Evolutionary Impact of Life Cycle and Mating System on Genetic and Functional Diversity in *Zygosaccharomyces rouxii* complex strains

A THESIS SUBMITTED TO THE

Ph.D. School in AgriFood Sciences, Technologies And Biotechnologies
Department of Life Sciences
UNIVERSITY OF MODENA AND REGGIO EMILIA
FOR THE DEGREE OF DOCTORAL OF PHILOSOPHY

Supervisor:
Prof. PAOLO GIUDICI
Co-supervisor:
Dr. LISA SOLIERI

Ph.D. student:
TIKAM CHAND DAKAL
(Matr. n. 60349)

Dean of the Ph.D. School:
Prof. ANDREA PULVIRENTI

Academic Year 2011-2014

I dedicate this thesis to my mother……

Table of contents

Acknowledgement

First and foremost, I thank my Ph.D. supervisor, Prof. Paolo Giudici, for introducing me to the fascinating world of genes and chromosomes. Words are insufficient to express my deep and sincere gratitude to him for his all time friendly approach, support, valuable guidance (that would be helpful in both, research and professional career), deep scholarly insights (especially his opinions regarding the existence of hybrids, HGT, non-genetic variations, less degree of genetic variations in *Saccharomyces* impressed me lot of time). His constant motivations (by saying I need papers… I am starving..!!) always kept my momentum high during my entire period, which led to the completion of this work successfully.

"I consider it an honor to work under his guidance"

With immense pleasure, I owe indebtedness and obligation to my Ph.D. co-supervisor, Dr. Lisa Solieri, for her dynamic guidance throughout my Ph.D. research period with dedicated and systematic approach, extensive discussions, expertise comments, constant encouragement, scholarly insight and friendly concern.

I will always remember all the efforts she put into my work

"Indeed, without her guidance and support, I would not be able to put the thesis together."

I share the credit of my work with Dr. Lisa Solieri

I would like to thank Dr. Luciana de Vero, for her generous, polite, friendly nature. I will always remember her for emotional encouragement during my hard time.

She stays happy and make other happy too

Whenever I will drink coffee, eat chocolates or ice creams, I will always remember her

I would like to thank Prof. Andrea Pulvirenti (Dean, Ph.D. School) for his *wonderful personality and nature* that is difficult to find. He always helped me, encouraged me whenever I needed…..

'**Thank you very much Sir**'

I would also like to thank Dr. Maria Gullo for her **polite nature, friendly behavior**, and her husband for liking the **'Indian Tea'**.

I am also thankful to Dr. Stefano Cassanelli. He helped me, whenever I needed his assistance (especially during many **emergencies**). I will always appreciate his efforts that he put into my work, especially in teaching me the Southern blotting experiments and characterization of mating type gene. **For me, he is among those persons who are highly dedicated to research.**

My heartfelt thanks to caring seniors: **Dr. Federico Lemmetti, Dr. Elena Verzelloni, Dr. Marwa Al Moghazi, Dr. Silvio Boveri, Dr. Davide Giovanardi and Dr. Veronica Sberveglieri**.

My special thanks to Francesco Mezzetti, my true Italian friend, colleague and little brother, for his all time support. Thanks to my previous lab mate and friends, Aldo Bianchi, for his caring nature during my starting period in Italy. Also thanks to Luca Laviano (*my wonderful and best friend from Bari,* ex-room mate) for always helping me and reminding me that there is a world outside the Lab.

The following people all helped me during my time at the Unimore, whether with scientific advice, discussions or personal/practical help: Dr. Denise Garavini (Sorveglianza Sanitaria, Azienda Ospedaliera Universitaria Policlinico di Modena, for her **generous nature**), Dr. Letizia Musto (In-charge, Student's Apartments-Via. Mascagni, **many many thanks** for helping me and for making my stay comfortable and peaceful in Via. Mascagni), Dr. Luisa Antonella Volpelli (for her generous nature, **beautiful & nobel heart**, and of course, for her technical support during Seminars). I am also grateful to all my Ph.D. batchmates, especially Carlotta Ferrari, Davide Saldo and Valentina Monti, for their wonderful friendship and filling joy in my research life.

