	-0,58	-0,03	0,84	0,06	ADP,ATP carrier protein 2	ATP/ADP exchange
YBL050W	**SEC17**	0,69	0,70	-0,04	-0,27	-1,28	0,81	0,39	transport vesicle fusion protein	ER to Golgi transport
YBL051C	**PIN4**	-0,34	1,55	0,39	0,03	-4,98	-0,18	-0,33	molecular_function unknown	biological_process unknown
YBL052C	**SAS3**	-0,31	-0,13	0,06	-0,20	-1,90	-0,10	-0,14	similarity to Sas2p	chromatin modification; chromatin silencing at telomere; chromatin silencing at HML and HMR (sensu Saccharomyces)
YBL072C	**RPS8A**	0,00	-0,12	-0,24	-0,11	0,41	0,09	-0,01	ribosomal protein S8,e	protein biosynthesis
YBL073W		-0,25	-0,32	-0,35	-0,07	1,16	0,06	-0,23	questionable ORF	biological_process unknown
YBL075C	**SSA3**	-0,29	-0,40	-0,10	-1,71	8,53	-0,17	-0,10	cytoplasmic heat shock protein	SRP-dependent cotranslational membrane targeting, translocation; response to stress; protein folding; protein-mitochondrial targeting

YBL078C	**ATG8**	-0,58	-0,33	-0,74	-4,92	4,17	-0,05	-0,18	homology to unknown C,elegans protein	autophagy; protein-vacuolar targeting
YBL079W	**NUP170**	-0,26	0,12	0,01	1,42	-0,33	0,26	-0,27	nuclear pore protein	protein-nucleus import, docking; nuclear pore organization and biogenesis; protein-nucleus export; ribosomal protein-nucleus import
YBL100C		-0,95	-0,81	0,11	0,01	0,21	-0,73	-0,16	questionable ORF	biological_process unknown
YBL101W-A	**YBL101W-A**	0,03	0,57	0,61	0,18	-0,64	0,77	-1,94	TY2A protein	Ty element transposition
YBL109W		-0,67	-1,00	0,43	-0,89	0,00	-1,52	-0,48	similarity to hypothetical proteins YDR5	biological_process unknown
YBL111C		-0,96	-0,16	0,03	0,18	-0,04	-0,42	0,29	homology to other subtelomeric encoded p	biological_process unknown
YBR012W-A	**YBR012W-A**	-1,12	-0,44	0,11	-1,29	-2,09	-1,88	-0,98	TY1A protein	Ty element transposition
YBR012W-B	**YBR012W-B**	-1,55	-0,16	-0,04	-0,41	-0,01	-0,67	-0,36	TY1B protein	Ty element transposition
YBR032W		0,01	-0,01	0,17	-0,38	-0,09	0,39	-1,01	hypothetical protein	biological_process unknown
YBR036C	**CSG2**	0,17	0,22	-0,02	-0,13	-0,73	0,41	-0,86	calcium dependent regulatory protein	calcium ion homeostasis
YBR076W	**ECM8**	0,04	-0,40	-0,12	-0,12	0,34	0,24	0,86	hypothetical protein	cell wall organization and biogenesis
YBR077C	**YBR077C**	-0,11	0,41	0,20	-0,09	0,18	-0,30	-0,87	hypothetical protein	biological_process unknown
YBR079C	**RPG1**	-0,31	0,00	0,17	0,02	0,22	-0,21	0,09	translation initiation factor activity	translational initiation
YBR089C-A	**NHP6B**	0,84	0,71	-0,23	-0,14	-0,81	0,66	0,58		regulation of transcription from Pol II promoter; establishment and/or maintenance of chromatin architecture; regulation of transcription from Pol III promoter
YBR093C	**PHO5**	-0,24	-0,27	1,10	0,17	1,09	-0,49	-0,55	repressible acid phosphatase precursor	cellular response to phosphate starvation; phosphate metabolism
YBR103W	**SIF2**	-0,02	0,02	0,02	0,05	0,40	4,24	-0,07	weak similarity to YCR057p	chromatin silencing at telomere; histone deacetylation; negative regulation of transcription from Pol II promoter; negative regulation of meiosis
YBR105C	**VID24**	0,07	0,31	0,13	1,13	-1,69	-1,18	-0,13	similarity to YGR066c	vesicle-mediated transport; negative regulation of gluconeogenesis
YBR122C	**MRPL36**	0,62	0,00	0,49	0,30	-1,35	1,00	-0,28	structural constituent of ribosome	protein biosynthesis
YBR123C	**TFC1**	0,88	1,19	0,52	0,07	-0,09	0,76	0,22	RNA polymerase transcription factor III	transcription initiation from Pol III promoter
YBR124W		0,74	1,96	0,45	0,24	-2,49	1,34	-0,20	questionable ORF	biological_process unknown
YBR125C	**PTC4**	0,90	0,75	0,63	0,17	0,55	0,92	0,36	homology to protein phosphatase 2C	biological_process unknown
YBR128C	**ATG14**	0,69	2,10	0,60	0,17	0,15	1,11	-0,09	molecular_function unknown	autophagy
YBR129C	**OPY1**	0,80	1,35	0,51	0,12	0,45	0,90	0,39	molecular_function unknown	conjugation with cellular fusion
YBR130C	**SHE3**	1,08	1,36	0,30	-0,07	-0,23	1,27	0,39	required for mother cell-specific expres	mRNA localization, intracellular
YBR131W	**CCZ1**	0,91	1,08	0,33	-0,10	-0,40	0,69	0,12	hypothetical protein	vacuolar transport; autophagic vacuole fusion; autophagy; protein-vacuolar targeting
YBR144C		-0,35	0,26	-0,68	-0,61	-1,27	0,04	-0,48	hypothetical protein	biological_process unknown

YBR153W	**RIB7**	-0,54	-0,67	-0,05	1,25	-0,05	-0,77	-0,28	HTP reductase	vitamin B2 biosynthesis
YBR174C		-0,68	-1,32	0,12	-0,03	-0,21	-1,12	-0,19	molecular_function unknown	biological_process unknown
YBR250W	**YBR250W**	0,99	0,08	-0,06	-0,20	0,33	-0,12	0,06	hypothetical protein	biological_process unknown
YBR269C	**FMP21**	-0,61	-0,55	0,18	0,26	-2,51	-0,34	-0,05	hypothetical protein	biological_process unknown
YBR295W	**PCA1**	-0,20	-0,30	-0,43	-1,52	-0,52	-0,22	0,08	P-type Cu2+-transporting ATPase	copper ion homeostasis
YBR298C	**MAL31**	0,26	0,81	1,10	0,82	1,97	0,33	0,01	maltose permease	alpha-glucoside transport
YCL020W	**YCL020W**	-0,09	0,64	0,29	0,15	-0,73	0,89	-1,40	TY2A protein	Ty element transposition
YCL065W		-0,55	-1,30	0,43	0,32	-0,13	-0,67	-0,42	molecular_function unknown	biological_process unknown
YCL066W	**HMLALPHA1**	-0,71	-1,30	-0,67	0,10	0,23	-0,64	-0,32	transcription coactivator activity	regulation of transcription from Pol II promoter; regulation of transcription, mating-type specific
YCL069W	**YCL069W**	-0,52	-2,63	0,47	0,52	0,44	-0,97	0,00	molecular_function unknown	biological_process unknown
YCL073C	**YCL073C**	-0,55	-2,36	-0,12	-0,10	-0,21	-0,69	-0,15	transporter activity	transport
YCL074W	**YCL074W**	1,02	-0,85	0,67	0,68	0,44	1,18	-2,56	homology to retrotransposon and retrovir	Ty element transposition
YCL075W	**YCL075W**	0,24	-1,27	0,46	0,65	0,25	0,88	-2,90	Ty5-1 transposon-encoded protein	Ty element transposition
YCL076W		0,79	-0,90	0,92	1,01	0,43	0,73	-1,52	hypothetical protein	biological_process unknown
YCR020C	**PET18**	-0,47	-0,51	1,59	0,59	-0,06	-1,01	-0,07	hypothetical protein	mitochondrion organization and biogenesis; thiamin metabolism
YCR020C-A	**MAK31**	-0,61	-0,63	1,68	0,61	0,59	-1,02	-0,03	involved in stability of L-A double-stra	host-pathogen interaction; N-terminal protein amino acid acetylation
YCR021C	**HSP30**	-0,40	-0,45	1,15	0,54	0,46	-0,80	-0,07	heat shock protein	response to stress; protein folding
YCR022C		-0,13	-0,08	1,00	0,23	-0,36	-0,09	0,26	hypothetical protein	biological_process unknown
YCR023C	**YCR023C**	-0,51	-0,67	1,47	0,35	0,49	-0,66	-0,04	member of major facilitator superfamily	biological_process unknown
YCR024C	**YCR024C**	-0,75	-0,70	1,40	0,64	0,84	-0,80	0,02	mitochondrial asn-tRNA synthetase	protein biosynthesis; biological_process unknown
YCR024C-A	**PMP1**	-0,22	-0,53	1,27	0,04	0,02	-0,11	0,17	H+-transporting ATPase subunit	cation transport
YCR025C		-0,65	-0,63	1,04	0,34	0,29	-0,44	-0,04	hypothetical protein	biological_process unknown
YCR026C	**YCR026C**	-0,50	-0,57	1,25	0,18	0,54	-0,54	-0,07	weak similarity to human autotaxin precu	biological_process unknown
YCR027C	**RHB1**	-0,75	-0,87	1,13	0,29	0,21	-0,66	-0,16	putative GTP-binding protein	L-lysine transport; L-arginine transport
YCR040W	**MATALPHA1**	-1,00	-1,50	-0,39	0,03	0,42	-0,73	-0,39	regulatory protein for mating type speci	regulation of transcription from Pol II promoter; regulation of transcription, mating-type specific
YCR073W-A	**SOL2**	-0,50	0,00	0,25	0,45	0,21	-0,49	0,79		tRNA processing
YCR085W		-0,82	-0,52	1,37	0,81	-0,66	-0,79	-0,28	hypothetical protein	biological_process unknown

YCR095C	**YCR095C**	-1,04	-1,05	-0,05	0,56	-0,12	-1,07	-0,28	hypothetical protein	biological_process unknown
YCR098C	**GIT1**	-1,00	-0,94	-0,24	0,41	-0,11	-1,00	-0,25	similarity to transporter proteins	phospholipid transport
YCR105W	**ADH7**	-1,15	-0,03	-3,46	-4,50	-3,33	-0,37	-2,81	putative alcohol dehydrogenase	alcohol metabolism
YCR106W	**RDS1**	-1,19	-0,18	-2,46	-3,77	-2,69	-0,35	-2,20	putative transcription factor	response to xenobiotic stimulus
YCR107W	**AAD3**	-0,69	-0,46	-1,51	-2,26	-0,91	-0,64	-1,46	putative aryl-alcohol reductase	aldehyde metabolism
YDL010W	**YDL010W**	0,11	1,20	0,17	0,06	0,04	1,36	-0,13	similarity to YBR014p and glutaredoxins	biological_process unknown
YDL013W	**HEX3**	0,19	1,40	0,66	0,02	0,35	0,34	-0,04	DNA binding	DNA recombination; response to DNA damage stimulus; sporulation (sensu Saccharomyces)
YDL014W	**NOP1**	0,10	0,83	0,21	-0,40	-0,04	0,79	0,83	fibrillarin	35S primary transcript processing; ribosomal large subunit assembly and maintenance; rRNA modification; processing of 20S pre-rRNA; RNA methylation
YDL016C		-0,22	1,22	0,24	0,24	-0,41	1,47	-0,29	questionable ORF	biological_process unknown
YDL017W	**CDC7**	0,20	1,33	0,58	-0,08	0,37	0,38	-0,39	protein serine/threonine kinase activity	protein amino acid phosphorylation; regulation of DNA replication; DNA replication initiation
YDL018C	**ERP3**	0,14	1,30	0,47	0,08	-0,18	1,17	-0,15	molecular_function unknown	secretory pathway
YDL020C	**RPN4**	0,47	1,25	0,16	0,15	-0,31	0,95	0,18	endopeptidase activity	ubiquitin-dependent protein catabolism
YDL021W	**GPM2**	1,17	1,62	0,06	-0,45	-0,79	1,51	0,24	homology to phosphoglycerate mutase	glycolysis; gluconeogenesis
YDL023C		0,42	0,98	0,97	-0,56	0,46	0,72	0,76	questionable ORF	biological_process unknown
YDL032W		0,61	1,19	0,34	-0,25	-0,61	0,86	0,41	molecular_function unknown	biological_process unknown
YDL038C	**YDL038C**	-1,06	-0,02	0,19	-0,08	0,04	0,05	0,05	weak similarity to hypothetical protein	biological_process unknown
YDL067C	**COX9**	0,66	1,24	-0,12	0,40	-0,30	0,58	0,36	cytochrome-c oxidase activity	aerobic respiration
YDL068W		0,88	-0,12	-0,22	-0,05	-0,80	0,15	0,11	questionable ORF	biological_process unknown
YDL093W	**PMT5**	-0,42	-0,52	0,33	0,42	0,64	4,12	-0,07	similarity to O-mannosyltransferases Pm	O-linked glycosylation
YDL111C	**RRP42**	1,12	0,19	-0,54	-0,30	-0,28	0,57	0,48	similarity to unknown human ORF	mRNA catabolism; 35S primary transcript processing
YDL113C	**ATG20**	1,01	0,00	-0,23	-0,69	0,13	0,15	-0,10	weak similarity to YDR425w	protein-vacuolar targeting; autophagy
YDL117W	**CYK3**	1,18	-0,11	0,05	-0,11	-0,29	0,12	0,34	hypothetical protein	cytokinesis
YDL222C	**FMP45**	-0,21	-0,05	1,07	-0,28	-0,50	-0,31	-0,08	homology to hypothetical proteins YNL194	cell wall organization and biogenesis
YDL229W	**SSB1**	-0,39	0,00	0,08	-0,18	-0,58	0,10	-0,13	chaperone activity; ATPase activity	protein biosynthesis
YDL242W		-0,79	-1,01	-0,59	0,30	-0,04	-1,51	-0,62	homology to hypothetical protein YPR079w	biological_process unknown
YDL243C	**AAD4**	-0,41	-1,12	-0,48	-0,39	-0,21	-1,19	-0,36	putative aryl-alcohol dehydrogenase	aldehyde metabolism

YDL244W	**THI13**	-0,80	-1,45	-0,07	-0,23	0,34	-1,67	-0,14	homology to Thi5p,YJR156p,and NMT1 prot	thiamin biosynthesis
YDL245C	**HXT15**	-0,04	-3,22	-2,19	-0,51	-1,77	-2,00	-1,10	putative hexose permease	hexose transport
YDL246C	**SOR2**	-0,64	-5,39	-5,34	-0,70	-4,86	-2,59	-3,54	putative sugar dehydrogenase	hexose metabolism
YDL247W	**MPH2**	-0,63	-3,61	-3,62	-1,06	-3,80	-2,42	-2,54	homology to sugar transport proteins	carbohydrate transport
YDR022C	**CIS1**	1,01	0,81	-0,11	-0,35	-0,72	1,01	0,34	hypothetical protein	regulation of CDK activity
YDR034C	**LYS14**	0,84	0,42	-0,46	-0,30	-0,03	0,65	0,16	transcriptional activator of lysine path	lysine biosynthesis, aminoadipic pathway
YDR036C	**EHD3**	1,16	-0,65	0,31	0,42	0,58	-0,73	1,04	putative enoyl CoA hydratase	endocytosis; fatty acid beta-oxidation
YDR038C	**ENA5**	-0,02	-1,03	-1,18	-1,69	-2,65	-0,91	-1,48	P-type ATPase involved in Na+ efflux	sodium ion transport
YDR039C	**ENA2**	-0,39	-1,08	-1,03	-1,91	-2,44	-1,00	-1,40	P-type ATPase involved in Na+ efflux	sodium ion transport
YDR040C	**ENA1**	-0,30	-1,04	-1,11	-1,90	-2,09	-1,06	-1,43	P-type ATPase involved in Na+ and Li+ ef	sodium ion transport
YDR094W		0,77	0,00	1,24	-0,61	0,58	0,44	1,68	questionable ORF	biological_process unknown
YDR095C		0,51	0,00	1,21	-0,65	-0,08	0,45	1,25	hypothetical protein	biological_process unknown
YDR104C	**SPO71**	-0,46	-0,08	-0,06	-1,39	0,03	0,11	-0,15	hypothetical protein	spore wall assembly (sensu Saccharomyces)
YDR135C	**YCF1**	-0,28	-0,21	0,11	0,02	-1,18	2,13	-0,06	vacuolar glutathione S-conjugate transpo	bilirubin transport; cadmium ion transport; arsenite transport; response to mercury ion
YDR152W	**GIR2**	0,54	0,53	-0,25	-0,27	-0,39	1,45	0,39	hypothetical protein	biological_process unknown
YDR154C		0,12	0,31	-0,29	1,25	-0,25	0,34	0,07	questionable ORF	biological_process unknown
YDR155C	**CPR1**	1,02	0,64	-0,37	-0,33	-0,52	0,47	0,73	cyclophilin (peptidylprolyl isomerase)	protein metabolism
YDR162C	**NBP2**	0,91	0,76	0,00	-0,18	-0,28	0,63	0,30	Nap1p-binding protein	biological_process unknown
YDR184C	**ATC1**	1,20	0,68	-0,62	0,00	-0,89	1,08	0,66	hypothetical protein	response to stress; polar budding
YDR204W	**COQ4**	0,40	1,05	-1,18	1,29	1,07	0,88	-0,21	hypothetical protein	ubiquinone metabolism
YDR208W	**MSS4**	0,48	1,16	0,18	0,31	-0,34	0,81	0,07	1-phosphatidylinositol-4-phosphate 5-kinase activity	actin cable assembly; actin cytoskeleton organization and biogenesis
YDR212W	**TCP1**	0,50	1,31	0,26	0,15	-0,15	1,29	-0,18	component of chaperonin-containing T-com	cytoskeleton organization and biogenesis; protein folding
YDR215C		0,21	0,00	0,55	-0,63	0,84	0,32	1,17	hypothetical protein	biological_process unknown
YDR246W	**TRS23**	0,16	0,13	0,91	-0,59	-1,33	0,35	0,39		ER to Golgi transport
YDR251W	**PAM1**	1,22	0,16	0,37	0,04	0,22	0,49	0,11	coiled-coil protein multicopy suppressor	pseudohyphal growth
YDR262W	**YDR262W**	0,86	0,59	0,24	-1,22	0,27	4,26	-0,33	hypothetical protein	biological_process unknown
YDR347W	**MRP1**	1,03	2,65	1,60	-1,98	0,59	0,39	0,11	mitochondrial ribosomal protein of the s	protein biosynthesis