My heartfelt thanks to everyone in our department (Via Kennedy), for making my stay enjoyable and memorable in Italy. I really feel pleasure meeting you all.

I thank all my friends from different countries (America, Egypt, Spain, Italy, and Turkey) who were my neighbors at Student's residence, Via Mascagni. I had a great time and fun with them (**the golden period of Via Mascagni**). Especially, I miss all my Turkish friends (**my Super Super Best friends**), Italian friends (Lorenzo, Sara, Onofrio and Elio), Kurdish friend Ramazan Düzen (**my little brother**), Karim Ibrahim (**my fantastic Indian-Portuguese party friend**), and my lovely Jordanian friend, Thaer Zaitoun (**ghost of Via Mascagni**), for his wonderful friendship, love and care.

Thanks to Dr. Subhash Maharshi (a person full of wisdom) for his wonderful company and nice friendship. Tons and millions of thanks to Santokh Singh Thandi (Sukha paji) and Devindra Singh (Dev paji) for taking care of my health, and helping me in cooking during my last months.

Thanks to Indian friends and teachers……

I really don't want hundreds of friends on Facebook nor do I want thousands of followers of Twitter. Instead, I want some real friends and well wishers as below:

Akshat Rana (noble and the most handsome friend I have), Anil Kalia (my money Bank, nice supporting friend), Dr. Arshad Jawed (thanks for teaching me Protein purification), Aurobind Vidyarathi (wonderful friend), Balveer S. Rathore, Chandresh Upadhyay, Gurdeep Singh (Jane'ann, best and supporting friend), Dr. Hemant K. Verma, Jitendra Pareek, Kamal Jain, Dr. Kamna Ailawadi (nice and supporting senior), Kshipra Sharma, Lipsy Chopra (beautiful little sister), Love Jain, Manisha Sharma (memorable friend), Mukesh Yadav, Nancy George, Dr. Narendra K. Sharma (thanks for improving my Proteomics skills), Nandita Kohli, Narendra Goyal, Navjot Athwal, Dr. Neelam Khatri (most admirable friend), Nishant Thakur, Pankaj Pujari, Payal Sanadhaya (best supporting friend), Rajendra Verma (thanks for improving my Bioinfo skills), Ramit Rikhi, Satbir Singh, Suksham Pal (thanks for listening my adventurous stories), Dr. Vaibhav Pandya, Vikas Choudhary (Jane'ann, best and oldest friend), Vikas CLC, Vikas Sharma, Dr. Vikrant Upadhyay, and Dr. Vinod Yadav (Bhallu, a fantastic friend and person).

Especially, Akshat, Anil, Gurdeep, Lipsy, Neelam, Payal, Vikas JPC and Vinod, their best wishes, encouragement and support made my Ph.D. journey easy and possible.....Thanks for adding truck loads of joy and happiness in my life..... I feel proud and happy to have friends like them.

Thanks to all my SB-07 IMTech-batchmates….. (**Great future scientists**) and all my supporting Seniors, Scientists (Especially Dr. Anirban Rai Chaudhary, Dr. Shanmugam Mayilraj, Dr. Swaranjit S. Cameotra and Dr. Vijay Sonawane) and Staff members at IMTech.

I also thanks my Professors who trained me during my Master Degree at MLS Univ., Udaipur: Dr. S. D. Purohit (Head, Dept of Biotech), Dr. Harshada Joshi, Dr. Shailendra Vyas, Dr. Kanika Sharma, and Dr. D. K. Bhatt; and during my Master project (Dr. Ravindra Kumar, Hindu College, Univ. of Delhi). Many thanks to Dr. Sujata Magdum (KTHM College, Nashik) for her wonderful nature, best wishes and encouragement.

Thanks to my family members and relatives……

This thesis would be just impossible without the blessing from Maa Sherawali and my family members. I thank my family for all their blessings and support, especially my mom-dad, who invested their whole life for my education. That's the reason….**I like to dedicate this thesis to my Mom**

I am thankful for my elder sisters, Sita didi, Sushama didi and Meena didi; and my younger brother, Suresh K. Dakal and his family (especially, cute "*Aalu*"), for their all time support, encouragement and love. Also, I thank my ~~elder brother~~ uncle Brajesh Arya, for his unparallel love, care, and for constantly encouraging me to enjoy life in Italy.