YDR348C	**YDR348C**	1,35	1,73	0,02	1,45	0,05	1,58	0,14	similarity to hypothetical protein YHR09	biological_process unknown
YDR349C	**YPS7**	8,53	1,60	1,26	-1,99	0,98	0,32	-0,90	weak similarity to YLR121c and Mkc7p	biological_process unknown
YDR361C	**BCP1**	0,87	1,07	-0,04	-0,18	0,16	0,70	0,41	hypothetical protein	biological_process unknown
YDR379W	**RGA2**	0,10	0,28	0,00	1,28	0,43	0,10	-0,01	putative GTPase activating protein	actin filament organization; signal transduction during conjugation with cellular fusion; invasive growth (sensu Saccharomyces); pseudohyphal growth; isotropic bud growth; apical bud growth; establishment of cell polarity (sensu Saccharomyces); small GTPa
YDR387C	**YDR387C**	0,06	0,32	0,29	4,81	0,20	0,29	-0,19	similarity to Itr1p and Itr2p	biological_process unknown
YDR394W	**RPT3**	0,96	1,02	-0,04	0,03	-0,31	1,07	-0,13	subunit of 26S protease	ubiquitin-dependent protein catabolism
YDR395W	**SXM1**	1,08	1,29	0,44	0,00	-0,17	0,67	-0,04	similarity to NMD and CSE1 proteins	mRNA-nucleus export; nucleocytoplasmic transport
YDR396W		0,83	1,28	0,06	-0,20	-0,16	1,13	0,00	hypothetical protein	biological_process unknown
YDR397C	**NCB2**	1,21	1,37	0,64	-0,07	0,39	1,08	0,08	similarity to human TATA-binding protein	negative regulation of transcription from Pol II promoter
YDR398W	**UTP5**	1,18	1,49	0,37	0,18	-0,09	1,45	0,13	hypothetical protein	processing of 20S pre-rRNA
YDR399W	**HPT1**	0,99	1,32	1,02	0,60	0,40	0,84	0,08	homology to hypothetical protein YJR133w	purine nucleotide biosynthesis
YDR400W	**URH1**	1,14	1,53	-0,10	-0,19	-0,20	1,29	-0,16	similarity to E,coli hypothetical 33,7 k	pyrimidine salvage; uridine catabolism
YDR401W		0,68	1,29	0,34	-0,34	0,07	1,05	0,02	molecular_function unknown	biological_process unknown
YDR402C	**DIT2**	0,77	1,47	0,05	-0,16	-0,21	0,93	0,24	catalytic activity	spore wall assembly (sensu Saccharomyces)
YDR403W	**DIT1**	0,71	1,47	0,97	0,90	0,80	0,98	0,43	spore wall maturation protein	spore wall assembly (sensu Saccharomyces)
YDR404C	**RPB7**	0,51	1,44	0,78	0,56	0,25	1,30	0,37	RNA polymerase II subunit	transcription from Pol II promoter
YDR405W	**MRP20**	0,29	1,39	1,15	0,73	0,36	1,15	0,43	mitochondrial ribosomal protein of the l	protein biosynthesis
YDR408C	**ADE8**	0,48	1,22	0,96	0,70	0,65	0,66	0,21	phosphoribosylglycinamide formyltransfer	adenine biosynthesis; purine nucleotide biosynthesis
YDR410C	**STE14**	0,78	1,18	0,11	0,15	0,01	1,22	0,12	farnesyl cysteine carboxyl-methyltransfe	peptide pheromone maturation
YDR411C	**YDR411C**	1,12	1,11	0,05	-0,20	-0,32	0,95	0,15	hypothetical protein	biological_process unknown
YDR412W	**YDR412W**	1,28	1,59	0,05	-0,20	-0,28	1,00	0,32	questionable ORF	rRNA processing
YDR413C		1,44	1,51	0,27	-0,26	0,02	0,79	0,46	similarity to NADH dehydrogenase	biological_process unknown
YDR432W	**NPL3**	0,39	0,60	0,61	-1,17	-0,77	0,46	0,43	nucleolar protein	mRNA-nucleus export
YDR433W	**KRE22**	-0,03	0,31	1,30	-0,89	-0,20	0,13	0,43	questionable ORF	biological_process unknown
YDR483W	**KRE2**	-0,34	-0,23	-0,20	1,28	0,05	0,22	-0,17	alpha-1,2-mannosyltransferase	cell wall mannoprotein biosynthesis; O-linked glycosylation; N-glycan processing
YDR494W	**RSM28**	0,94	0,88	-0,36	-0,35	-0,45	0,76	0,43	hypothetical protein	biological_process unknown

YDR537C		-0,08	-0,18	-0,44	1,43	-0,35	0,08	-0,29	questionable ORF	biological_process unknown
YDR540C	**YDR540C**	0,33	-0,06	-0,41	-0,44	-0,50	-0,04	-2,72	hypothetical protein	biological_process unknown
YDR541C	**YDR541C**	-0,44	-0,61	0,00	-0,15	0,33	-0,70	-1,88	weak similarity to dihydroflavonol-4-red	biological_process unknown
YDR543C		-0,55	-0,79	0,13	0,16	0,37	-1,22	-0,45	homology to other subtelomeric encoded p	biological_process unknown
YEL002C	**WBP1**	-0,15	0,08	-0,15	0,09	0,00	0,04	-0,07	oligosaccharyl transferase beta subunit	cell cycle; N-linked glycosylation
YEL045C		-0,18	0,00	2,26	-1,19	-0,18	-0,58	0,13	similarity to cytochrome c oxidase III o	biological_process unknown
YEL066W	**HPA3**	-0,67	-0,80	-0,28	-0,36	0,31	-1,17	-0,54	homology to hypothetical protein YPR193c	histone acetylation
YEL068C		-0,86	-0,87	0,32	0,01	0,72	-1,37	-0,36	hypothetical protein	biological_process unknown
YEL069C	**HXT13**	-0,27	-1,44	-0,43	-0,67	-0,49	-1,37	-0,30	high-affinity hexose transporter	hexose transport
YEL070W	**YEL070W**	0,08	-0,67	-0,29	-1,32	0,11	-0,94	0,20	similarity to R,sphaeroides D-mannitol 2	biological_process unknown
YEL071W	**DLD3**	-0,49	-0,33	-0,54	-5,65	0,15	-0,61	0,91	similarity to Aip2p and Dld1p	lactate metabolism
YEL072W	**RMD6**	-0,33	-0,47	-0,22	-5,48	-0,05	-0,54	0,71	hypothetical protein	biological_process unknown
YEL073C	**YEL073C**	-0,73	-1,48	-0,84	-1,68	-0,81	-1,21	-0,15	hypothetical protein	biological_process unknown
YEL075C	**YEL075C**	-1,22	-0,31	-0,16	-0,05	0,23	-0,44	0,16	homology to other subtelomeric encoded p	biological_process unknown
YEL075W-A		-1,02	-0,33	0,00	-0,39	0,31	-0,77	0,24		
YEL076C	**YEL076C**	-0,36	-0,33	-0,18	-0,34	-0,17	-0,73	-0,15		biological_process unknown
YER023W	**PRO3**	-0,09	0,12	0,55	-1,16	0,52	0,10	-0,12	delta 1-pyrroline-5-carboxylate reductas	proline biosynthesis
YER056C	**FCY2**	0,43	0,30	1,27	0,08	-0,24	0,23	0,25	purine-cytosine permease	cytosine transport; purine transport
YER089C	**PTC2**	0,00	0,29	1,40	-0,10	-1,20	0,01	0,11	similarity to phosphoprotein phosphatase	DNA damage response, signal transduction resulting in cell cycle arrest; response to unfolded protein; inactivation of MAPK during osmolarity sensing; protein amino acid dephosphorylation; regulation of CDK activity; G1/S transition of mitotic cell cycle
YER146W	**LSM5**	-0,08	0,88	1,66	0,23	-0,13	0,14	-0,30	putative snRNA-associated protein	nuclear mRNA splicing, via spliceosome; rRNA processing
YER148W	**SPT15**	0,64	1,60	0,60	0,01	0,61	0,59	-0,01	DNA binding; general RNA polymerase II transcription factor activity; RNA polymerase III transcription factor activity; RNA polymerase I transcription factor activity	transcription initiation from Pol II promoter; transcription initiation from Pol III promoter; transcription from Pol I promoter
YER151C	**UBP3**	0,37	1,15	0,40	0,00	-0,14	0,68	0,22	ubiquitin-specific protease activity	protein deubiquitination
YER153C	**PET122**	0,27	0,00	0,74	-1,37	-0,48	0,59	0,69	translation regulator activity	protein biosynthesis
YER172C	**BRR2**	-0,52	-0,56	0,05	-1,95	-0,76	-0,94	0,05	RNA helicase-related protein	U2-type spliceosome conformational change to release U4 and U1
YER187W	**YER187W**	-1,22	-2,13	-1,19	-0,03	0,12	-1,31	-0,26	similarity to killer toxin KHS precursor	biological_process unknown

YER188W	**YER188W**	-1,05	-1,93	-1,48	-10,61	0,44	-2,68	-3,70	hypothetical protein	biological_process unknown
YER189W	**YER189W**	-1,39	-0,36	-0,28	0,07	0,30	-0,61	0,01	homology to other subtelomeric encoded p	biological_process unknown
YFL010C	**WWM1**	-0,84	-0,86	0,25	0,31	0,53	-1,33	-0,46	questionable ORF	response to dessication
YFL010W-A	**AUA1**	-0,51	-0,42	-0,15	-0,13	-0,09	-0,57	-0,11		amino acid transport
YFL016C	**MDJ1**	-0,53	-0,24	0,44	4,80	0,05	-0,86	-0,44	heat shock protein	proteolysis and peptidolysis; protein folding
YFL017W-A	**SMX2**	-1,24	-0,73	-0,10	-0,23	0,03	-1,66	-0,42		nuclear mRNA splicing, via spliceosome
YFL036W	**RPO41**	-0,48	-0,50	0,20	1,99	-0,50	-0,69	-0,06	mitochondrial DNA-directed RNA polymeras	transcription from mitochondrial promoter; mitochondrial genome maintenance
YFL041W	**FET5**	-0,96	-0,36	-0,04	0,16	0,42	-0,49	-0,19	homology to cell surface ferroxidase pre	iron ion transport
YFL045C	**SEC53**	-0,92	-0,76	0,05	-0,17	0,94	-1,32	-0,39	phosphomannomutase	protein-ER targeting
YFL046W	**FMP32**	-0,89	-0,93	0,02	0,01	0,51	-1,33	-0,18	weak similarity to myosin heavy chain A	biological_process unknown
YFL050C	**ALR2**	-0,46	-0,42	-0,30	-0,39	-0,02	-0,73	-1,02	aluminum resistance protein	di-, tri-valent inorganic cation transport; magnesium ion transport
YFL051C	**YFL051C**	-0,66	-0,24	0,28	0,14	0,47	-0,89	-0,44	homology to flocculation Flo1p/putative	biological_process unknown
YFL052W	**YFL052W**	-0,46	-0,42	-0,30	-0,39	-0,02	-0,73	-1,02	homology to Mal63p, Mal23p and Mal33p	biological_process unknown
YFL053W	**DAK2**	-0,98	-1,44	-0,18	-0,41	0,91	-1,94	-2,66	similarity to C,freundii dihydroxyaceton	glycerol catabolism; response to stress
YFL054C	**YFL054C**	-0,55	-4,68	-0,61	-5,56	0,32	-5,41	-2,19	similarity to channel proteins	water transport
YFL055W	**AGP3**	-0,32	-3,14	-0,29	-2,97	0,08	-2,91	-1,63	similarity to Gap1p and other amino acid	amino acid transport
YFL056C	**AAD6**	-0,40	-3,68	-0,41	-3,83	-0,09	-3,89	-1,95	homology to aryl-alcohol dehydrogenases	aldehyde metabolism
YFL057C	**AAD16**	-0,67	-6,64	-0,21	-6,97	0,40	-6,81	-3,71	homology to aryl-alcohol dehydrogenases	aldehyde metabolism
YFL058W	**THI5**	-0,89	-0,69	-0,22	-0,30	0,59	-1,24	-0,03	pyrimidine biosynthesis protein	thiamin biosynthesis
YFL059W	**SNZ3**	-0,19	-0,68	-0,77	-1,34	0,78	-0,86	-0,14	homology to YNL333w and YMR096w	pyridoxine metabolism; thiamin biosynthesis
YFL061W	**YFL061W**	-0,54	-0,57	-0,25	-1,20	0,54	-0,84	-0,79	homology to M,verrucaria cyanamide hydra	biological_process unknown
YFL063W		-0,75	-1,35	0,24	0,07	0,23	-1,55	-0,38	homology to other subtelomeric encoded p	biological_process unknown
YFL064C	**YFL064C**	-1,12	-0,23	-0,09	0,00	-0,10	-0,44	0,40	homology to other subtelomeric encoded p	biological_process unknown
YFL065C	**YFL065C**	-1,00	-0,71	0,37	-0,04	0,08	-0,47	-0,07	homology to other subtelomeric encoded p	biological_process unknown
YFL066C	**YFL066C**	-1,14	-0,55	0,15	-0,09	0,20	-0,52	0,02	homology to other subtelomeric encoded p	biological_process unknown
YFL068W	**YFL068W**	-1,06	0,13	-0,20	-0,06	0,04	-0,51	0,13	hypothetical protein in Y' repeat region	biological_process unknown
YFL-TYA		0,22	0,71	0,34	0,18	-0,68	0,72	-1,53		

YFL-TYB		-0,08	0,41	0,32	0,11	-0,50	0,24	-1,13		
YFR019W	**FAB1**	-0,53	-0,10	-0,04	1,24	0,24	-0,49	-0,33	putative PIP 5-kinase	phospholipid metabolism; vacuole organization and biogenesis; response to stress
YFR026C	**YFR026C**	-0,68	-0,16	0,11	-1,54	-0,09	-1,10	-2,09	hypothetical protein	biological_process unknown
YFR033C	**QCR6**	0,16	0,36	0,54	-0,74	0,27	0,22	0,95	ubiquinol--cytochrome-c reductase 17K pr	mitochondrial electron transport, ubiquinol to cytochrome c; aerobic respiration
YFR037C	**RSC8**	0,37	1,38	0,29	-0,04	-0,06	0,49	-0,19	molecular_function unknown	chromatin remodeling
YFR054C		-0,29	-0,20	-0,25	-0,35	-0,22	-0,43	0,00	hypothetical protein	biological_process unknown
YFR055W	**YFR055W**	-0,85	-0,52	0,32	0,29	0,62	-1,30	-1,51	homology to beta-cystathionases	sulfur metabolism; copper ion homeostasis
YFR056C		-1,05	-0,68	0,37	0,38	-0,03	-0,84	-2,43	questionable ORF	biological_process unknown
YFR057W	**YFR057W**	-0,28	-0,51	-0,15	-0,45	-3,81	-0,36	-2,73	hypothetical protein	biological_process unknown
YGL034C		-0,42	-0,38	1,21	0,31	0,14	-0,36	-0,37	questionable ORF	biological_process unknown
YGL043W	**DST1**	-1,75	-0,41	3,07	1,54	-0,15	0,00	0,00	transcription elongation factor S-II	meiotic recombination; RNA elongation from Pol II promoter
YGL051W	**MST27**	-0,90	-0,18	-3,20	-0,68	0,10	-0,62	-2,21	similarity to FUN59 protein	vesicle organization and biogenesis
YGL052W	**YGL052W**	-1,12	-0,51	-4,37	-0,14	0,28	-0,79	-3,20	questionable ORF	biological_process unknown
YGL053W	**PRM8**	-1,40	-0,88	-2,85	-1,20	-0,10	-2,10	-2,32	homology to hypothetical protein YAR031	conjugation with cellular fusion
YGL090W	**LIF1**	0,85	0,61	-0,44	-0,37	-0,21	0,59	0,65	hypothetical protein	double-strand break repair via nonhomologous end-joining
YGL109W		0,87	0,78	-0,06	0,04	-0,33	0,65	0,26	questionable ORF	biological_process unknown
YGL115W	**SNF4**	1,02	1,15	-0,42	-0,13	-0,34	1,18	0,30	nuclear regulatory protein	peroxisome organization and biogenesis; regulation of transcription from Pol II promoter
YGL123W	**RPS2**	0,93	1,61	0,27	0,29	-0,10	1,19	0,55	ribosomal protein	regulation of translational fidelity; protein biosynthesis
YGL124C	**MON1**	0,84	0,70	0,65	0,30	0,08	1,21	0,21	hypothetical protein	protein-vacuolar targeting; autophagy
YGL125W	**MET13**	0,99	1,58	0,21	0,27	0,03	1,46	0,57	putative methylenetetrahydrofolate reduc	methionine metabolism; sulfur amino acid biosynthesis; protein biosynthesis
YGL126W	**SCS3**	0,40	-0,30	0,67	-0,37	0,26	0,18	1,55	inositol phospholipid synthesis protein	phospholipid metabolism
YGL127C	**SOH1**	0,82	1,19	0,51	0,12	-0,18	1,23	1,25	allows hpr1 null mutant to grow at 37 de	DNA repair; transcription from Pol II promoter
YGL128C	**CWC23**	0,76	1,27	0,66	0,26	0,14	1,14	0,54	molecular_function unknown	biological_process unknown
YGL129C	**RSM23**	0,80	1,68	0,57	0,46	-0,29	1,17	0,50	structural constituent of ribosome	protein biosynthesis
YGL130W	**CEG1**	1,06	0,00	0,68	0,31	0,25	1,21	0,38	mRNA guanylyltransferase (mRNA capping	mRNA capping
YGL188C		-0,06	0,00	-0,05	-0,10	0,31	0,24	-0,23	molecular_function unknown	biological_process unknown
YGL204C		-0,78	-0,52	-0,26	-1,01	0,34	-1,39	-0,84	questionable ORF	biological_process unknown