My heartiest thanks to my in-laws for their blessing, love and supports. Thanks to Devika Nagar (Mannu), Chetan Nagar and all his friends (Abhinav, Nemi and Chetan Kumawat).

Last but not least, I thank my loving wife and cute daughter Priyanshi ("Nannu/Babu/Didi") for caring me, loving me, boosting me up with their love, and for standing all time beside me, all the time, all the way, whether it's a good time or a hard time. **-Jai Mata Di**

List of publications

Journal publications

1. Solieri L, **Dakal TC**, Giudici P (2012) Next-generation sequencing and its potential impact on food microbial genomics. Ann Microbiol 63: 21-37, 2013. **doi 10.1007/s13213-012-0478-8**

2. Solieri L, **Dakal TC**, Giudici P (2013a) *Zygosaccharomyces sapae* sp. nov., a novel yeast species isolated from Italian traditional balsamic vinegar. Int J Syst Evol Microbiol 63: 364-371. **doi:10.1099/ijs.0.043323-0**

3. Solieri L, **Dakal TC**, Croce MA, Giudici P (2013b) Unraveling genomic diversity of the *Zygosaccharomyces rouxii* complex with a link to its life cycle. FEMS Yeast Res 13: 245-258.

4. Solieri L, **Dakal TC**, Bicchiato S (2014) Quantitative phenotypic analysis of multi-stress response in *Zygosaccharomyces*. FEMS Yeast Res. Manuscript accepted.

5. **Dakal TC**, Solieri L, Giudici P. (2014) Osmotolerance and halotolerance in yeast genus *Zygosaccharomyces*. International Journal of Food Microbiology. Manuscript accepted.

6. Solieri L, **Dakal TC**, Giudici P, Cassanelli S (2014) Sex-determination system in diploid *Zygosaccharomyces sapae.* G3 (Bethesda). Manuscript accepted.

Conference abstracts

1. **Dakal TC** (2013) Life cycle and reticulate evolution in *Zygosaccharomyces rouxii* complex. In: 18th Workshop on the Development of Italian PhD Research on Food Science Technology and Biotechnology (Edited by Corich V and Celotti E). pp. 75-79. 25-27 Sep, 2013. University of Padova and Udine, Conegliano, Italy.

Awarded: A grade in oral presentation

2. **Dakal TC** (2013) Getting insights from genomic complexity in *Zygosaccharomyces rouxii* complex. In: 26th International Conference on Yeast Genetics and Molecular Biology. 29 Aug- 03 Sep 2013, Geothe University, Frankfurt, Germany.

Awarded: FEMS Young Scientist Meeting Grant

3. **Dakal TC** (2012) *Zygosaccharomyces sapae*: from species delineation to mating type genes characterization. In: 17th Workshop on the Development of Italian PhD Research on Food Science Technology and Biotechnology (Edited by M Petracci and T G Toschi). pp. 205-206. 19-21 Sep, 2012. Cesena, Italy (ISBN 978-88-6541-224-4)

4. Solieri L, **Dakal TC**, Giudici P (2012) *Zygosaccharomyces sapae* from sugary-low acid niches: delineation of a new yeast species. In: Third National Conference SIMTREA. 26-28 Jun, 2012, Italy.

5. Solieri L, **Dakal TC**, de Vero L (2011) Complex nature of *Zygosaccharomyces* wild strains: a genotypic and functional characterization. In: Proceedings of the First International Conference on Microbial diversity 2011: Environmental Stress and Adaptation (Edited by A Casella, D. Daffonchio, M. Gobbetti, and E. Parente). pp. 33-36. 26-28 Oct, 2011. Milan, Italy.

List of figures

Supplementary Figures

List of tables

Supplementary tables

Abstract of the thesis

Zygosaccharomyces rouxii traditionally includes osmotolerant and halotolerant haploid yeasts, which emerged from the *S. cerevisiae* lineage prior to a whole genome duplication (WGD) event occurred in Saccharomycetaceae family 100 million years ago. In this thesis, I applied multi-phasic approaches to elucidate the molecular basis of genomic complexity and functional diversity in a pool of selected *Zygosaccharomyces* strains, referred as *Z. rouxii* complex.