YGL220W	**YGL220W**	0,93	0,79	-0,13	-0,38	-0,77	0,66	0,55	weak similarity to V,alginolyticus bolA	biological_process unknown
YGL226W	**YGL226W**	-0,88	-1,31	0,20	0,12	0,96	-1,46	-0,44	weak similarity to N,crassa cytochrome-c	biological_process unknown
YGL235W	**YGL235W**	-0,63	-0,66	-0,16	0,12	0,18	-0,53	-1,02	questionable ORF	biological_process unknown
YGL236C	**MTO1**	-0,18	-0,61	-0,03	-0,13	0,23	-0,31	-0,81	homology to gidA E,coli protein	protein biosynthesis
YGL237C	**HAP2**	-0,33	-0,44	-0,42	-0,11	-0,46	-0,11	-0,81	component of heterotrimeric CCAAT-bindin	regulation of carbohydrate metabolism; transcription
YGL255W	**ZRT1**	-0,70	-1,06	-0,89	-0,23	0,20	-1,19	-0,34	high-affinity zinc transport protein	high-affinity zinc ion transport
YGL262W	**YGL262W**	-1,49	0,07	1,11	-0,82	1,56	-1,78	-3,00	similarity to hypothetical protein YER18	biological_process unknown
YGL262W	**YGL262W**	-1,49	0,07	1,11	-0,82	1,56	-1,78	-3,00	similarity to hypothetical protein YER18	biological_process unknown
YGL263W	**COS12**	-1,22	0,10	0,79	0,13	1,71	-0,99	-2,09	similarity to other subtelomeric encoded	biological_process unknown
YGR008C	**STF2**	-0,04	-0,40	-0,32	-0,18	0,14	0,22	-1,21	ATPase stabilizing factor	ATP synthesis coupled proton transport; response to dessication
YGR029W	**ERV1**	-0,15	0,20	0,04	1,25	0,59	-0,49	-0,24	mitochondrial biogenesis and cell viabil	mitochondrion organization and biogenesis; iron ion homeostasis
YGR030C	**POP6**	-0,40	0,12	1,22	-0,09	-0,13	-0,16	-0,03	hypothetical protein	tRNA processing; rRNA processing
YGR031W	**YGR031W**	-0,36	0,10	1,31	0,53	-0,93	-0,46	-0,11	similarity to hypothetical protein YGR01	biological_process unknown
YGR035C	**YGR035C**	0,04	0,30	-0,25	1,26	0,24	-0,58	-0,19	hypothetical protein	biological_process unknown
YGR049W	**SCM4**	1,23	0,97	0,07	-0,44	-0,51	1,15	0,35	cdc4 suppressor	cell cycle
YGR050C	**YGR050C**	0,42	0,23	-1,05	-1,17	-1,24	0,36	0,01	questionable ORF	biological_process unknown
YGR051C		0,18	0,32	1,88	-0,10	0,39	0,19	0,15	questionable ORF	biological_process unknown
YGR054W	**YGR054W**	0,13	0,51	0,11	0,35	0,03	0,64	-0,94	hypothetical protein	translational initiation
YGR062C	**COX18**	0,58	1,68	0,53	0,43	0,49	1,20	0,09	hypothetical protein	cytochrome c oxidase biogenesis
YGR063C	**SPT4**	0,13	1,13	1,06	0,50	0,27	0,89	-0,06	transcription initiation protein	regulation of transcription, DNA-dependent; chromosome segregation; establishment and/or maintenance of chromatin architecture; RNA elongation from Pol II promoter
YGR064W		0,05	1,26	0,81	0,49	0,41	1,06	0,06	molecular_function unknown	biological_process unknown
YGR065C	**VHT1**	0,57	1,38	0,55	0,26	0,02	1,24	0,14	similarity to hypothetical protein YAL06	biotin transport
YGR066C	**YGR066C**	0,58	1,69	0,60	0,23	-0,04	1,25	-0,04	similarity to hypothetical protein YBR10	biological_process unknown
YGR067C	**YGR067C**	0,89	1,65	0,43	0,34	-0,43	1,68	0,20	hypothetical protein	biological_process unknown
YGR068C	**YGR068C**	0,22	1,46	0,58	0,31	-0,32	1,38	0,16	weak similarity to Rod1p	biological_process unknown
YGR069W		0,09	1,30	0,66	0,37	0,64	1,09	0,03	molecular_function unknown	biological_process unknown

YGR071C	**YGR071C**	0,59	1,43	0,25	0,21	0,31	0,98	0,20	molecular_function unknown	biological_process unknown
YGR073C		0,73	1,32	-0,08	-0,41	0,04	1,21	-0,14	questionable ORF	biological_process unknown
YGR074W	**SMD1**	0,79	1,33	-0,41	-0,51	-0,43	1,39	-0,16	snRNA-associated protein	nuclear mRNA splicing, via spliceosome
YGR075C	**PRP38**	0,97	1,30	0,15	-0,24	0,31	1,11	0,18	pre-mRNA splicing factor	nuclear mRNA splicing, via spliceosome
YGR081C	**SLX9**	1,08	0,99	-0,09	0,86	-0,09	0,08	0,02	hypothetical protein	DNA metabolism
YGR086C	**PIL1**	1,07	0,91	-0,19	-0,40	-0,60	1,00	0,26	similarity to hypothetical protein YPL00	biological_process unknown
YGR105W	**VMA21**	-0,15	-0,34	0,08	0,12	1,17	-0,04	-0,02	vacuolar ATPase assembly integral membra	protein complex assembly
YGR114C		0,51	0,58	-0,47	-0,42	-1,19	0,91	0,38	questionable ORF	biological_process unknown
YGR116W	**SPT6**	1,50	-0,13	-0,04	0,59	-0,26	0,11	-0,01	transcription initiation protein	nucleosome assembly; regulation of transcription, DNA-dependent; establishment and/or maintenance of chromatin architecture; RNA elongation from Pol II promoter
YGR129W	**SYF2**	0,88	0,92	-0,42	0,03	-0,40	0,67	0,36	hypothetical protein	nuclear mRNA splicing, via spliceosome; cell cycle
YGR133W	**PEX4**	0,89	1,15	-0,10	-0,59	-0,24	0,94	-0,01	ubiquitin-conjugating enzyme	peroxisome organization and biogenesis; protein monoubiquitination; protein polyubiquitination
YGR146C	**YGR146C**	0,78	1,14	-0,30	-0,44	-0,52	1,30	0,30	hypothetical protein	biological_process unknown
YGR152C	**RSR1**	0,03	0,67	0,24	-0,29	-0,39	1,27	-0,04	GTP-binding protein	polar budding; axial budding; bud site selection; small GTPase mediated signal transduction
YGR153W	**YGR153W**	0,77	1,15	0,15	-0,24	0,02	1,06	0,08	molecular_function unknown	biological_process unknown
YGR161C	**RTS3**	0,69	1,50	0,27	-0,09	-0,25	1,22	-0,03	hypothetical protein	protein amino acid dephosphorylation
YGR203W	**YGR203W**	0,52	0,19	1,65	-1,29	0,15	0,67	0,23	similarity to hypothetical protein YGR20	biological_process unknown
YGR235C	**YGR235C**	-0,28	-0,34	0,38	0,32	0,17	-0,33	-1,27	hypothetical protein	biological_process unknown
YGR236C	**SPG1**	0,06	0,00	-0,13	-0,21	0,06	0,35	-1,22	questionable ORF	biological_process unknown
YGR288W	**MAL13**	-0,42	-0,29	0,09	-6,12	-0,32	-5,22	-2,43	maltose pathway regulatory protein	regulation of transcription, DNA-dependent; carbohydrate metabolism
YGR289C	**MAL11**	-0,23	0,07	0,23	-4,39	-0,04	-4,15	-2,69	general alpha-glucoside permease	alpha-glucoside transport; trehalose transport
YGR290W	**YGR290W**	-0,12	0,08	-0,07	-4,07	-0,06	-3,27	-1,75	hypothetical protein	biological_process unknown
YGR295C	**COS6**	-0,28	-0,57	-0,30	-0,84	0,32	-0,63	-1,05	homology to other subtelomeric encoded p	biological_process unknown
YGR296W	**YRF1-3**	-1,09	-0,34	-0,11	-0,05	-0,09	-0,42	-0,06	Y' long ORF with intron	telomerase-independent telomere maintenance
YHL008C	**YHL008C**	-1,15	-1,92	-0,72	-0,97	-0,94	-1,88	-1,39	similarity to M,formicicum formate dehyd	biological_process unknown
YHL009C	**YAP3**	-0,53	-0,52	0,04	-0,26	-0,19	-0,39	-1,70	bZip DNA binding protein	regulation of transcription from Pol II promoter
YHL010C	**YHL010C**	-0,29	-0,27	0,19	-0,02	-0,07	-0,18	-1,00	similarity to unknown C,elegans protein	biological_process unknown

YHL012W	**YHL012W**	-0,34	-0,53	-0,39	-0,50	-0,13	-0,27	-1,60	homology to UDP Glucose pyrophosphorylas	biological_process unknown
YHL013C	**YHL013C**	0,26	-0,37	-0,08	-0,69	0,30	0,17	-0,90	hypothetical protein	biological_process unknown
YHL014C	**YLF2**	-0,12	-0,14	-0,17	-0,11	-0,26	-0,05	-1,24	putative GTP-binding protein	biological_process unknown
YHL017W	**YHL017W**	-0,47	-0,50	0,77	0,20	0,21	-0,44	-1,06	putative transmembrane protein PTM1 homo	biological_process unknown
YHL018W	**YHL018W**	0,54	0,35	-0,10	-0,01	-1,21	0,57	-0,74	weak similarity to human pterin-4-alpha-	biological_process unknown
YHL019C	**APM2**	-0,16	-0,28	0,33	0,06	0,08	-0,36	-0,90	homology to clathrin AP medium chain AP5	vesicle-mediated transport
YHL020C	**OPI1**	-0,26	-0,26	-0,09	-0,10	-0,33	-0,29	-0,87	negative regulator of phospholipid biosy	negative regulation of transcription from Pol II promoter; phospholipid biosynthesis; positive regulation of transcription from Pol II promoter
YHL025W	**SNF6**	-0,54	-0,53	1,87	-0,64	-0,63	-0,89	-0,10	global transcription activator	chromatin remodeling
YHL041W		-0,64	-1,08	0,60	0,18	0,50	-1,48	-0,34	weak similarity to Drosophila hypothetic	biological_process unknown
YHL045W		-0,20	0,00	0,16	-0,07	-0,01	-0,45	0,11	molecular_function unknown	biological_process unknown
YHL047C	**ARN2**	-0,56	-0,95	0,45	0,25	0,67	-0,81	-1,16	putative multidrug resistance protein	iron-siderochrome transport; iron ion homeostasis; siderochrome metabolism
YHL048W	**COS8**	-0,79	-1,01	-0,41	-0,48	0,16	-1,06	-1,44	homology to other subtelomeric encoded protein	response to unfolded protein
YHL049C	**YHL049C**	-1,01	-0,52	0,08	-0,02	0,26	-0,30	0,30	hypothetical protein in Y' repeat region	biological_process unknown
YHL050C	**YHL050C**	-1,16	-0,04	0,20	-0,04	0,38	-0,63	0,01	hypothetical protein in Y' repeat region	biological_process unknown
YHR033W	**YHR033W**	-0,03	-0,26	0,50	0,14	0,83	-0,58	-0,19	putative glutamate 5-kinase	biological_process unknown
YHR043C	**DOG2**	-0,17	-0,04	0,54	0,02	1,00	-0,64	0,10	2-deoxyglucose-6-phosphate phosphatase	response to stress; glucose metabolism
YHR053C	**CUP1-1**	-1,64	-1,32	-3,36	-4,04	-3,65	-1,19	-2,39	metallothionein	response to copper ion
YHR054C	**YHR054C**	-1,61	-1,31	-3,37	-4,28	-3,64	-2,00	-1,45	weak similarity to YOR262w	biological_process unknown
YHR055C	**CUP1-2**	-1,69	-1,25	-3,54	-4,01	-3,99	-1,18	-2,39	metallothionein	response to copper ion
YHR056C	**RSC30**	-0,71	-0,78	-1,25	-1,43	-1,29	-0,84	-0,93	weak similarity to YHR054c	regulation of transcription, DNA-dependent
YHR117W	**TOM71**	0,29	0,18	-0,01	0,13	-0,30	0,39	0,30	similarity to Tom70p/Mas70p	biological_process unknown
YHR142W	**CHS7**	-0,07	-0,10	0,42	-0,08	-0,07	0,04	0,04	putative transmembrane protein	ER to Golgi transport; cell wall chitin biosynthesis
YHR164C	**DNA2**	-0,46	-0,58	0,07	0,42	0,22	-0,40	-0,27	DNA helicase	DNA dependent DNA replication; DNA repair; lagging strand elongation; replicative cell aging
YHR180W		0,32	0,59	0,25	-3,97	-0,15	0,70	-0,63	hypothetical protein	biological_process unknown
YHR181W	**SVP26**	-0,43	-0,42	0,45	0,22	0,58	-0,42	-0,81	homology to mouse TEG-261 protein	biological_process unknown
YHR182W	**YHR182W**	-0,49	-0,33	0,89	0,28	-0,11	-0,45	-1,47	hypothetical protein	biological_process unknown
YHR207C	**SET5**	-0,50	-0,70	0,24	-0,12	0,52	-0,73	-0,85	hypothetical protein	biological_process unknown

YHR209W	**YHR209W**	-0,99	-0,84	-0,09	0,07	0,38	-1,16	-1,81	homology to hypothetical protein YER175c	biological_process unknown
YHR210C	**YHR210C**	-0,26	-0,25	-0,31	-0,28	-0,01	-0,62	-1,35	putative UDP-glucose-4-epimerase	biological_process unknown
YHR212C	**YHR212C**	-2,74	-1,00	-1,37	-6,86	-5,96	-5,81	-2,22	homology to YAR060c	biological_process unknown
YHR213W	**YHR213W**	-0,95	-0,82	-0,09	-1,58	-0,97	-2,19	-1,72	homology to Flo1p/putative pseudogene	biological_process unknown
YHR214C-B	**YHR214C-B**	-1,17	-0,62	0,27	-1,85	-2,05	-1,79	-0,67		Ty element transposition
YHR214W-A		0,00	0,33	0,75	0,31	1,71	-0,36	-0,98		biological_process unknown
YHR215W	**PHO12**	0,21	0,55	0,74	0,58	1,29	-0,21	-0,83	secreted acid phosphatase	biological_process unknown
YIL014C-A	**YIL014C-A**	-0,90	-0,26	-3,03	-3,24	-0,42	-3,66	0,20		biological_process unknown
YIL029C	**YIL029C**	-0,71	-0,79	0,20	0,17	-2,58	-0,66	-2,19		biological_process unknown
YIL070C	**MAM33**	-0,04	-0,05	-1,00	-0,47	-0,50	0,02	0,28	hypothetical protein	aerobic respiration
YIL080W	**YIL080W**	-1,53	-1,47	0,90	-0,91	-0,46	-1,66	-1,88	Ty3-2 orf C fragment	Ty element transposition
YIL082W	**YIL082W**	-3,45	-0,88	0,87	-0,76	-0,24	-1,37	-3,07	Ty3A protein	biological_process unknown
YIL082W-A	**YIL082W-A**	-1,53	-0,76	0,30	-0,49	0,01	-0,82	-1,26		Ty element transposition
YIL120W	**QDR1**	-0,58	-0,86	-0,15	-0,10	-0,07	-0,76	-0,87	homology to antibiotic resistance protei	multidrug transport
YIL139C	**REV7**	-0,27	-0,16	0,30	0,10	1,13	-0,60	-0,07	required for DNA damage induced mutagene	DNA repair; mutagenesis
YIL153W	**RRD1**	-0,07	-0,04	1,03	0,01	-0,24	-0,30	0,05	similarity to phosphotyrosyl phosphatase	response to osmotic stress; mitotic spindle assembly (sensu Saccharomyces); DNA repair
YIL167W	**YIL167W**	-0,50	-1,64	-0,11	-0,23	0,31	-0,51	-0,26	molecular_function unknown	biological_process unknown
YIL168W	**SDL1**	-0,52	-4,68	-0,45	-0,43	-0,47	-0,39	-0,22	L-serine ammonia-lyase activity	serine family amino acid metabolism
YIL169C	**YIL169C**	-1,01	-2,31	-0,26	-0,66	0,81	-1,74	-0,44	homology to glucan 1,4-alpha-glucosidas	biological_process unknown
YIL170W	**HXT12**	-1,81	-1,75	-1,53	-1,83	0,32	-1,69	-0,56	putative pseudogene; homology to sugar t	biological_process unknown
YIL171W	**HXT12**	-1,39	-1,36	-1,18	-1,37	0,58	-1,43	-0,54	putative pseudogene; homology to sugar t	biological_process unknown
YIL172C	**YIL172C**	-0,88	-0,92	-2,12	-1,50	0,23	-1,10	-0,66	identical to Fsp2p	biological_process unknown
YIL174W	**YIL174W**	-0,53	-0,86	-0,95	-0,05	-0,94	-0,44	-0,14	putative pseudogene	biological_process unknown
YIR030C	**DCG1**	0,91	0,63	-0,34	-0,53	-0,24	0,46	0,52	involved in nitrogen-catabolite metaboli	nitrogen metabolism
YIR039C	**YPS6**	-0,80	-0,81	-0,30	-0,09	0,00	-1,18	-0,33	similarity to Yap3p	biological_process unknown
YIR041W	**YIR041W**	-0,83	-1,07	-0,98	-0,48	-0,52	-1,05	-0,63	similarity to YIL176c,YIL011w and other	biological_process unknown
YIR042C	**YIR042C**	-2,47	-2,35	-1,84	-0,45	-1,25	-2,18	-2,58	hypothetical protein	biological_process unknown

YIR043C	**YIR043C**	-0,70	-0,90	-0,95	-0,38	-0,20	-1,00	-1,10	putative pseudogene	biological_process unknown
YJL002C	**OST1**	0,04	0,22	0,07	1,94	-0,09	-0,15	-0,36	oligosaccharyltransferase, alpha subuni	N-linked glycosylation via asparagine; N-linked glycosylation
YJL020C	**BBC1**	0,43	0,64	0,47	-0,10	-0,29	0,56	-0,92	hypothetical protein	actin cytoskeleton organization and biogenesis
YJL022W		0,40	0,44	0,08	-0,48	-0,60	0,76	-2,14	questionable ORF	biological_process unknown
YJL023C	**PET130**	0,00	0,58	-0,13	-0,33	0,17	0,71	-0,85	mitochondrial protein synthesis protein	biological_process unknown
YJL028W		0,45	-0,65	0,89	-0,74	0,25	0,50	0,89	hypothetical protein	biological_process unknown
YJL043W	**YJL043W**	-0,09	0,17	1,26	-0,16	-0,48	-0,16	-0,04	similarity to YKR015c	biological_process unknown
YJL114W	**YJL114W**	-3,66	-0,17	-1,57	-4,98	1,03	-2,08	-3,00	TY4A protein	Ty element transposition
YJL142C		-0,11	0,01	0,10	0,18	0,09	-0,16	-1,06	questionable ORF	biological_process unknown
YJL182C		-0,77	-1,20	-0,02	-0,14	0,11	-1,09	-0,50	molecular_function unknown	biological_process unknown
YJL211C		-0,52	-0,46	-0,14	-0,23	0,28	-0,66	-1,07	questionable ORF	biological_process unknown
YJL214W	**HXT8**	-0,62	-1,12	-0,30	-0,31	-4,29	-1,04	-2,75	hexose transport protein	hexose transport
YJL215C	**YJL215C**	-0,55	-1,18	0,15	-0,22	-4,11	-0,88	-2,66	hypothetical protein	biological_process unknown
YJL216C	**YJL216C**	-0,39	-0,40	-0,58	-0,59	-2,32	-0,37	-1,63	similarity to Mal62p	biological_process unknown
YJL217W	**YJL217W**	-3,22	-0,31	-4,75	-0,37	-4,53	0,07	-3,13	hypothetical protein	biological_process unknown
YJL218W	**YJL218W**	-4,17	-0,85	-4,79	0,24	-0,45	-0,86	-3,19	similarity to E,coli galactoside O-acety	biological_process unknown
YJL219W	**HXT9**	-1,54	-0,78	-1,22	-1,60	0,26	-1,43	-0,73	hexose transport protein	hexose transport
YJL220W	**YJL220W**	-1,24	-1,05	-1,63	-1,68	-0,08	-1,24	-0,61	homology to hypothetical protein YIL172w	biological_process unknown
YJL221C	**FSP2**	-1,13	-0,27	-2,04	-1,32	0,46	-0,98	-0,86	homology to alpha-D-glucosidase	biological_process unknown
YJR019C	**TES1**	-0,57	0,43	-0,33	-1,32	-0,13	0,31	0,28	hypothetical protein	fatty acid oxidation
YJR020W		-0,08	-0,09	-0,03	-2,08	-0,14	-0,08	-0,30	questionable ORF	biological_process unknown
YJR026W	**YJR026W**	-1,26	-0,54	0,14	-3,60	-2,44	-2,37	-0,98	TY1A protein	Ty element transposition
YJR027W	**YJR027W**	-0,40	-0,43	0,18	-1,19	-0,97	-0,97	-0,75	TY1B protein	Ty element transposition
YJR028W	**YJR028W**	-1,32	-0,66	0,15	-3,71	-2,83	-2,33	-1,01	TY1A protein	Ty element transposition
YJR029W	**YJR029W**	-0,39	-0,55	0,16	-1,42	-1,23	-1,15	-0,63	TY1B protein	Ty element transposition
YJR032W	**CPR7**	1,21	0,75	-0,12	-0,20	-0,01	0,37	0,32	similarity to peptidylprolyl isomerase	response to stress
YJR038C		0,85	0,72	0,17	-0,42	-0,18	0,60	0,51	questionable ORF	biological_process unknown

YJR044C	**VPS55**	0,89	0,62	-0,18	-0,43	-0,64	0,98	0,38	similarity to putative transport protein	late endosome to vacuole transport
YJR056C	**YJR056C**	1,04	0,70	0,20	-0,16	-0,49	1,18	0,52	hypothetical protein	biological_process unknown
YJR073C	**OPI3**	0,94	0,93	0,13	-0,03	-0,33	0,75	0,62	methylene-fatty-acyl-phospholipid syntha	phosphatidylcholine biosynthesis
YJR104C	**SOD1**	1,21	1,26	0,57	0,11	-0,11	0,70	0,44	copper-zinc superoxide dismutase	copper ion homeostasis; superoxide metabolism; zinc ion homeostasis
YJR110W	**YJR110W**	0,77	1,21	0,77	0,53	0,02	0,76	-0,19	phosphoric monoester hydrolase activity	biological_process unknown
YJR111C	**YJR111C**	0,33	1,34	0,65	0,51	0,25	0,82	-0,52	molecular_function unknown	biological_process unknown
YJR113C	**RSM7**	-0,23	0,76	1,56	0,59	-0,92	0,68	-0,58	putative mitochondrial ribosomal protein	protein biosynthesis
YJR114W	**SRF2**	-0,20	0,63	1,50	0,68	-0,44	0,58	-0,62	questionable ORF	biological_process unknown
YJR116W	**YJR116W**	0,63	1,31	0,48	0,27	-0,23	1,07	-0,41	molecular_function unknown	biological_process unknown
YJR117W	**STE24**	0,62	1,49	0,67	0,53	0,30	0,83	-0,58	metalloendopeptidase activity; prenyl-dependent CAAX protease activity	peptide pheromone maturation
YJR118C	**ILM1**	0,51	1,35	0,30	0,40	0,28	0,69	-0,60	molecular_function unknown	biological_process unknown
YJR119C	**YJR119C**	0,55	0,80	-0,03	0,00	-0,51	0,69	-0,98	similarity to human retinoblastoma bindi	biological_process unknown
YJR122W	**CAF17**	0,32	0,70	0,79	0,64	0,35	0,26	-1,52	hypothetical protein	biological_process unknown
YJR124C	**YJR124C**	0,46	0,91	0,63	0,20	-0,16	0,39	-1,06	weak similarity to hexose transport prot	biological_process unknown
YJR126C	**VPS70**	0,54	0,66	0,53	0,00	0,13	0,15	-0,99	weak similarity to transferrin receptor	protein-vacuolar targeting
YJR127C	**ZMS1**	0,04	0,23	0,41	0,22	0,00	-0,01	-1,06	putative regulatory protein	biological_process unknown
YJR128W		0,01	0,23	0,22	0,04	-0,15	-0,05	-1,21	questionable ORF	biological_process unknown
YJR129C	**YJR129C**	0,07	0,34	0,07	-0,14	-0,14	0,11	-1,18	similarity to hypothetical proteins YNL0	biological_process unknown
YJR152W	**DAL5**	-0,67	-0,95	-0,03	0,17	-5,79	-0,87	-0,22	allantoate and ureidosuccinate permease	allantoate transport
YJR153W	**PGU1**	-0,38	-0,19	-0,65	-0,29	-5,66	-0,56	0,01	putative polygalacturonase	pectin catabolism; pseudohyphal growth
YJR154W	**YJR154W**	0,10	-0,23	-0,88	-0,62	-2,63	-0,10	0,20	hypothetical protein	biological_process unknown
YJR155W	**AAD10**	-0,30	-0,62	-1,05	-0,23	-2,49	-0,47	-0,10	putative aryl-alcohol dehydrogenase	aldehyde metabolism
YJR156C	**THI11**	-0,42	-1,13	-0,33	-0,31	0,85	-1,62	0,11	homology to Thi5p, Nmt1p and YDL244w	thiamin biosynthesis
YJR157W	**YJR157W**	-0,29	-4,51	-4,98	-1,21	-5,73	-2,21	-3,19	hypothetical protein	biological_process unknown
YJR158W	**HXT16**	-0,30	-3,48	-2,24	-0,96	-1,66	-2,33	-1,32	homology to sugar transport protein	hexose transport
YJR159W	**SOR1**	-0,67	-6,07	-5,68	-0,58	-5,97	-2,77	-3,86	sorbitol dehydrogenase	mannose metabolism; fructose metabolism
YJR160C	**MPH3**	-0,32	-2,56	-1,92	-0,66	-1,70	-2,12	-1,75	homology to Mal3Tp	carbohydrate transport

YKL024C	**URA6**	-0,29	0,00	-0,06	0,01	0,32	-0,08	0,95	uridine-monophosphate kinase	nucleobase, nucleoside, nucleotide and nucleic acid metabolism
YKL078W	**DHR2**	0,95	1,10	0,38	0,01	-0,13	0,78	0,24	similarity to ATP-dependent RNA helicase	ribosome biogenesis
YKL080W	**VMA5**	1,09	1,51	0,33	-0,07	-0,36	1,25	0,34	vacuolar H+-transporting ATPase,chain C	vacuolar acidification
YKL081W	**TEF4**	0,67	1,35	0,44	0,06	0,60	1,06	0,07	translation elongation factor activity	translational elongation
YKL082C	**RRP14**	0,76	1,58	0,68	0,35	-0,04	1,18	0,47	hypothetical protein	establishment of cell polarity (sensu Saccharomyces)
YKL084W	**YKL084W**	0,36	1,54	1,00	0,76	0,04	0,91	0,32	hypothetical protein	biological_process unknown
YKL085W	**MDH1**	0,61	1,34	0,48	0,22	0,40	1,12	0,23	L-malate dehydrogenase activity	malate metabolism; tricarboxylic acid cycle
YKL087C	**CYT2**	0,32	1,25	1,13	0,05	0,23	1,18	0,30	holocytochrome-c synthase (cytochrome-c1	cytochrome c-heme linkage
YKL088W	**YKL088W**	0,71	1,62	0,33	-0,05	-0,26	1,06	0,37	purine nucleotide binding; phosphopantothenoylcysteine decarboxylase activity	salinity response; coenzyme A biosynthesis
YKL089W	**MIF2**	0,79	1,45	0,49	0,06	0,20	1,06	0,29	centromeric DNA binding	mitotic spindle assembly (sensu Saccharomyces)
YKL090W	**CUE2**	1,05	1,57	-0,04	-0,22	-0,19	0,97	0,41	hypothetical protein	biological_process unknown
YKL091C	**YKL091C**	0,56	1,18	0,66	0,53	0,12	0,92	0,57	molecular_function unknown	biological_process unknown
YKL097C	**YKL097C**	0,38	1,26	1,47	-0,89	0,09	0,21	0,77	hypothetical protein	biological_process unknown
YKL111C		0,84	1,60	-0,18	-0,31	-0,51	0,69	0,30	questionable ORF	biological_process unknown
YKL213C	**DOA1**	-0,08	0,13	-0,32	-0,19	-4,57	-0,07	0,13	involved in ubiquitin-dependent proteoly	ubiquitin-dependent protein catabolism; double-strand break repair via nonhomologous end-joining
YKL221W	**MCH2**	-0,38	-0,63	-0,01	1,62	-5,77	-0,06	-0,10	similarity to monocarboxylate transporte	transport
YKL222C	**YKL222C**	0,26	-0,81	0,04	1,73	-2,56	-0,28	-0,15	putative transcription factor protein	biological_process unknown
YKL225W		-0,65	-1,10	0,13	-0,07	-0,10	-1,17	-0,48	homology to other subtelomeric encoded protein	biological_process unknown
YKR102W	**FLO10**	-0,08	-1,73	-0,37	-0,02	-0,36	-0,09	0,12	molecular_function unknown	biological_process unknown
YKR103W	**NFT1**	-0,40	-1,32	0,09	0,09	0,09	-0,39	-0,14	ATP-binding cassette (ABC) transporter activity	transport
YKR104W		-0,52	-4,72	-0,53	-0,42	-0,10	-0,43	-0,01	not yet annotated	biological_process unknown
YKR105C	**YKR105C**	-0,34	-2,31	-0,15	0,56	0,43	-0,81	0,07	molecular_function unknown	biological_process unknown
YKR106W	**YKR106W**	-0,56	-2,18	-0,36	-0,05	-0,54	-0,62	-0,19	transporter activity	transport
YLL053C	**YLL053C**	-0,40	-0,67	-0,27	0,75	0,00	-0,26	-1,44	putative water channel protein	biological_process unknown
YLL063C	**AYT1**	-0,67	-0,39	-0,09	-4,59	0,50	-4,81	0,14	hypothetical protein	secondary metabolism
YLR038C	**COX12**	-0,34	-0,05	-0,05	1,85	-0,10	-0,42	-0,45	cytochrome-c oxidase, subunit VIB	cytochrome c oxidase biogenesis

YLR056W	**ERG3**	0,69	0,86	0,29	-0,55	0,35	0,69	0,75	C-5 sterol desaturase	ergosterol biosynthesis
YLR066W	**SPC3**	-0,33	0,22	0,74	0,58	1,30	0,15	-0,10	similarity to signal peptidase	signal peptide processing
YLR067C	**PET309**	-0,14	0,08	1,01	0,81	0,41	-0,01	-0,24	required for stability and translation o	aerobic respiration; RNA metabolism; protein biosynthesis
YLR087C	**CSF1**	0,91	0,29	-0,16	0,05	-0,24	0,23	0,19	hypothetical protein	fermentation
YLR133W	**CKI1**	0,88	1,16	0,05	0,03	-0,73	1,20	0,14	choline kinase	phosphatidylcholine biosynthesis
YLR135W	**SLX4**	0,67	1,49	0,27	0,32	-0,33	0,93	0,00	molecular_function unknown	DNA replication
YLR136C	**TIS11**	1,16	1,74	0,49	0,11	0,13	1,50	0,15	member of the inducible ccch zinc-finger	biological_process unknown
YLR137W	**YLR137W**	0,89	1,61	0,53	0,43	0,09	1,41	0,41	hypothetical protein	biological_process unknown
YLR138W	**NHA1**	0,83	1,56	0,25	0,12	-0,49	1,17	-0,07	homology to NA+/H+ antiporters of S,pomb	monovalent inorganic cation homeostasis
YLR139C	**SLS1**	0,79	1,71	0,37	0,29	-0,17	1,79	0,50	suppresses lethality of SSM4 deletion	aerobic respiration; protein biosynthesis
YLR140W	**YLR140W**	0,25	1,70	1,04	0,58	0,43	1,78	0,04	questionable ORF	biological_process unknown
YLR141W	**RRN5**	0,16	1,46	0,55	0,41	-0,01	1,74	0,58	RNA polymerase I-specific transcription	transcription from Pol I promoter
YLR142W	**PUT1**	0,87	1,67	0,54	-0,14	-0,12	1,32	0,10	proline oxidase	proline catabolism; glutamate biosynthesis
YLR143W	**YLR143W**	0,90	1,72	0,50	0,46	-0,13	1,36	0,41	hypothetical protein	biological_process unknown
YLR144C	**ACF2**	0,78	1,36	0,54	0,11	-0,23	1,14	-0,08	glucan 1,3-beta-glucosidase activity	actin cytoskeleton organization and biogenesis
YLR145W	**YLR145W**	0,38	1,58	0,66	0,70	0,34	1,14	0,25	molecular_function unknown	tRNA processing
YLR146C	**SPE4**	0,72	1,48	0,53	0,31	0,61	1,27	0,14	homology to H,sapiens spermidine synthas	spermine biosynthesis; pantothenate biosynthesis
YLR155C	**ASP3-1**	-3,17	-5,60	-4,57	-4,74	-0,52	-5,46	-0,91	L-asparaginase II	cellular response to nitrogen starvation; asparagine catabolism
YLR156W	**YLR156W**	-2,36	-1,61	-2,94	-3,03	-1,55	-3,42	-0,47	homology to hypothetical proteins YLR159	biological_process unknown
YLR157C	**ASP3-2**	-2,82	-4,99	-4,40	-4,39	-0,17	-5,16	-0,79	L-asparaginase II	cellular response to nitrogen starvation; asparagine catabolism
YLR158C	**ASP3-3**	-3,48	-2,50	-4,98	-5,21	-2,06	-4,97	-1,11	L-asparaginase II	cellular response to nitrogen starvation; asparagine catabolism
YLR159W	**YLR159W**	-1,46	-3,10	-2,98	-3,20	-3,72	-3,48	-0,25	homology to YLR161w and YLR156w	biological_process unknown
YLR160C	**ASP3-4**	-3,45	-5,38	-5,06	-5,27	-4,87	-5,53	-1,03	L-asparaginase II	cellular response to nitrogen starvation; asparagine catabolism
YLR161W	**YLR161W**	-1,83	-3,14	-2,95	-2,92	-3,79	-3,46	-0,31	homolog to YLR156w and YLR161w	biological_process unknown
YLR162W	**YLR162W**	0,25	0,42	0,77	0,74	0,33	0,43	1,23	hypothetical protein	biological_process unknown
YLR164W	**YLR164W**	0,78	0,98	-0,05	0,30	-0,37	1,18	0,30	putative succinate dehydrogenase	biological_process unknown
YLR191W	**PEX13**	0,82	0,89	-0,29	-0,31	-0,67	1,04	0,54	peroxisomal protein involved in protein	protein-peroxisome targeting; peroxisome organization and biogenesis