Firstly, I delineated a novel diploid species related to *Z. rouxii*, namely *Zygosaccharomyces sapae* (**Chapter 2**). Subsequently, a polyphasic comparative analysis of *Z. sapae* (strains ABT301^{T} and ABT601) with strains deposited in culture collection as *Z. rouxii* (CBS 732^{T}, ATCC 42981, CBS 4837, CBS 4838 and OUT7136) was performed, which revealed that the *Z. rouxii* complex strains show variability with respect to chromosome number, genome size, ploidy and sequences of housekeeping and phylogenetic markers (**Chapter 3**). My finding demonstrated that *Z. rouxii* complex could be split up into three lineage, namely haploid lineage *Z. rouxii* (CBS 732^{T}), diploid Z. sapae, and allodiploid/aneuploid lineage comprising ATCC 42981, CBS 4837, CBS 4838 and OUT 7136. Moreover, an extent pattern of rDNA heterogeneity was observed in intragenomic copies of rDNA array in all strains of *Z. rouxii* complex, with exception of CBS 732^{T}, with indels/substitutions occurring mainly in internal transcribed spacer regions (ITS). These finding indicated that the strains are defective in mechanisms driving intra-strain homogenization by concerted evolution, suggesting a birth and death process of evolution.

When a wider pool of *Z. rouxii* strains were screened on the basis of rDNA heterogeneity in D1-D2 26S and ITS-5.8S rDNA, 6 out of 78 strains showed lack of concerted evolution in rDNA array (**Chapter 6, Chapter 7**). The multiple sequence alignment of all cloned sequences identified

variability (mostly heterogeneous) at 103 sites. The majority of them grouped in two short regions, which we have referred as variable region 1 (VR1) and variable region 2 (VR2). Twenty seven variable sites of the D1 domain formed VR1 and sixty-six sites of the D2 domain comprised VR2. At maximum sites, twenty five sites in VR1 and sixty one sites in VR2, usually two nucleotides alternated. Moreover, at maximum of these variable sites, 15 sites (each in VR1 and VR2), the nucleotide remained conserved in all heterogeneous *Zygosaccharomyces* strains and alternation has taken place just with respect to *S. cerevisiae*. From comparative sequence analysis and *in-silico* RNA structure prediction, I found one variable site of VR1 and five in VR2 of *Z. rouxii* CBS 732^T (and corresponding sites in the aligned sequence from the heterogeneous strains) at position 91, 455, 456, 469, 476, and 516 paired with stable guanines in the hybridizing segment of the hairpin. Most of them had either T(U) or C in the cloned sequences. Their transitions did not alter the stem structure in many cloned sequences, confirming that wobble pairing of nucleotides did neutralize many substitutions indeed. However, in some cloned sequences the transition has structural alterations. So, the structural variability could be ascribed to substitutions in sites, where wobble pairing was not possible.

A quantitative profiling of *Z. rouxii* complex strains was carried out in response to 15 environmental perturbations using biostatistical tool "*grofit*" (**Chapter 5**). Allodiploid/aneuploid strains ATCC 42981, CBS 4837 and CBS 4838 displayed alkali metal cations resistance and glycerol respiration higher than *Z. sapae* $ABT301^T$ and AB601. μ-based logarithmic phenotypic index highlighted that ABT601 is a slow-growing organism insensitive to stress perturbation, whereas $ABT301^T$ grows fast on rich medium and is sensitive to sub-optimal conditions.