YLR279W		0,23	0,73	-0,02	-0,57	-0,33	1,31	0,33	questionable ORF	biological_process unknown
YLR280C		0,34	0,74	0,47	-0,71	-0,16	1,19	0,41	questionable ORF	biological_process unknown
YLR281C	**YLR281C**	0,32	1,10	0,45	-0,70	0,03	1,16	0,82	hypothetical protein	biological_process unknown
YLR283W	**YLR283W**	0,12	-0,01	0,51	0,09	0,07	1,30	-0,06	hypothetical protein	biological_process unknown
YLR285W	**NNT1**	0,54	1,49	0,61	0,08	0,45	1,26	0,06	hypothetical protein	chromatin silencing at ribosomal DNA; nicotinamide metabolism
YLR293C	**GSP1**	0,03	0,48	0,64	0,42	1,20	0,32	-0,17	GTP-binding protein of the ras superfami	nuclear organization and biogenesis; rRNA processing; nucleocytoplasmic transport
YLR325C	**RPL38**	-0,09	-0,06	0,17	1,41	0,34	-0,05	-0,08	putative ribosomal protein L38	protein biosynthesis
YLR331C		1,09	1,37	-0,17	-0,29	-0,10	0,82	0,50	questionable ORF	biological_process unknown
YLR371W	**ROM2**	-0,26	-0,21	-0,02	1,39	-0,15	-0,12	-0,38	GDP-GTP exchange protein for Rho1p	actin filament organization; establishment of cell polarity (sensu Saccharomyces); small GTPase mediated signal transduction; bud growth; cell wall organization and biogenesis
YLR374C		-0,14	-0,13	1,50	0,43	0,02	-0,26	-0,12	questionable ORF	biological_process unknown
YLR375W	**STP3**	0,05	0,03	1,89	0,28	0,24	-0,06	-0,02	similarity to hypothetical protein YDL04	biological_process unknown
YLR385C	**AWS1**	0,88	0,44	0,30	-0,29	0,11	0,54	0,28	hypothetical protein	biological_process unknown
YLR415C	**YLR415C**	-0,34	0,12	1,67	-0,56	-1,23	0,17	-0,09	questionable ORF	biological_process unknown
YLR416C		-0,38	0,06	1,59	-0,76	-1,42	-0,08	-0,04	hypothetical protein	biological_process unknown
YLR419W	**YLR419W**	-0,11	0,10	0,05	1,70	-0,17	0,02	-0,35	similarity to helicases	biological_process unknown
YLR422W	**YLR422W**	1,80	-0,06	0,02	0,04	-0,29	0,59	0,54	hypothetical protein	biological_process unknown
YLR434C		0,91	1,42	0,08	-0,14	-0,50	0,89	-0,09	questionable ORF	biological_process unknown
YLR438W	**CAR2**	0,19	0,07	-1,87	3,90	-0,08	0,00	-0,12	ornithine aminotransferase	arginine catabolism
YLR439W	**MRPL4**	0,40	1,22	0,36	0,03	-0,06	1,10	-0,30	structural constituent of ribosome	protein biosynthesis
YLR464W	**YLR464W**	-0,99	-0,31	-0,14	-0,64	-0,08	-0,61	0,25	homology to other subtelomeric encoded p	biological_process unknown
YLR465C		-0,91	0,00	-0,20	-2,14	-0,16	-1,24	0,43	questionable ORF	biological_process unknown
YML039W	**YML039W**	-1,28	-0,79	0,53	-1,10	-1,17	-1,47	-0,89		Ty element transposition
YML040W	**YML040W**	-1,44	-0,74	-0,11	-3,09	-2,35	-1,97	-1,08	TY1A protein	Ty element transposition
YML045W	**YML045W**	-1,25	-0,72	-0,20	-2,99	-3,50	-2,01	-1,04	TY1A protein	Ty element transposition
YML057C-A		0,81	0,95	-0,02	0,03	-0,68	0,55	0,80		
YML065W	**ORC1**	-0,43	-0,58	0,13	0,14	-4,95	-0,24	-0,19	origin recognition complex, large subun	chromatin silencing at HML and HMR (sensu Saccharomyces); DNA replication initiation; pre-replicative complex formation and maintenance

YML066C	**SMA2**	-0,52	0,00	0,50	0,14	-1,35	-0,33	-0,12	hypothetical protein	spore wall assembly (sensu Saccharomyces)
YML067C	**ERV41**	-0,45	-0,45	0,23	0,33	-4,77	-0,76	-0,04		ER to Golgi transport
YML068W	**ITT1**	-0,16	0,00	0,14	-0,01	-2,15	-0,24	-0,08	hypothetical protein	regulation of translational termination
YML101C	**CUE4**	0,92	1,29	0,51	-1,08	-0,17	0,22	1,49	hypothetical protein	biological_process unknown
YML101C-A		0,64	0,00	0,48	-0,63	0,40	0,17	0,87		
YMR045C	**YMR045C**	-1,03	-0,41	0,17	-0,87	-1,03	-0,95	-0,74		Ty element transposition
YMR046C	**YMR046C**	-1,34	-0,68	0,01	-3,38	-2,25	-2,17	-0,87	TY1A protein	Ty element transposition
YMR051C	**YMR051C**	-1,20	-0,70	-0,20	-3,65	-3,23	-2,14	-0,73	TY1A protein	Ty element transposition
YMR052C-A		1,80	0,00	0,28	-1,74	-0,01	-0,32	0,37		biological_process unknown
YMR063W	**RIM9**	0,88	0,88	0,15	-0,38	0,05	0,90	0,06	hypothetical protein	sporulation (sensu Saccharomyces)
YMR065W	**KAR5**	0,89	1,08	0,16	-0,27	0,28	0,76	0,07	weak similarity to S,pombe hypothetical	karyogamy during conjugation with cellular fusion
YMR107W	**SPG4**	-0,16	-0,06	1,36	-0,22	0,23	-0,01	0,13	hypothetical protein	biological_process unknown
YMR144W	**YMR144W**	0,96	1,17	-0,17	-0,38	-0,32	0,76	0,43	hypothetical protein	biological_process unknown
YMR169C	**ALD3**	-0,18	-1,23	0,16	0,13	0,67	-0,40	0,27	aldehyde dehydrogenase activity	response to stress; polyamine catabolism; beta-alanine biosynthesis
YMR173W	**DDR48**	1,73	2,02	0,01	0,57	-0,28	1,31	1,16	stress protein	DNA repair
YMR173W-A	**YMR173W-A**	2,26	1,76	-0,30	-0,10	-0,15	1,10	1,50		biological_process unknown
YMR185W	**YMR185W**	-0,22	-0,17	-0,23	1,25	-0,04	-0,24	-0,22	hypothetical protein	biological_process unknown
YMR255W	**GFD1**	0,96	1,02	-0,34	-0,26	-0,25	0,90	0,48	hypothetical protein	mRNA-nucleus export
YMR320W		0,60	-0,91	-0,53	-2,33	-0,71	-0,23	-0,26	hypothetical protein	biological_process unknown
YMR322C	**SNO4**	-0,65	-1,31	-0,01	-0,02	-0,09	-0,69	-0,10	molecular_function unknown	pyridoxine metabolism
YMR326C		-0,73	-1,85	0,10	-0,14	0,37	-1,55	-0,36	similarity to other subtelomeric encoded	biological_process unknown
YNL184C		-0,55	-0,53	0,00	1,08	0,51	-0,25	-0,18	questionable ORF	biological_process unknown
YNL236W	**SIN4**	-0,10	-0,11	-0,01	1,28	0,32	-0,20	-0,09	global regulator protein	transcription from Pol II promoter
YNL285W		-0,63	0,36	-0,57	-1,27	-0,02	-0,19	-0,06	hypothetical protein	biological_process unknown
YNL328C	**MDJ2**	-0,82	-0,84	-0,07	-0,13	0,20	-1,18	-0,47	weak similarity to dnaJ homolog proteins	protein folding
YNL332W	**THI12**	-0,76	-1,44	-0,29	-0,19	0,38	-1,76	0,10	homology to Thi5p and YJR156c	thiamin biosynthesis
YNL333W	**SNZ2**	-0,52	-0,66	-1,07	-1,06	0,24	-0,76	-0,41	homology to YFL059w	pyridoxine metabolism; thiamin biosynthesis

YNL335W	**YNL335W**	-0,34	-0,68	-0,24	-1,26	0,60	-0,70	-0,48	homology to YFL061w	biological_process unknown
YNL336W	**COS1**	0,60	-0,49	0,40	3,25	-1,13	-0,33	-0,70	homology to other subtelomeric encoded p	biological_process unknown
YNR014W	**YNR014W**	-0,45	-0,03	1,51	-0,15	-0,22	0,00	-0,18	hypothetical protein	biological_process unknown
YNR040W	**YNR040W**	0,82	-0,07	-0,16	-2,44	-0,27	-0,04	-0,27	hypothetical protein	biological_process unknown
YNR060W	**FRE4**	-0,16	-0,58	-0,14	-0,15	-0,33	-0,27	-1,06	homology to Fre2p, YOR384w and YLL051c	iron-siderochrome transport
YNR061C	**YNR061C**	-0,61	0,00	0,08	-0,20	-0,37	-0,51	-1,89	similarity to hypothetical protein YDL21	biological_process unknown
YNR062C	**YNR062C**	-0,66	-0,74	-0,05	0,16	-0,19	-0,65	-0,98	similarity to H,influenzae L-lactate per	biological_process unknown
YNR065C	**YSN1**	-2,18	-0,35	-1,14	-2,89	-1,41	-1,91	-1,85	homology to carboxypeptidase Y-sorting P	biological_process unknown
YNR066C	**YNR066C**	-0,10	-0,23	-0,59	-0,71	-0,46	-0,18	-1,26	homology to carboxypeptidase Y-sorting P	biological_process unknown
YNR071C	**YNR071C**	-0,37	-1,22	-0,38	-0,19	0,12	-0,54	0,19	molecular_function unknown	biological_process unknown
YNR072W	**YNR072W**	-0,30	-1,45	-0,68	-0,66	-0,60	-1,22	-0,33	sugar transport protein	hexose transport
YNR073C	**YNR073C**	0,14	-0,63	-0,47	-1,24	0,38	-0,83	0,19	homology to E,coli D-mannonate oxidoredu	biological_process unknown
YNR074C	**YNR074C**	-0,56	-1,56	-5,90	-0,07	-5,97	-1,73	-3,73	weak similarity to B,subtilis nitrite re	response to singlet oxygen
YNR075W	**COS10**	-0,39	-0,46	-4,45	-0,06	-1,85	-1,29	-3,10	homology to YFL062w, YBR302c, YHL048w,	endocytosis
YOL046C	**YOL046C**	0,85	1,28	0,05	-0,22	-0,21	1,19	0,14	hypothetical protein	biological_process unknown
YOL047C	**YOL047C**	0,46	1,24	0,30	-0,12	-0,51	0,95	-0,02	molecular_function unknown	biological_process unknown
YOL050C	**YOL050C**	0,15	1,26	0,45	0,07	-0,23	1,49	-0,14	hypothetical protein	biological_process unknown
YOL052C	**SPE2**	0,35	1,23	0,43	0,14	0,00	1,10	0,00	adenosylmethionine decarboxylase activity	pantothenate biosynthesis
YOL054W	**PSH1**	0,62	1,24	-0,06	-0,06	-0,62	1,22	0,11	putative Zn-finger (C3HC4) protein	RNA elongation from Pol II promoter
YOL061W	**PRS5**	0,53	0,84	0,14	1,12	-0,25	0,37	0,22	putative ribose-phosphate pyrophosphokin	'de novo' pyrimidine base biosynthesis; 'de novo' IMP biosynthesis; purine salvage; tryptophan biosynthesis; histidine biosynthesis
YOL069W	**NUF2**	0,92	0,89	0,03	-0,33	0,30	1,05	0,11	spindle pole body protein	microtubule nucleation; chromosome segregation
YOL131W	**YOL131W**	1,01	0,19	-0,24	-0,35	-0,46	0,80	0,21	hypothetical protein	biological_process unknown
YOL147C	**PEX11**	0,43	0,11	-0,85	-0,48	-0,84	0,33	-0,85	peroxisomal membrane protein	peroxisome organization and biogenesis
YOL148C	**SPT20**	-0,14	-0,48	-0,19	-0,14	0,30	-0,38	-0,81	putative transcription factor	histone acetylation; chromatin modification
YOL152W	**FRE7**	-0,45	-0,48	-0,02	0,72	0,25	-0,71	-0,81	weak similarity to Fre1p and Fre2p	biological_process unknown
YOL153C	**YOL153C**	-0,53	-0,63	-0,44	0,01	-0,14	-0,47	-0,93		biological_process unknown
YOL156W	**HXT11**	-1,34	-1,92	-1,61	-1,62	0,19	-0,65	-0,20	low affinity glucose transport protein	hexose transport

YOL157C	**YOL157C**	-0,92	-0,98	-1,96	-1,28	0,16	-1,06	-0,55	putative alpha-glucosidase	biological_process unknown
YOL158C	**ENB1**	-0,18	0,20	-2,87	-0,30	0,14	-0,42	0,40	homology to YEL065p,YKR106p,YCL070p,YHL	ferric-enterobactin transport
YOL159C	**YOL159C**	-0,52	-0,64	-3,90	-0,72	0,56	-1,12	-0,33	weak similarity to O,aries melatonin rec	biological_process unknown
YOL160W		-0,35	-0,13	-4,83	0,28	1,61	-0,68	-3,40	hypothetical protein	biological_process unknown
YOL162W	**YOL162W**	-2,94	-4,11	-3,06	-4,11	-3,24	-4,01	-2,87	homology to hypothetical protein YIL166c	transport
YOL163W	**YOL163W**	-2,67	-4,07	-4,85	-6,47	-5,19	-5,16	-3,26	similarity to Pseudomonas putida phthala	transport
YOL164W	**YOL164W**	-2,45	-3,56	-2,65	-2,74	-2,40	-3,59	-2,34	homology to P,sp,alkyl sulfatase	biological_process unknown
YOL165C	**AAD15**	-1,04	-1,28	-2,13	-2,52	-0,97	-1,45	-1,62	putative aryl-alcohol dehydrogenase (NAD	aldehyde metabolism
YOL166C	**YOL166C**	-1,26	-0,68	-0,93	-1,05	-1,09	-1,73	-1,54	hypothetical protein	biological_process unknown
YOR008C	**SLG1**	0,98	0,68	-0,05	-0,07	0,01	0,76	0,64	similarity to L,mexicana lmsap2 gene (se	response to osmotic stress; actin cytoskeleton organization and biogenesis; response to heat; establishment of cell polarity (sensu Saccharomyces); Rho protein signal transduction; cell wall organization and biogenesis
YOR038C	**HIR2**	-0,32	-0,49	-0,05	1,09	0,04	-0,13	-0,01	histone transcription regulator	regulation of transcription from Pol II promoter
YOR041C		0,83	0,60	-0,51	-0,69	-0,97	0,08	0,45	hypothetical protein	biological_process unknown
YOR072W		0,05	0,72	1,46	-0,17	-0,45	0,29	0,12	hypothetical protein	biological_process unknown
YOR079C	**ATX2**	0,75	1,49	-0,05	-0,16	-0,23	1,14	0,24	manganese ion transporter activity	manganese ion homeostasis
YOR085W	**OST3**	0,61	1,34	0,02	0,02	-0,22	1,09	0,17	dolichyl-diphosphooligosaccharide-protein glycotransferase activity	protein amino acid glycosylation; N-linked glycosylation; protein complex assembly
YOR146W		1,08	0,61	0,06	-0,17	0,25	0,29	0,11	hypothetical protein	biological_process unknown
YOR155C	**ISN1**	-0,25	-0,24	0,07	0,09	-0,05	0,03	-0,29	hypothetical protein	biological_process unknown
YOR206W	**NOC2**	-0,19	-0,44	0,12	0,14	-1,22	-0,21	-0,07	hypothetical protein	ribosome-nucleus export; ribosome assembly
YOR225W		0,98	-0,45	-4,68	-0,01	-0,50	0,23	-1,97	hypothetical protein	biological_process unknown
YOR343C		0,26	0,05	-0,22	-1,62	0,24	-0,17	0,53	hypothetical protein	biological_process unknown
YOR382W	**FIT2**	-0,33	-0,10	0,07	-4,12	0,49	-0,03	-0,59	hypothetical protein	siderochrome transport
YOR383C	**FIT3**	-0,33	-0,02	0,20	-3,07	0,60	-0,13	-0,70	homology to hypothetical protein YDR534c	siderochrome transport
YOR384W	**FRE5**	-0,15	0,03	-0,31	-4,82	-0,07	0,17	-0,37	homology to ferric reductase FRE2 precur	biological_process unknown
YOR385W	**YOR385W**	-0,30	-0,41	-0,03	-3,68	0,54	-0,17	-0,45	homology to hypothetical protein YMR316w	biological_process unknown
YOR386W	**PHR1**	-0,50	-0,37	-0,17	-3,77	0,19	-0,21	-0,41	deoxyribodipyrimidine photo-lyase	photoreactive repair
YOR391C	**HSP33**	-0,49	-1,37	0,43	0,05	0,91	-0,74	0,02	molecular_function unknown	biological_process unknown

YPL034W	**YPL034W**	0,84	0,80	0,08	0,15	0,23	0,80	0,62	questionable ORF	biological_process unknown
YPL035C		1,18	1,12	0,17	0,12	0,09	0,74	0,50	questionable ORF	biological_process unknown
YPL058C	**PDR12**	0,03	0,45	0,37	0,35	-0,38	0,53	-1,23	homology to Snq2p	organic acid transport; propionate metabolism; transport
YPL062W		-0,25	-0,14	1,57	-0,35	-1,12	-0,01	0,16	hypothetical protein	biological_process unknown
YPL167C	**REV3**	-0,46	-0,23	0,34	1,67	-0,15	-0,07	-0,35	DNA-directed DNA polymerase	DNA repair; mutagenesis
YPL170W	**DAP1**	1,31	0,13	-0,12	0,18	-0,02	-0,09	-0,57	hypothetical protein	sterol metabolism
YPL247C	**YPL247C**	-0,15	-0,11	8,25	1,56	1,62	-0,12	-0,04	hypothetical protein	biological_process unknown
YPL254W	**HFI1**	-0,23	0,02	1,04	-1,30	-0,38	-0,03	-0,12	interacts functionally with histone H2A	transcription from Pol II promoter; histone acetylation; chromatin modification
YPL257W	**YPL257W**	-0,22	0,25	-0,83	-4,09	0,25	-4,68	-3,39	hypothetical protein	biological_process unknown; Ty element transposition
YPL267W	**YPL267W**	-0,56	-0,52	-0,05	1,22	0,50	-0,49	-0,58	hypothetical protein	biological_process unknown
YPL280W	**HSP32**	-0,64	-1,25	0,35	0,16	0,67	-0,75	-0,31	molecular_function unknown	biological_process unknown
YPR030W	**CSR2**	0,00	0,22	0,02	1,18	-0,17	0,36	-0,16	similarity to hypothetical protein YBL10	cell wall organization and biogenesis; regulation of transcription from Pol II promoter
YPR032W	**SRO7**	0,25	0,26	-0,10	1,13	-0,12	0,40	-0,03	homology to hypothetical protein YBL106c	Golgi to plasma membrane transport; exocytosis
YPR035W	**GLN1**	0,61	1,73	0,53	0,21	0,28	1,05	0,17	glutamate-ammonia ligase activity	nitrogen metabolism; glutamine biosynthesis
YPR036W	**VMA13**	0,58	1,60	0,38	0,36	0,33	1,32	0,21	vacuolar H+-transporting ATPase, 54K ch	vacuolar acidification
YPR037C	**ERV2**	0,01	1,44	0,68	0,37	0,47	1,10	-0,10	thiol oxidase activity	protein thiol-disulfide exchange
YPR038W		0,16	1,33	0,82	0,25	0,15	1,18	-0,14	questionable ORF	biological_process unknown
YPR039W		-0,01	1,44	0,77	0,25	0,04	1,14	-0,23	molecular_function unknown	biological_process unknown
YPR040W	**TIP41**	0,60	1,53	0,08	-0,03	0,32	1,14	0,32	molecular_function unknown	signal transduction
YPR041W	**TIF5**	-0,04	1,24	0,71	0,35	0,46	0,86	-0,06	translation initiation factor activity; GTPase activator activity	mature ribosome assembly; regulation of translational initiation
YPR060C	**ARO7**	0,62	0,86	-0,11	-2,63	-0,01	0,51	0,28	chorismate mutase	aromatic amino acid family biosynthesis
YPR128C	**ANT1**	0,13	0,24	-0,31	1,26	0,13	-0,16	0,06	weak similarity to carrier protein FLX1	peroxisome organization and biogenesis; fatty acid beta-oxidation; ATP transport
YPR130C		0,30	0,19	-0,55	1,20	-0,23	0,13	-0,14	questionable ORF	biological_process unknown
YPR140W	**YPR140W**	-0,05	0,14	0,02	0,19	1,18	0,14	-0,27	hypothetical protein	phospholipid biosynthesis
YPR142C		0,88	0,54	0,03	-0,35	-0,35	0,80	0,03	questionable ORF	biological_process unknown
YPR189W	**SKI3**	-0,16	0,48	0,20	-0,16	-0,12	3,86	0,00	antiviral protein	mRNA catabolism
YPR196W	**YPR196W**	0,05	0,06	1,07	0,32	0,77	-0,05	-0,34	putative transcription factor	biological_process unknown

YPR202W	YPR202W	-1,12	-0,51	-0,30	0,18	0,10	-0,48	-0,03	homology to other subtelomeric encoded protein	biological_process unknown
YPR203W	YPR203W	-1,09	-0,51	0,23	0,27	-0,47	-0,30	-0,15	homology to other subtelomeric encoded protein	biological_process unknown

Table S1.2. Validation of microarray data by qPCR to determine GCN*

Gene	CECT10131	T73	Temohaya-26	PE35M
HAP3	5.03 ± 0.22	3.39 ± 0.24	1.64 ± 0.05	0,81 ± 0,02
PUT1	6.59 ± 0.40	4.96 ± 0.09	1.94 ± 0.03	1,77 ± 0,008
SNF4	4.97 ± 0.40	3.34 ± 0.03	0.88 ± 0.13	0,54 ± 0,09
VMA5	2.87 ± 0.15	5.66 ± 0.30	1.08 ± 0.12	0,5 ± 0,04

*Gene copy number. Average ± standard deviation is shown

Suplementary data of Chapter 2

Table S2.1. Amplified (A) and depleted (B) genes shared by the different *S. uvarum* strains.

(A)

ID	Gene	Strains				Function	Process
		CECT1969	CECT12600	BMV58	NCAIM 789		
YBR012C		●	●	●	●	molecular_function unknown	biological_process unknown
YCL067C	HMLALPHA2	●	●	●	●	transcription corepressor activity	donor preference; regulation of transcription from Pol II promoter;
YCR039C	MATALPHA2	●	●	●	●	transcription corepressor activity	donor preference; regulation of transcription from Pol II promoter;
YDR134C	YDR134C	●	●	●	●	molecular_function unknown	biological_process unknown
YGR192C	TDH3	●	●	●	●	glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)	glycolysis; gluconeogenesis
YJR009C	TDH2	●	●	●	●	glyceraldehyde-3-phosphate dehydrogenase (NAD+) (phosphorylating) activity	Apoptotic process; gluconeogenesis; glycolysis; reactive oxygen species metabolic process
YJR123W	RPS5	●	●	●	●	structural constituent of ribosome	Cytoplasmic translation; regulation of translational fidelity; Rrna
YKL060C	FBA1	●	●	●	●	fructose-bisphosphate aldolase	Gluconeogenesis; glycolysis
YKL144C	RPC25	●	●	●	●	contributes_to: DNA-directed RNA polymerase activity	transcription initiation from RNA polymerase III promoter; tRNA transcription from RNA polymerase III promoter
YLR110C	CCW12	●	●	●	●	molecular_function unknown	agglutination involved in conjugation with cellular fusion; conjugation
YOR167C	RPS28A	●	●	●	●	structural constituent of ribosome	protein biosynthesis
YBL026W	LSM2	●	●		●	pre-mRNA splicing factor activity	nuclear mRNA splicing, via spliceosome; rRNA processing
YBR103W	SIF2	●		●	●	NAD-dependent histone deacetylase activity; NAD-independent histone deacetylase	chromatin silencing at telomere; histone deacetylation; negative regulation of transcription from Pol II promoter; negative regulation of meiosis
YDL192W	ARF1	●	●		●	ARF small monomeric GTPase	intra-Golgi transport; ER to Golgi transport
YDR037W	KRS1	●	●		●	lysine-tRNA ligase activity	lysyl-tRNA aminoacylation

YFR031C-A	**RPL2A**	●	●		●	structural constituent of ribosome	protein biosynthesis
YGL220W	**YGL220W**	●	●	●		molecular_function unknown	negative regulation of transcription from RNA polymerase II promoter
YGR027C	**RPS25A**	●	●		●	structural constituent of ribosome	protein biosynthesis
YHL015W	**RPS20**	●	●	●		structural constituent of ribosome	Cytoplasmic translation; maturation of SSU-rRNA from tricistronic
YJL052W	**TDH1**	●	●		●	glyceraldehyde-3-phosphate	Gluconeogenesis; glycolysis
YJR044C	**VPS55**	●	●	●		molecular_function unknown	late endosome to vacuole transport via multivesicular body sorting
YKL152C	**GPM1**	●	●		●	phosphoglycerate mutase activity	Gluconeogenesis; glycolysis
YKL080W	**VMA5**	●		●	●	proton-transporting ATPase	vacuolar acidification
YKL153W		●	●	●		molecular_function unknown	biological_process unknown
YLR029C	**RPL15A**	●	●		●	RNA binding; structural	cytoplasmic translation
YLR044C	**PDC1**	●	●		●	pyruvate decarboxylase activity	aromatic amino acid family catabolic process to alcohol via Ehrlich
YLR167W	**RPS31**	●		●	●	Protein tag; structural constituent	cytoplasmic translation; maturation of SSU-rRNA from tricistronic
YLR264W	**RPS28B**	●	●	●		structural constituent of ribosome	cytoplasmic translation; rRNA export from nucleus
YLR333C	**RPS25B**	●		●	●	structural constituent of ribosome	cytoplasmic translation
YML063W	**RPS1B**	●	●		●	structural constituent of ribosome	cytoplasmic translation;maturation of SSU-rRNA from tricistronic
YOL120C	**RPL18A**	●	●	●		structural constituent of ribosome	cytoplasmic translation
YOR053W		●		●	●	molecular_function unknown	biological_process unknown
YPR154W	**PIN3**	●	●	●		molecular_function unknown	actin cytoskeleton organization and biogenesis
YPR165W	**RHO1**	●	●	●		signal transducer activity; Rho	actin filament organization; establishment of cell polarity (sensu
YAL005C	**SSA1**	●		●		chaperone activity; ATPase	SRP-dependent cotranslational membrane targeting, translocation;
YBL030C	**PET9**	●		●		ATP:ADP antiporter activity	ATP/ADP exchange
YBL072C	**RPS8A**			●	●	structural constituent of ribosome	protein biosynthesis
YBL075C	**SSA3**	●			●	heat shock protein activity	SRP-dependent cotranslational membrane targeting, translocation;
YBL078C	**ATG8**		●		●	microtubule binding	autophagy; protein-vacuolar targeting
YBR054W	**YRO2**	●	●			molecular_function unknown	biological_process unknown
YBR089C-A	**NHP6B**	●	●			chromatin binding	regulation of transcription from Pol II promoter; establishment and/or

YBR099C				●	●	molecular_function unknown	biological_process unknown
YBR101C	**FES1**			●	●	adenyl-nucleotide exchange	protein biosynthesis
YBR113W		●	●			molecular_function unknown	biological_process unknown
YBR135W	**CKS1**	●	●			protein kinase activator activity	regulation of cell cycle
YBR149W	**ARA1**			●	●	D-arabinose 1-dehydrogenase	carbohydrate metabolism
YBR181C	**RPS6B**		●		●	structural constituent of ribosome	protein biosynthesis
YCR009C	**RVS161**	●	●			cytoskeletal protein binding	polar budding; response to osmotic stress; endocytosis
YDL014W	**NOP1**			●	●	methyltransferase activity	35S primary transcript processing; ribosomal large subunit assembly
YDL081C	**RPP1A**	●	●			structural constituent of ribosome	translational elongation; protein biosynthesis
YDL130W	**RPP1B**	●		●		structural constituent of ribosome	translational elongation; protein biosynthesis
YDR008C		●	●			molecular_function unknown	biological_process unknown
YDR054C	**CDC34**	●	●			ubiquitin-protein ligase activity;	G2/M transition of mitotic cell cycle; G1/S transition of mitotic cell
YDR099W	**BMH2**	●	●			protein binding; DNA binding	glycogen metabolism; signal transduction during filamentous growth;
YDR155C	**CPR1**	●	●			peptidyl-prolyl cis-trans	protein metabolism
YDR215C		●	●			molecular_function unknown	biological_process unknown
YDR225W	**HTA1**		●		●	DNA binding	chromatin assembly/disassembly
YDR382W	**RPP2B**	●	●			structural constituent of ribosome	translational elongation; protein biosynthesis
YDR433W	**KRE22**		●		●	molecular_function unknown	response to DNA damage stimulus
YEL036C	**ANP1**	●	●			mannosyltransferase activity	N-linked glycosylation
YER177W	**BMH1**	●	●			protein binding; DNA binding	glycogen metabolism; signal transduction during filamentous growth;
YGL123W	**RPS2**			●	●	structural constituent of ribosome	regulation of translational fidelity; protein biosynthesis
YGL135W	**RPL1B**			●	●	structural constituent of ribosome	protein biosynthesis
YGR216C	**GPI1**		●		●	phosphatidylinositol N-	GPI anchor biosynthesis
YGR264C	**MES1**	●	●			methionine-tRNA ligase activity	methionyl-tRNA aminoacylation
YHL020C	**OPI1**	●	●			transcription corepressor activity	endoplasmic reticulum unfolded protein response; negative
YHL033C	**RPL8A**			●	●	structural constituent of ribosome	cytoplasmic translation

YIL022W	**TIM44**	●	●			Chaperone binding; protein	protein import into mitochondrial matrix
YJL159W	**HSP150**			●	●	ATPase activity; structural	fungal-type cell wall organization
YJL190C	**RPS22A**		●		●	structural constituent of ribosome	biological_process unknown
YJR041C	**URB2**	●			●	molecular_function unknown	Ribosome biogenesis;rRNA metabolic process
YJR065C	**ARP3**	●			●	ATP binding	actin filament organization; Arp2/3 complex-mediated actin
YJR086W	**STE18**	●	●			signal transducer activity	heterotrimeric G-protein complex cycle; pheromone-dependent
YKL016C	**ATP7**	●	●			ATPase activity; proton-	ATP synthesis coupled proton transport; protein complex assembly
YKL056C	**TMA19**			●	●	molecular_function unknown	cellular response to oxidative stress; cytoplasmic translation
YLL044W				●	●	molecular_function unknown	biological_process unknown
YLL045C	**RPL8B**			●	●	structural constituent of ribosome	cytoplasmic translation
YLR075W	**RPL10**			●	●	structural constituent of ribosome	cytoplasmic translation; ribosomal large subunit assembly;
YLR293C	**GSP1**			●	●	GTPase activity	exonucleolytic trimming to generate mature 3'-end of 5.8S rRNA from
YLR325C	**RPL38**			●	●	structural constituent of ribosome	cytoplasmic translation
YLR339C				●	●	molecular_function unknown	biological_process unknown
YLR340W	**RPP0**			●	●	LSU rRNA binding; structural	cytoplasmic translation; ribosomal large subunit assembly;
YLR437C	**YLR437C**	●	●			molecular_function unknown	biological_process unknown
YLR441C	**RPS1A**	●	●			structural constituent of ribosome	Cytoplasmic translation; maturation of SSU-rRNA from tricistronic
YMR273C	**ZDS1**	●	●			protein phosphatase type 2A	Cell aging; cell shape checkpoint; chromatin silencing; establishment
YNL337W		●	●			molecular_function unknown	biological_process unknown
YNR028W	**CPR8**	●	●			peptidyl-prolyl cis-trans	biological_process unknown
YNR030W	**ECM39**		●		●	alpha-1,6-mannosyltransferase	dolichol-linked oligosaccharide biosynthetic process; protein
YNR034W	**SOL1**			●	●	molecular_function unknown	tRNA export from nucleus
YOL001W	**PHO80**			●	●	cyclin-dependent protein	Cellular metal ion homeostasis; negative regulation of calcium-
YOL039W	**RPP2A**			●	●	structural constituent of ribosome	cytoplasmic translation; translational elongation
YOL040C	**RPS15**			●	●	structural constituent of ribosome	cytoplasmic translation; rRNA export from nucleus
YOL086C	**ADH1**			●	●	alcohol dehydrogenase (NAD)	amino acid catabolic process to alcohol via Ehrlich pathway; ethanol

YOR083W	**WHI5**	●	●			RNA polymerase II transcription	negative regulation of transcription involved in G1/S phase of mitotic
YOR369C	**RPS12**	●	●			structural constituent of ribosome	protein biosynthesis
YOR372C	**NDD1**	●	●			transcriptional activator activity	G2/M-specific transcription in mitotic cell cycle
YPL234C	**TFP3**	●	●			chromatin binding	regulation of transcription from Pol II promoter; establishment and/or
YPR016C	**TIF6**		●		●	chromatin binding	regulation of transcription from Pol II promoter; establishment and/or
YPR044C				●	●	chromatin binding	regulation of transcription from Pol II promoter; establishment and/or
YAL003W	**EFB1**			●		translation elongation factor	translational elongation
YAL012W	**CYS3**			●		cystathionine gamma-lyase	cysteine metabolism; sulfur amino acid metabolism; transsulfuration
YBL002W	**HTB2**			●		DNA binding	chromatin assembly/disassembly
YBL021C	**HAP3**		●			transcriptional activator activity	regulation of carbohydrate metabolism; transcription
YBL025W	**RRN10**				●	RNA polymerase I transcription	transcription from Pol I promoter
YBL029W	**YBL029W**				●	molecular_function unknown	biological_process unknown
YBL031W	**SHE1**				●	microtubule binding	mitotic spindle elongation
YBL074C	**AAR2**				●	molecular_function unknown	assembly of spliceosomal tri-snRNP
YBL077W					●	molecular_function unknown	biological_process unknown
YBR118W	**TEF2**				●	translation elongation factor	translational elongation
YBR147W	**YBR147W**				●	cationic amino acid	basic amino acid transmembrane export from vacuole
YBR162C	**TOS1**			●		molecular_function unknown	biological_process unknown
YBR249C	**ARO4**			●		3-deoxy-7-phosphoheptulonate	aromatic amino acid family biosynthesis
YCL009C	**ILV6**			●		enzyme regulator activity;	branched chain family amino acid biosynthesis
YCR003W	**MRPL32**	●				structural constituent of ribosome	protein biosynthesis
YCR071C	**IMG2**			●		structural constituent of ribosome	protein biosynthesis
YCR096C	**HMRA2**		●			sequence-specific DNA binding	biological_process unknown
YDL062W					●	molecular_function unknown	biological_process unknown
YDL067C	**COX9**			●		cytochrome-c oxidase activity	aerobic respiration
YDL075W	**RPL31A**				●	structural constituent of ribosome	protein biosynthesis

YDL084W	**SUB2**			●		pre-mRNA splicing factor activity;	U2-type nuclear mRNA branch site recognition; mRNA-nucleus
YDL100C	**ARR4**			●		ATPase activity	response to heat; response to metal ion
YDL126C	**CDC48**			●		ATPase activity	cell cycle; ubiquitin-dependent protein catabolism; vesicle fusion;
YDL134C	**PPH21**			●		protein phosphatase type 2A	protein biosynthesis; protein amino acid dephosphorylation; bud
YDL165W	**CDC36**			●		3'-5' exoribonuclease activity	regulation of transcription from Pol II promoter; regulation of cell
YDR095C			●			molecular_function unknown	biological_process unknown
YDR224C	**HTB1**				●	DNA binding	chromatin assembly/disassembly
YDR385W	**EFT2**			●		translation elongation factor	translational elongation
YDR432W	**NPL3**		●			mRNA binding	mRNA-nucleus export
YEL027W	**CUP5**				●	hydrogen ion transporter activity	vacuole organization and biogenesis; endocytosis; protein-vacuolar
YER031C	**YPT31**		●			GTPase activity	exocytosis; vesicle-mediated transport
YFR008W	**FAR7**		●			molecular_function unknown	cell cycle arrest in response to pheromone
YFR055W	**YFR055W**				●	cystathionine beta-lyase activity	sulfur metabolism; copper ion homeostasis
YGL030W	**RPL30**	●				structural constituent of ribosome	negative regulation of translation; negative regulation of nuclear
YGL043W	**DST1**			●		positive transcription elongation	meiotic recombination; RNA elongation from Pol II promoter
YGL058W	**RAD6**		●			ubiquitin conjugating enzyme	DNA repair; ubiquitin-dependent protein catabolism; histone
YGL102C				●		molecular_function unknown	biological_process unknown
YGR086C	**PIL1**	●				lipid binding	Eisosome assembly; endocytosis; negative regulation of protein
YGR159C	**NSR1**			●		RNA binding; single-stranded	rRNA processing; ribosomal small subunit assembly and
YGR212W	**YGR212W**	●				N-acetyltransferase activity	response to drug
YHR008C	**SOD2**			●		superoxide dismutase activity	age-dependent response to oxidative stress involved in chronological
YHR041C	**SRB2**			●		Subunit of the RNA polymerase II	negative regulation of ribosomal protein gene transcription from RNA
YHR087W	**YHR087W**			●		molecular_function unknown	RNA metabolic process
YHR093W	**AHT1**			●		molecular_function unknown	biological_process unknown
YHR152W	**SPO12**		●			molecular_function unknown	Meiosis; mitotic cell cycle; regulation of exit from mitosis
YIL053W	**RHR2**	●				glycerol-1-phosphatase activity	glycerol biosynthetic process; response to osmotic stress