The life cycle and the mating system have been reported as the major evolutionary forces driving genetic and functional diversity in yeasts. In context to *Z. rouxii* complex strains, I

hypothesized that mating between the highly divergent haploid, heterothallic parent cells could have resulted in diploid progenies with high genome complexity that subsequently lost the ability to undergo meiosis and given rise to diploid (*Z. sapae*) and allodiploid/aneuploid (ATCC 42981, CBS 4837, CBS 4838, OUT 7136) strains. To test this hypothesis, I investigated sex-determination system in *Z. sapae* ABT301^{T} (**Chapter 4**). I cloned four mating type-like loci (*MTL*): three of them were divergent *MATα*-idiomorph loci (ZsMTLα2α1_1, ZsMTLα2α1_2, and ZsMTLα2α1_3) and the remaining one was a single ZsMTLa2a1. ABT301^{T} possesses two divergent *HO* endonuclease genes (92% and 100% aa sequence identities compared to *Z. rouxii*). To establish the three-cassette structure (*MAT*, *HML*, and *HMR*), I analyzed the flanking regions of Zs*MTLα* and ZsMTLa-idiomorph loci. The ZsMTLa2a1 locus and three ZsMTLα2α1 loci exhibited a conserved *DIC1-MAT-SLA2* organization, identical to *Z. rouxii MAT* expression locus. Moreover, four putative *HML* cassettes were identified: one containing sequence ZsMTLα2α1_1 and; the remaining three, HML_D_1, HML_D_2, and HML_D_3 harbouringZsMATα2α1_1, ZsMATα2α1_2, and ZsMATα2α1_3, respectively. In conclusion, ABT301^{T} is a highly heterozygous aααα strain which has lost *HMR*.

The results presented in this thesis globally provide evidences of evolutionary impact of life cycle and mating systems on genetic and functional diversity in *Z. rouxii* complex strains, which represents a promising model for studying genome evolution in pre-WGD yeasts species. I demonstrate that mating system in *Z. rouxii* complex is a hypermutagenic process that contributes to stress adaptation by generating progenies with different genetic and phenotypic outcomes, of which some might be industrially important.

Chapter 1:

Review of Literature

1.1. Genus *Zygosaccharomyces*: an overview

Genus *Zygosaccharomyces* comprises of non-conventional, osmotolerant/halotolerant, food yeast species that belong to Saccharomycetaceae family of large phylum Ascomycota placed under fungal kingdom (**Figure 1.1**). The family Saccharomycetaceae also accommodates the genus *Saccharomyces*, which includes the well-studied yeast *Saccharomyces cerevisiae*. The species of this genus have shown marked genetic and functional diversity when compared with species of other Saccharomycetaceae genus. The peculiar life cycle (**refer section 1.3.1**) and mating system (**refer section 1.3.2**) in this group of yeasts as well as exceptional tolerance to high osmotic conditions (**refer section 1.4**) have accelerated genetic studies in this genus.

Zygosaccharomyces	
Kingdom	Fungi
Phylum	Ascomycota
Sub-phylum	Saccharomycotina
Class	Saccharomycetes
Order	Saccharomycetales
Family	Saccharomycetaceae
Genus	*Zygosaccharomyces*

Figure 1.1. Taxonomic status of *Zygosaccharomyces*

Moreover, species of the genus *Zygosaccharomyces* are referred as a 'protoploid' for reason that they have diverged prior whole-genome duplication (WGD) event approximately 100 milion years ago from the *Saccharomyces* lineage, which have given rise to genus *Saccharomyces* 20-25 million years ago. In this sense genus *Zygosaccharomyces* is most closely related to putative ancestral genome of *S. cerevisiae*, and therefore, this genus has become an exciting pre-WGD model for studying genome evolution and speciation as an alternative to pre-WGD ancestral genome of *S. cerevisiae*.

Species Name	Type-strain code	Isolation source	Country of Isolation	Reference
Z. bailii	CBS 680*	Institute of Brewing	Japan	James and Stratford, 2011
Z. bisporus	CBS 702	--	Japan	James and Stratford, 2011
Z. kombuchaensis	CBS 8849	Kombucha tea	Russia	James and Stratford, 2011
Z. lentus	CBS 8574	Spoiled whole-organe drink	UK	James and Stratford, 2011
Z. mellis	CBS 736	Honey	USA	James and Stratford, 2011
Z. rouxii	CBS 732*	Concentrated grape must	Italy	James and Stratford, 2011
Z. gambellarensis	CBS 12191	Italian 'passito' style wine	Italy	Torriani et al., 2011
Z. machadoi	CBS 10264	Nest of the stingless bee *Tetragonisca angustula*	Brazil	Rosa and Lachance, 2005
Z. parabailii	CBS 12809	Wild-type isolate associated with fermentation	--	Suh et al., 2013
Z. pseudobailii	CBS 2856	Worcestershire sauce	Japan	Suh et al., 2013
Z. sapae#	CBS 12607	Italian traditional balsamic vinegar (TBV)	Italy	Solieri et al., 2013a
Z. siamensis	CBS 12273	Honey	Malasia	Saksinchai et al., 2012

* Neotype of the species indicated where type material of the species not exist
\# Species isolated and characterized in our lab (Solieri et al., 2007; Solieri et al., 2013a)
Rows in red shade are species described in current yeast taxonomy (Kurtzman et al., 2011) and in blue shade are recently isolated species

Table 1.1. The complete list of *Zygosaccharomyces* species identified so far

This genus, from taxonomical point of view, is very unstable. The taxonomic status and species number of genus *Zygosaccharomyces* have taken several moves in recent years. Approximately one dozen of species were known to belong to this genus. Following phylogenetic reconstruction by Kurtzman and Robnett (2013), the number of species in this genus reduced to six. However, exploration of some food niches led to delineation of several novel yeast species as mentioned in **Table 1.1**. Presently, *Zygosaccharomyces* genus comprises of 12 species. Besides this, several strains of this genus have been found as postulated hybrid, allodiploid and aneuploid strains. Altogether, it indicates that *Zygosaccharomyces* genus, as compared to other hemiascomycete yeasts genus, offer

great opportunity for studying many aspects of genome evolution, including in particular the whole-genome duplication, polyploidization, postulated hybridization (James et al., 2005) and gene transfers.

In this Ph.D. thesis, I investigated and discussed about the impact of evolutionary motive forces on genetic and functional diversity in a pool of *Zygosaccharomyces* strains comprising of haploid type strain *Z. rouxii* CBS 732^{T}, *Z. rouxii* allodiploid ATCC 42981, diploid *Z. sapae*, and some taxonomical uncertain aneuploid strains (CBS 4837, CBS 4838, OUT7136) collectively referred as *Zygosaccharomyces rouxii* complex (Solieri et al., 2013b). Besides this, a wider pool comprising of 78 *Zygosaccharomyces* strains, obtained from diverse ecological niche, were genotyped using different molecular techniques as discussed in **section 1.6**. Six strains out of 78 strains analyzed showed rDNA heterogeneity based on polymorphism in their 5.8S ITS gene or 26S D1-D2 region. The polymorphic copies were cloned and subsequently sequenced. The presence of polymorphic copies of 5.8S ITS gene and 26S D1-D2 region was confirmed by chemiluminiscent detection by Southern blotting. The polymorphism suggested relaxation of concerted evolution of rDNA regions these strains. Concerted evolution describes the mechanism by which repeat families such as rDNA repeats within genomes are often maintained with similar sequences. The repeats in an array tend to evolve “in concert” with the same sequence via continual turnover of repeats by recombination. Evidence of relaxation of concerted evolution has been seen in other species of genus *Metschnikowia*, where the species has been found to share a pool of diverse rRNA genes differing in regions that determine hairpin-loop structures and is supposed to be an outcome of reticulate evolution (Sipiczki et al., 2013).

1.1.1. Systematic and phylogenetic circumscription of *Zygosaccharomyces*

The assignment of yeast species to genera and families has long been based primarily on morphology of vegetative cells and sexual states, and secondarily on physiological responses on the ca. 60-80 fermentation and growth tests commonly used in yeast systematic (Kurtzman et al., 2011). Conventional physiological and biochemical test intended to identify, characterize and differentiate of species of the genus *Zygosaccharomyces* posed considerable problems owing to relatively shallow nature of the genus in evolutionary terms (James et al. 1994). These analyses have been proved inadequate as they do not provide data sufficient for accurate placement of a species within a genus or sometimes may lead to misidentification or misplacement of species.