YIL105C	**LIT2**	●				phosphatidylinositol binding;	actin cytoskeleton organization; actin filament bundle assembly;
YJL065C	**DLS1**		●			molecular_function unknown	chromatin silencing at telomere
YJL138C	**TIF2**			●		ATP-dependent RNA helicase	regulation of translational initiation
YJL158C	**CIS3**			●		structural constituent of cell wall	fungal-type cell wall organization
YJL166W	**QCR8**			●		ubiquinol-cytochrome-c	Aerobic respiration; mitochondrial electron transport, ubiquinol to
YJR038C			●			molecular_function unknown	biological_process unknown
YJR045C	**SSC1**		●			ATPase activity; enzyme	positive regulation of endodeoxyribonuclease activity; protein import
YJR047C	**ANB1**				●	Ribosome binding; RNA binding;	positive regulation of translational elongation; positive regulation of
YJR048W	**CYC1**			●		electron carrier activity	mitochondrial electron transport, cytochrome c to oxygen;
YKL117W	**SBA1**			●		chaperone binding	positive regulation of telomere maintenance via telomerase; protein
YKL190W	**CNB1**		●			Calcium ion binding	adaptation of signaling pathway by response to pheromone involved
YKR039W	**GAP1**			●		Amino acid transmembrane	Amino acid transport; polyamine transport
YKR085C	**MRPL20**			●		structural constituent of ribosome	mitochondrial translation
YKR092C	**SRP40**			●		molecular_function unknown	nucleocytoplasmic transport
YLL027W	**ISA1**			●		iron ion binding	Iron-sulfur cluster binding
YLR008C	**PAM18**			●		ATPase activator activity;	protein import into mitochondrial matrix
YLR076C					●	molecular_function unknown	biological_process unknown
YLR104W	**YLR104W**	●				molecular_function unknown	ER-associated protein catabolic process
YLR134W	**PDC5**				●	pyruvate decarboxylase activity	aromatic amino acid family catabolic process to alcohol via Ehrlich
YLR150W	**STM1**				●	DNA binding; telomeric DNA	Negative regulation of apoptotic process; regulation of translational
YLR162W	**YLR162W**				●	molecular_function unknown	biological_process unknown
YLR198C		●				molecular_function unknown	biological_process unknown
YLR355C	**ILV5**				●	Double-stranded DNA binding;	branched-chain amino acid biosynthetic process; mitochondrial
YLR388W	**RPS29A**		●			structural constituent of ribosome	cytoplasmic translation
YLR465C	**BSC3**				●	molecular_function unknown	biological_process unknown
YMR121C	**RPL15B**	●				molecular_function unknown	biological_process unknown

YMR173W-A			●			molecular_function unknown	biological_process unknown
YMR242C	**RPL20A**	●				structural constituent of ribosome	cytoplasmic translation
YNL070W	**TOM7**			●		protein channel activity	protein import into mitochondrial matrix; protein import into
YNL075W	**IMP4**	●				rRNA primary transcript binding;	rRNA processing
YNL112W	**DBP2**			●		RNA helicase activity; RNA-	nuclear polyadenylation-dependent mRNA catabolic process;
YNL178W	**RPS3**				●	DNA-(apurinic or apyrimidinic	cytoplasmic translation; ribosomal small subunit export from nucleus;
YNL208W	**YNL208W**			●		molecular_function unknown	biological_process unknown
YNL222W	**SSU72**		●			CTD phosphatase activity;	dephosphorylation of RNA polymerase II C-terminal domain; mRNA
YNR043W	**MVD1**		●			diphosphomevalonate	sterol biosynthetic process
YNR057C	**BIO4**		●			dethiobiotin synthase activity	biotin biosynthetic process
YOL134C				●		molecular_function unknown	biological_process unknown
YOL152W	**FRE7**		●			ferric-chelate reductase activity	Copper ion import; iron ion transport
YOL158C	**ENB1**		●			ferric-enterobactin	ferric-enterobactin transport
YOR034C	**AKR2**		●			palmitoyltransferase activity	Endocytosis; protein palmitoylation
YOR054C	**VHS3**		●			phosphopantothenoylcysteine	cellular monovalent inorganic cation homeostasis; coenzyme A
YOR099W	**KTR1**			●		alpha-1,2-mannosyltransferase	cell wall mannoprotein biosynthetic process; N-glycan processing;
YOR133W	**EFT1**	●				translation elongation factor	translational elongation
YOR145C	**PNO1**			●		unfolded protein binding	endonucleolytic cleavage in ITS1 to separate SSU-rRNA from 5.8S
YOR185C	**GSP2**			●		structural constituent of ribosome	protein biosynthesis
YOR187W	**TUF1**			●		structural constituent of ribosome	protein biosynthesis
YOR219C	**STE13**				●	structural constituent of ribosome	protein biosynthesis
YOR226C	**ISU2**		●			structural constituent of ribosome	protein biosynthesis
YOR235W			●			structural constituent of ribosome	protein biosynthesis
YOR250C	**CLP1**			●		structural constituent of ribosome	protein biosynthesis
YOR300W				●		structural constituent of ribosome	protein biosynthesis
YOR309C				●		structural constituent of ribosome	protein biosynthesis

YPL023C	**MET12**				●	chromatin binding	regulation of transcription from Pol II promoter; establishment and/or
YPL220W	**RPL1A**			●		chromatin binding	regulation of transcription from Pol II promoter; establishment and/or
YPL239W	**YAR1**			●		chromatin binding	regulation of transcription from Pol II promoter; establishment and/or
YPR010C	**RPA135**			●		chromatin binding	regulation of transcription from Pol II promoter; establishment and/or
YPR035W	**GLN1**				●	chromatin binding	regulation of transcription from Pol II promoter; establishment and/or
YPR102C	**RPL11A**			●		chromatin binding	regulation of transcription from Pol II promoter; establishment and/or
YPR188C	**MLC2**			●		Myosin II binding	actomyosin contractile ring contraction
YPR201W	**ARR3**			●		arsenite transporter activity	arsenite transport

(B)

ID	**Gene**	**Strains**				**Function**	**Process**
		CECT1969	**CECT12600**	**BMV58**	**NCAIM 789**		
YAR023C	**YAR023C**	●	●	●	●	molecular_function unknown	biological_process unknown
YAR031W		●	●	●	●	molecular_function unknown	conjugation with cellular fusion
YAR053W		●	●	●	●	molecular_function unknown	biological_process unknown
YAR060C		●	●	●	●	not yet annotated	biological_process unknown
YBL005W-A	**YBL005W-A**	●	●	●	●	protein binding; RNA binding	Ty element transposition
YBL101W-A	**YBL101W-A**	●	●	●	●	protein binding; RNA binding	Ty element transposition
YBL111C	**YBL111C**	●	●	●	●	molecular_function unknown	biological_process unknown
YBL112C	**YBL112C**	●	●	●	●	molecular_function unknown	biological_process unknown
YCL069W	**YCL069W**	●	●	●	●	basic amino acid transmembrane	basic amino acid transport
YDR179C	**CSN9**	●	●	●	●	molecular_function unknown	protein deneddylation
YDR540C	**YDR540C**	●	●	●	●	molecular_function unknown	mitotic recombination
YEL075C	**YEL075C**	●	●	●	●	molecular_function unknown	biological_process unknown
YEL075W-A		●	●	●	●		
YEL076C	**YEL076C**	●	●	●	●	molecular_function unknown	biological_process unknown

YEL077C	**YEL077C**	●	●	●	●	helicase activity	biological_process unknown
YER189W	**YER189W**	●	●	●	●	molecular_function unknown	biological_process unknown
YFL062W	**COS4**	●	●	●	●	molecular_function unknown	biological_process unknown
YFL064C	**YFL064C**	●	●	●	●	molecular_function unknown	biological_process unknown
YFL066C	**YFL066C**	●	●	●	●	molecular_function unknown	biological_process unknown
YFL068W	**YFL068W**	●	●	●	●	molecular_function unknown	biological_process unknown
YGL051W	**MST27**	●	●	●	●	protein binding	vesicle organization and biogenesis
YGL053W	**PRM8**	●	●	●	●	molecular_function unknown	conjugation with cellular fusion
YGL257C	**MNT2**	●	●	●	●	alpha-1,3-mannosyltransferase	O-linked glycosylation
YGR154C	**YGR154C**	●	●	●	●	glutathione transferase activity	glutathione metabolic process
YHL048W	**COS8**	●	●	●	●	molecular_function unknown	response to unfolded protein
YHL049C	**YHL049C**	●	●	●	●	molecular_function unknown	biological_process unknown
YHL050C	**YHL050C**	●	●	●	●	helicase activity	biological_process unknown
YHR053C	**CUP1-1**	●	●	●	●	copper ion binding	response to copper ion
YHR054C	**YHR054C**	●	●	●	●	molecular_function unknown	biological_process unknown
YHR055C	**CUP1-2**	●	●	●	●	**copper ion binding**	**response to copper ion**
YHR212C		●	●	●	●	molecular_function unknown	biological_process unknown
YHR218W	**YHR218W**	●	●	●	●	molecular_function unknown	biological_process unknown
YHR219W	**YHR219W**	●	●	●	●	molecular_function unknown	biological_process unknown
YIL060W	**YIL060W**	●	●	●	●	molecular_function unknown	biological_process unknown
YIL102C	**YIL102C**	●	●	●	●	molecular_function unknown	biological_process unknown
YIR040C		●	●	●	●	molecular_function unknown	biological_process unknown
YIR043C	**YIR043C**	●	●	●	●	molecular_function unknown	biological_process unknown
YIR044C	**YIR044C**	●	●	●	●	molecular_function unknown	biological_process unknown
YJL182C		●	●	●	●	molecular_function unknown	biological_process unknown
YJR026W	**YJR026W**	●	●	●	●	protein binding; RNA binding	Ty element transposition
YJR028W	**YJR028W**	●	●	●	●	protein binding; RNA binding	Ty element transposition
YJR157W		●	●	●	●	molecular_function unknown	biological_process unknown
YJR161C	**COS5**	●	●	●	●	molecular_function unknown	biological_process unknown

YKL219W	**COS9**	●	●	●	●	molecular_function unknown	biological_process unknown
YKR105C	**YKR105C**	●	●	●	●	molecular_function unknown	arginine transport
YLR155C	**ASP3-1**	●	●	●	●	asparaginase activity	cellular response to nitrogen starvation; asparagine
YLR157C	**ASP3-2**	●	●	●	●	asparaginase activity	cellular response to nitrogen starvation; asparagine
YLR158C	**ASP3-3**	●	●	●	●	asparaginase activity	cellular response to nitrogen starvation; asparagine
YLR160C	**ASP3-4**	●	●	●	●	asparaginase activity	cellular response to nitrogen starvation; asparagine
YLR462W	**YLR462W**	●	●	●	●	molecular_function unknown	biological_process unknown
YLR463C		●	●	●	●	molecular_function unknown	biological_process unknown
YLR464W	**YLR464W**	●	●	●	●	molecular_function unknown	biological_process unknown
YML040W	**YML040W**	●	●	●	●	protein binding; RNA binding	Ty element transposition
YML045W	**YML045W**	●	●	●	●	protein binding; RNA binding;	Ty element transposition
YMR046C	**YMR046C**	●	●	●	●	protein binding; RNA binding	Ty element transposition
YMR051C	**YMR051C**	●	●	●	●	protein binding; RNA binding	Ty element transposition
YNL335W	**YNL335W**	●	●	●	●	molecular_function unknown	biological_process unknown
YNL336W	**COS1**	●	●	●	●	molecular_function unknown	biological_process unknown
YNR068C	**YNR068C**	●	●	●	●	molecular_function unknown	biological_process unknown
YOR392W		●	●	●	●	molecular_function unknown	biological_process unknown
YPL257W	**YPL257W**	●	●	●	●	molecular_function unknown; protein	biological_process unknown; Ty element transposition
YPR202W	**YPR202W**	●	●	●	●	molecular_function unknown	biological_process unknown
YPR203W	**YPR203W**	●	●	●	●	molecular_function unknown	biological_process unknown
YAL027W	**YAL027W**	●	●	●		3'-flap-structured DNA binding; 5'-flap-	double-strand break repair via single-strand annealing,
YAL066W		●	●	●		molecular_function unknown	biological_process unknown
YAR009C	**YAR009C**	●	●	●		protein binding; RNA binding;	Ty element transposition
YAR027W	**UIP3**	●	●		●	molecular_function unknown	biological_process unknown
YAR064W	**YAR064W**	●		●	●	molecular_function unknown	biological_process unknown
YBR184W	**YBR184W**	●	●		●	molecular_function unknown	biological_process unknown
YCL073C	**YCL073C**		●	●	●	transporter activity	transport
YDL242W		●	●		●	molecular_function unknown	biological_process unknown
YDL248W	**COS7**	●	●	●		receptor activity	biological_process unknown

YDR543C		●	●	●		molecular_function unknown	biological_process unknown
YEL067C			●	●	●	molecular_function unknown	biological_process unknown
YEL072W	**RMD6**	●	●		●	molecular_function unknown	biological_process unknown
YFL-TYA		●	●	●			
YFL-TYB		●	●	●			
YFL061W	**YFL061W**	●	●		●	molecular_function unknown	biological_process unknown
YFL067W	**YFL067W**	●	●		●	molecular_function unknown	biological_process unknown
YGL057C	**YGL057C**	●	●	●		molecular_function unknown	biological_process unknown
YGL259W	**YPS5**	●	●	●		aspartic-type endopeptidase activity	biological_process unknown
YGR295C	**COS6**	●	●		●	molecular_function unknown	biological_process unknown
YHL041W		●	●		●	molecular_function unknown	biological_process unknown
YHL044W	**YHL044W**		●	●	●	molecular_function unknown	biological_process unknown
YHR022C	**YHR022C**	●		●	●	molecular_function unknown	biological_process unknown
YHR139C-A		●		●	●	molecular_function unknown	biological_process unknown
YIL029C	**YIL029C**	●	●		●	molecular_function unknown	biological_process unknown
YIL114C	**POR2**		●	●	●	voltage-dependent ion-selective channel	ion transport
YIL174W	**YIL174W**	●	●	●		molecular_function unknown	biological_process unknown
YKL225W		●	●	●		molecular_function unknown	biological_process unknown
YLL017W	**SDC25**	●	●		●	Ras guanyl-nucleotide exchange factor	Ras protein signal transduction
YLR466W	**YRF1-4**	●	●	●		DNA helicase activity	telomerase-independent telomere maintenance
YML039W	**YML039W**	●	●	●		protein binding; RNA binding;	Ty element transposition
YML132W	**COS3**	●	●	●		molecular_function unknown	cellular sodium ion homeostasis
YML133C	**YML133C**	●	●	●		helicase activity	biological_process unknown
YOL165C	**AAD15**	●	●		●	aryl-alcohol dehydrogenase activity	aldehyde metabolism
YOR105W		●	●	●		molecular_function unknown	biological_process unknown
YOR390W	**YOR390W**	●	●	●		molecular_function unknown	biological_process unknown
YPL283C	**YRF1-7**	●	●	●		DNA helicase activity	telomerase-independent telomere maintenance
YPR197C			●	●	●	molecular_function unknown	biological_process unknown
YAL026C	**DRS2**	●	●			ATPase activity; phospholipid-	post-Golgi transport; intracellular protein transport;

YAL069W		●	●			molecular_function unknown	biological_process unknown
YAR010C	**YAR010C**	●	●			protein binding; RNA binding	Ty element transposition
YAR047C		●	●			molecular_function unknown	biological_process unknown
YBL071C		●	●			molecular_function unknown	biological_process unknown
YBL113C	**YBL113C**	●	●			helicase activity	biological_process unknown
YBR012W-A	**YBR012W-A**	●	●			protein binding; RNA binding	Ty element transposition
YBR064W		●	●			molecular_function unknown	biological_process unknown
YCL020W	**YCL020W**	●	●			protein binding; RNA binding	Ty element transposition
YCL056C	**YCL056C**	●	●			molecular_function unknown	peroxisome organization
YCR001W		●	●			molecular_function unknown	biological_process unknown
YCR103C	**YCR103C**	●			●	molecular_function unknown	biological_process unknown
YDL026W		●	●			molecular_function unknown	biological_process unknown
YDL027C	**YDL027C**			●	●	molecular_function unknown	biological_process unknown
YDL068W		●	●			molecular_function unknown	biological_process unknown
YDR048C		●	●			molecular_function unknown	biological_process unknown
YDR360W		●	●			molecular_function unknown	biological_process unknown
YDR366C	**YDR366C**	●	●			molecular_function unknown	biological_process unknown
YDR469W	**SDC1**	●	●			chromatin binding; histone-lysine N-	chromatin silencing at telomere; histone methylation
YDR541C	**YDR541C**	●			●	dihydrokaempferol 4-reductase activity	biological_process unknown
YEL028W		●		●		molecular_function unknown	biological_process unknown
YEL045C				●	●	molecular_function unknown	biological_process unknown
YER066W	**YER066W**	●	●			molecular_function unknown	biological_process unknown
YER135C		●	●			molecular_function unknown	biological_process unknown
YER181C		●	●			molecular_function unknown	biological_process unknown
YER184C	**YER184C**			●	●	sequence-specific DNA binding	biological_process unknown
YER185W	**YER185W**			●	●	molecular_function unknown	heme transport
YER188W				●	●	molecular_function unknown	biological_process unknown
YFL055W	**AGP3**			●	●	amino acid transporter activity	amino acid transport
YFL057C	**AAD16**	●	●			aryl-alcohol dehydrogenase activity	aldehyde metabolism

YFL063W		●	●			molecular_function unknown	biological_process unknown
YFL065C	**YFL065C**	●	●			molecular_function unknown	biological_process unknown
YFR020W				●	●	molecular_function unknown	biological_process unknown
YFR026C	**YFR026C**			●	●	molecular_function unknown	endoplasmic reticulum unfolded protein response
YFR043C	**YFR043C**		●		●	molecular_function unknown	biological_process unknown
YFR054C				●	●	molecular_function unknown	biological_process unknown
YGL052W		●	●			molecular_function unknown	biological_process unknown
YGL108C	**YGL108C**			●	●	molecular_function unknown	biological_process unknown
YGL182C		●			●	molecular_function unknown	biological_process unknown
YGL223C	**COG1**			●	●	molecular_function unknown	retrograde transport, vesicle recycling within Golgi; intra-
YGL230C	**YGL230C**	●	●			molecular_function unknown	biological_process unknown
YGL258W	**VEL1**		●	●		molecular_function unknown	biological_process unknown
YGL263W	**COS12**	●	●			molecular_function unknown	biological_process unknown
YGR039W		●	●			molecular_function unknown	biological_process unknown
YGR122W	**YGR122W**			●	●	molecular_function unknown	negative regulation of transcription from RNA polymerase
YGR269W			●	●		molecular_function unknown	biological_process unknown
YHL005C		●			●	molecular_function unknown	biological_process unknown
YHL042W	**YHL042W**			●	●	molecular_function unknown	biological_process unknown
YHR035W	**YHR035W**			●	●	molecular_function unknown	biological_process unknown
YHR081W	**LRP1**	●			●	double-stranded DNA binding; double-	double-strand break repair via nonhomologous end-
YHR153C	**SPO16**	●			●	molecular_function unknown	sporulation (sensu Saccharomyces)
YHR156C	**LIN1**		●		●	protein binding	biological_process unknown
YHR214C-B	**YHR214C-B**	●	●			protein binding; RNA binding; peptidase	Ty element transposition
YIL082W			●	●		DNA binding	transposition, RNA-mediated
YIL092W	**YIL092W**			●	●	molecular_function unknown	biological_process unknown
YIL163C		●			●	molecular_function unknown	biological_process unknown
YIL167W	**SDL1**			●	●	L-serine ammonia-lyase activity	serine family amino acid metabolic process
YIR005W	**IST3**		●		●	pre-mRNA splicing factor activity	spliceosome assembly
YJL023C	**PET130**		●		●	molecular_function unknown	biological_process unknown

YJL086C		●	●			molecular_function unknown	biological_process unknown
YJL150W		●	●			molecular_function unknown	biological_process unknown
YJL181W	**RBH1**			●	●	molecular_function unknown	biological_process unknown
YJL202C		●	●			molecular_function unknown	biological_process unknown
YJR027W	**YJR027W**	●	●			protein binding; RNA binding; peptidase	Ty element transposition
YJR029W	**YJR029W**	●	●			protein binding; RNA binding; DNA-	Ty element transposition
YKL084W	**HOT13**			●	●	zinc ion binding	protein import into mitochondrial intermembrane space
YKL115C				●	●	molecular_function unknown	biological_process unknown
YKL131W				●	●	molecular_function unknown	biological_process unknown
YKL147C				●	●	molecular_function unknown	biological_process unknown
YKL223W			●	●		molecular_function unknown	biological_process unknown
YKR106W	**YKR106W**			●	●	transporter activity	transport
YLL065W		●	●			not yet annotated	biological_process unknown
YLR046C	**YLR046C**	●	●			molecular_function unknown	biological_process unknown
YLR112W				●	●	molecular_function unknown	biological_process unknown
YLR159W	**YLR159W**	●	●			molecular_function unknown	biological_process unknown
YLR465C	**BSC3**	●	●			molecular_function unknown	biological_process unknown
YLR458W				●	●	molecular_function unknown	biological_process unknown
YMR045C	**YMR045C**	●	●			protein binding; RNA binding; peptidase	Ty element transposition
YMR050C	**YMR050C**	●	●			protein binding; RNA binding; peptidase	Ty element transposition
YMR193C-A				●	●	molecular_function unknown	biological_process unknown
YMR316C-A				●	●	molecular_function unknown	biological_process unknown
YMR324C				●	●	molecular_function unknown	biological_process unknown
YMR326C		●	●			molecular_function unknown	biological_process unknown
YNL018C	**YNL018C**			●	●	molecular_function unknown	biological_process unknown
YNL028W				●	●	molecular_function unknown	biological_process unknown
YNL042W	**BOP3**	●	●			molecular_function unknown	cellular response to methylmercury
YNL195C	**YNL195C**			●	●	molecular_function unknown	biological_process unknown
YNL331C	**AAD14**	●	●			aryl-alcohol dehydrogenase activity	aldehyde metabolism

YNL339C	**YRF1-6**	●	●			DNA helicase activity	telomerase-independent telomere maintenance
YNR004W	**SWM2**	●			●	molecular_function unknown	biological_process unknown
YNR010W	**CSE2**	●	●			RNA polymerase II transcription mediator	mitotic chromosome segregation; transcription from Pol II
YNR044W	**AGA1**			●	●	cell adhesion receptor activity	agglutination during conjugation with cellular fusion
YNR065C	**YSN1**			●	●	molecular_function unknown	biological_process unknown
YNR075W	**COS10**			●	●	molecular_function unknown	endocytosis
YNR077C		●	●			molecular_function unknown	biological_process unknown
YOL015W	**IRC10**			●	●	molecular_function unknown	biological_process unknown
YOL031C	**SIL1**		●		●	adenyl-nucleotide exchange factor	SRP-dependent cotranslational membrane targeting,
YOL079W		●	●			molecular_function unknown	biological_process unknown
YOL106W		●	●			molecular_function unknown	biological_process unknown
YOL160W		●	●			molecular_function unknown	biological_process unknown
YOR015W		●	●			molecular_function unknown	biological_process unknown
YOR024W				●	●	molecular_function unknown	biological_process unknown
YOR044W	**IRC23**			●	●	molecular_function unknown	biological_process unknown
YOR189W	**IES4**			●	●	molecular_function unknown	response to DNA damage stimulus
YOR318C		●	●			molecular_function unknown	biological_process unknown
YOR385W	**YOR385W**		●		●	molecular_function unknown	biological_process unknown
YOR387C	**YOR387C**		●	●		molecular_function unknown	biological_process unknown
YOR391C	**HSP33**	●	●			cysteine-type peptidase activity; unfolded	biological_process unknown
YPL168W	**YPL168W**			●	●	molecular_function unknown	biological_process unknown
YPL192C	**PRM3**			●	●	molecular_function unknown	karyogamy
YPL200W	**CSM4**			●	●	molecular_function unknown	meiotic chromosome segregation
YPL279C	**YPL279C**	●			●	molecular_function unknown	biological_process unknown
YPL280W	**HSP32**	●	●			cysteine-type peptidase activity; unfolded	biological_process unknown
YPR071W	**YPR071W**			●	●	molecular_function unknown	biological_process unknown
YPR077C				●	●	molecular_function unknown	biological_process unknown
YPR198W	**SGE1**		●		●	xenobiotic-transporting ATPase activity	response to drug; drug transport
YPR204W	**YPR204W**	●	●			DNA helicase activity	biological_process unknown

YAL055W	**PEX22**				●	protein anchor	protein-peroxisome targeting
YAL064W	**YAL064W**			●		molecular_function unknown	biological_process unknown
YAR029W	**YAR029W**	●				molecular_function unknown	biological_process unknown
YAR052C					●		
YBL012C			●			molecular_function unknown	biological_process unknown
YBL029W	**YBL029W**	●		●		molecular_function unknown	biological_process unknown
YBL073W				●		molecular_function unknown	biological_process unknown
YBR037C	**SCO1**		●			copper ion binding; thioredoxin	protein complex assembly; copper ion transport
YBR063C	**YBR063C**				●	molecular_function unknown	biological_process unknown
YBR174C			●			molecular_function unknown	biological_process unknown
YCL062W	**VAC17**		●			protein anchor	vacuole inheritance
YCL075W	**YCL075W**			●		protein binding; RNA binding; peptidase	Ty element transposition
YCR105W	**ADH7**				●	alcohol dehydrogenase (NADP+) activity	alcohol metabolism
YDL241W	**YDL241W**				●	molecular_function unknown	biological_process unknown
YDR024W	**FYV1**	●				molecular_function unknown	biological_process unknown
YDR042C	**YDR042C**				●	molecular_function unknown	biological_process unknown
YDR149C				●		molecular_function unknown	biological_process unknown
YDR241W	**BUD26**		●			molecular_function unknown	biological_process unknown
YDR305C	**HNT2**				●	nucleoside-triphosphatase activity;	nucleoside catabolism
YDR488C	**PAC11**			●		microtubule motor activity	microtubule-based process
YDR506C	**YDR506C**			●		molecular_function unknown	Meiosis; synaptonemal complex organization
YEL014C			●			molecular_function unknown	biological_process unknown
YER028C	**MIG3**			●		transcription factor activity; DNA binding	transcription initiation
YER034W	**YER034W**			●		molecular_function unknown	biological_process unknown
YER063W	**THO1**				●	Chromatin binding; double-stranded DNA	transcription, DNA-dependent
YER092W	**IES5**			●		molecular_function unknown	biological_process unknown
YER163C	**GCG1**		●			gamma-glutamylcyclotransferase activity	glutathione catabolic process
YFL015C				●		molecular_function unknown	biological_process unknown
YFL025C	**BST1**	●				phosphatidylinositol deacylase activity	vesicle organization and biogenesis; ER-associated

YFL041W	**FET5**		●			ferroxidase activity	iron ion transport
YFL049W	**YFL049W**	●				molecular_function unknown	positive regulation of transcription from RNA polymerase II
YFR018C	**YFR018C**			●		molecular_function unknown	biological_process unknown
YFR056C		●				molecular_function unknown	biological_process unknown
YFR057W	**YFR057W**			●		molecular_function unknown	biological_process unknown
YGL081W	**YGL081W**			●		molecular_function unknown	biological_process unknown
YGL109W					●	molecular_function unknown	biological_process unknown
YGL117W	**YGL117W**			●		molecular_function unknown	biological_process unknown
YGL154C	**LYS5**			●		holo-[acyl-carrier protein] synthase	lysine biosynthesis, aminoadipic pathway; protein-cofactor
YGL204C					●	molecular_function unknown	biological_process unknown
YGL231C	**EMC4**		●			molecular_function unknown	protein folding in endoplasmic reticulum
YGL249W	**ZIP2**			●		molecular_function unknown	synaptonemal complex formation
YGR107W					●	molecular_function unknown	biological_process unknown
YGR153W	**YGR153W**				●	molecular_function unknown	biological_process unknown
YGR224W	**AZR1**			●		azole transporter activity	azole transport
YGR263C	**SAY1**	●				steryl deacetylase activity	response to toxic sustance; sterol deacetylation; sterol
YGR288W	**MAL13**			●		transcription factor activity	regulation of transcription, DNA-dependent; carbohydrate
YGR289C	**MAL11**				●	alpha-glucoside:hydrogen symporter	alpha-glucoside transport; trehalose transport
YHL008C	**YHL008C**			●		transporter activity	biological_process unknown
YHR056C	**RSC30**	●				DNA binding	regulation of transcription, DNA-dependent
YHR080C	**YHR080C**			●		molecular_function unknown	biological_process unknown
YHR130C		●				molecular_function unknown	biological_process unknown
YHR198C	**FMP22**			●		molecular_function unknown	biological_process unknown
YHR213W	**YHR213W**			●		molecular_function unknown	biological_process unknown
YIL012W					●	molecular_function unknown	biological_process unknown
YIL028W				●		molecular_function unknown	biological_process unknown
YIL032C				●		molecular_function unknown	biological_process unknown
YIL059C					●	molecular_function unknown	biological_process unknown
YIL079C	**AIR1**	●				polynucleotide adenylyltransferase	mRNA-nucleus export

YIL080W	**YIL080W**			●		protein binding; RNA binding; peptidase	Ty element transposition
YIL089W	**YIL089W**				●	molecular_function unknown	biological_process unknown
YIL113W	**SDP1**				●	MAP kinase phosphatase activity	MAPKKK cascade during cell wall biogenesis
YIL117C	**PRM5**				●	molecular_function unknown	conjugation with cellular fusion
YIL161W	**YIL161W**				●	molecular_function unknown	biological_process unknown
YIL168W	**SDL1**				●	L-serine ammonia-lyase activity	serine family amino acid metabolism
YIR021W	**MRS1**			●		RNA binding; endodeoxyribonuclease	Group I intron splicing
YJL135W					●	molecular_function unknown	biological_process unknown
YJL203W	**PRP21**				●	RNA binding	nuclear mRNA splicing, via spliceosome
YJR003C	**YJR003C**			●		molecular_function unknown	biological_process unknown
YJR083C	**ACF4**				●	molecular_function unknown	actin cytoskeleton organization and biogenesis
YJR149W	**YJR149W**			●		molecular_function unknown	biological_process unknown
YJR162C			●			molecular_function unknown	biological_process unknown
YKL023W	**YKL023W**				●	molecular_function unknown	biological_process unknown
YKL044W		●				molecular_function unknown	biological_process unknown
YKL123W				●		molecular_function unknown	biological_process unknown
YKR022C	**NTR2**				●	molecular_function unknown	nuclear mRNA splicing, via spliceosome
YLL033W	**RRG4**	●				molecular_function unknown	Ascospore formation; mitotic recombination
YLL037W					●	molecular_function unknown	biological_process unknown
YLL051C	**FRE6**	●				ferric-chelate reductase activity	copper ion import
YLL059C					●	molecular_function unknown	biological_process unknown
YLR010C	**TEN1**				●	single-stranded telomeric DNA binding	telomere capping
YLR031W	**YLR031W**			●		molecular_function unknown	biological_process unknown
YLR122C					●	molecular_function unknown	biological_process unknown
YLR124W					●	molecular_function unknown	biological_process unknown
YLR282C				●		molecular_function unknown	biological_process unknown
YLR311C			●			molecular_function unknown	biological_process unknown
YLR334C			●			molecular_function unknown	biological_process unknown
YLR400W				●		molecular_function unknown	biological_process unknown

YML036W	**CGI121**		●			molecular_function unknown	positive regulation of transcription from RNA polymerase II
YML047W-A					●		
YML050W	**AIM32**		●			molecular_function unknown	biological_process unknown
YML090W			●			molecular_function unknown	biological_process unknown
YMR018W	**YMR018W**			●		molecular_function unknown	biological_process unknown
YMR067C	**CUI1**				●	protein complex binding	ER-associated protein catabolic process; sporulation
YMR069W	**NAT4**				●	histone acetyltransferase activity	histone acetylation
YMR075C-A					●	molecular_function unknown	biological_process unknown
YMR086C-A					●	molecular_function unknown	biological_process unknown
YMR213W	**CEF1**			●		pre-mRNA splicing factor activity	nuclear mRNA splicing, via spliceosome
YMR299C	**YMR299C**			●		contributes to motor activity	nuclear migration along microtubule
YMR320W				●		molecular_function unknown	biological_process unknown
YMR322C	**SNO4**			●		cysteine-type peptidase activity; unfolded	biological_process unknown
YMR323W	**YMR323W**			●		phosphopyruvate hydratase activity	biological_process unknown
YNL034W	**YNL034W**			●		molecular_function unknown	biological_process unknown
YNL041C	**COG6**	●				molecular_function unknown	intra-Golgi transport
YNL146W	**YNL146W**				●	molecular_function unknown	biological_process unknown
YNL149C	**PGA2**				●	molecular_function unknown	protein transport
YNL171C		●				molecular_function unknown	biological_process unknown
YNL196C	**SLZ1**				●	molecular_function unknown	biological_process unknown
YNL260C	**LTO1**		●			molecular_function unknown	Ribosomal large subunit biogenesis; translational initiation
YNL285W			●			molecular_function unknown	biological_process unknown
YNL337W				●		molecular_function unknown	biological_process unknown
YNR017W	**MAS6**				●	protein transporter activity	mitochondrial matrix protein import
YNR024W	**MPP6**			●		poly(U) RNA binding	exonucleolytic trimming to generate mature 3'-end of 5.8S
YNR025C					●	molecular_function unknown	biological_process unknown
YNR030W	**ECM39**	●				alpha-1,6-mannosyltransferase activity	protein amino acid glycosylation; dolichol-linked
YNR034W	**SOL1**		●		●	6-phosphogluconolactonase activity	tRNA processing
YNR059W	**MNT4**				●	alpha-1,3-mannosyltransferase activity	O-linked glycosylation

YNR061C	**YNR061C**				●	molecular_function unknown	biological_process unknown
YNR066C	**YNR066C**				●	molecular_function unknown	biological_process unknown
YNR069C	**BSC5**				●	molecular_function unknown	biological_process unknown
YNR074C	**YNR074C**			●		oxidoreductase activity, acting on NADH	response to singlet oxygen
YOL024W	**YOL024W**				●	molecular_function unknown	biological_process unknown
YOL047C	**YOL047C**			●		molecular_function unknown	biological_process unknown
YOL082W	**ATG19**				●	protein binding	protein-vacuolar targeting
YOL085C					●	molecular_function unknown	biological_process unknown
YOL114C	**YOL114C**				●	molecular_function unknown	biological_process unknown
YOL131W	**YOL131W**				●	molecular_function unknown	biological_process unknown
YOL154W	**ZPS1**			●		molecular_function unknown	biological_process unknown
YOL162W	**YOL162W**				●	transporter activity	transport
YOR034C	**AKR2**				●	palmitoyltransferase activity	Endocytosis; protein palmitoylation
YOR072W					●	molecular_function unknown	biological_process unknown
YOR183W	**FYV12**			●		molecular_function unknown	biological_process unknown
YOR263C						molecular_function unknown	biological_process unknown
YOR288C	**MPD1**	●				protein disulfide isomerase activity	protein folding
YOR308C	**SNU66**			●		pre-mRNA splicing factor activity	nuclear mRNA splicing, via spliceosome
YOR371C	**GPB1**	●				signal transducer activity	signal transduction
YOR373W	**NUD1**	●				structural constituent of cytoskeleton	microtubule nucleation
YOR393W	**ERR1**		●			phosphopyruvate hydratase activity	biological_process unknown
YPL068C	**YPL068C**			●		molecular_function unknown	biological_process unknown
YPL072W	**UBP16**				●	ubiquitin-specific protease activity	protein deubiquitination
YPL073C					●	molecular_function unknown	biological_process unknown
YPL281C	**ERR2**				●	phosphopyruvate hydratase activity	biological_process unknown
YPR193C	**HPA2**				●	histone acetyltransferase activity	histone acetylation
YPR199C	**ARR1**				●	RNA polymerase II transcription factor	positive regulation of transcription from Pol II promoter

VNIVERSITAT DE VALÈNCIA