Recent progress in molecular biology has led to the development of new techniques for yeast identification based on similarity or dissimilarity of DNA, RNA or proteins. In particular, comparative analysis of partial 26S rRNA gene (Kurtzman and Robnett, 1998) and 18S rRNA gene sequencing (James et al., 1994; 1996) for phylogenetic based studies have gained considerable attention in past decade. Exploitation of DNA/protein sequences for systematic and phylogenetic studies have gained importance primarily for two reasons: firstly because they resolve species into well circumscribed phylogenetic genera; and secondly these analyses define species that couldn't be defined merely by analyzing morphology features and/or recording physiological growth responses. From single gene analyses, such as the D1-D2 phylogenetic trees presented by Kurtzman and Robnett (1998), it became apparent that many of the ascomycetous yeast genera are not well circumscribed and actual boundaries are often not well-defined. Comparative sequence analysis showed that the genus is phylogenetically very heterogeneous. This phylogenetic heterogeneity resulted in wrong identification and positioning of the species of the genus *Zygosaccharomyces* within the genus and in other closely related genera.

Sequence analysis of the small-subunit rRNA internal transcribed spacer region revealed that species of this genus are phylogenetically intermixed with species of other genera such as *Kluyveromyces, Saccharomyces* and *Torulaspora* (James et al., 1994, 1996; Cai et al., 1996). Since, multiple genes based phylogenetic reconstruction often receives far greater bootstrap support than single-gene do, Kurtzman and Robnett (2003) analyzed relationships among species of the '*Saccharomyces* complex' based on sequences of 18S, ITS, 5.8S and 26S rDNAs, translation elongation factor 1-K (*EF1-K*), mitochondrial small-subunit rDNA and *COX2*. These studies demonstrated the polyphyletic nature of this genus. "*Saccharomyces* complex" includes *Saccharomyces*, *Kluyveromyces*, *Tetrapisispora*, *Torulaspora*, and *Zygosaccharomyces* together with the neighboring genera *Eremothecium*, *Hanseniaspora*, and *Saccharomycodes* (Kurtzman and Robnett, 2003). In this multigene study, approximately 80 species were compared from the combined signal of seven genes. The analysis gave 14 phylogenetically defined clades, most of which had strong bootstrap support. From this study, the major genera *Saccharomyces*, *Kluyveromyces* and *Zygosaccharomyces* were shown to be polyphyletic, leading to reclassification of certain species in the new genera *Naumovia*, *Nakaseomyces*, *Vanderwaltozyma*, *Zygotorulaspora* and *Lachancea* expansion of the earlier described genus *Kazachstania* (Kurtzman, 2003).

On the basis of multiple-gene sequence analysis, the eleven *Zygosaccharomyces* species were resolved and ascribed to four phylogenetically circumscribed genera, viz. *Zygosaccharomyces* (*Z. rouxii, Z. mellis, Z. bailii, Z. bisporus, Z. kombuchaensis* and *Z. lentus*) (as described in previous yeast taxonomy; Kurtzman and Fell, 1998), *Zygotorulaspora* (*Z. mrakii* and *Z. florentinus*) and *Lachancea* (*Z. cidri* and *Z. fermentati*), *Torulospora* (*Z. microellipsoides*) (Kurtzman and Robnett, 2003; Kurtzman, 2003). Multigene based phylogenetic reconstruction assigned *Zygosaccharomyces* a well-

supported phyletic group of species, designated as *Zygosaccharomyces sensu strict*, that have overall phenotypic similarities (Kurtzman and Robnett, 2003; Steels et al., 2002).

In the current edition of yeast taxonomy, genus *Zygosaccharomyces* comprises of six species viz. *Z. bailii, Z. bisporus*, *Z. kombuchaensis*, *Z. lentus*, *Z. mellis* and *Z. rouxii* (James and Stratford, 2011). Other recently identified *Zygosaccharomyces* species includes *Z. machadoi* (Rosa and Lachance, 2005), *Z. gambellarensis* (Torriani et al., 2011), *Z. siamensis* (Saksinchai et al., 2012), *Z. sapae* (Solieri et al., 2013a), *Z. parabailii* and *Z. pseudobailii* (Suh et al., 2013).

1.1.2. Genomic exploration

Z. rouxii and *Z. bailii* are the most common food species contributing to food-grade fermentation as well as spoilage (Deak and Beuchat, 1996). Genome sequence of these food yeasts may generate valuable knowledge that could be used for metabolic engineering, improving cell factories and development of novel spoilage preservation methodologies. However, from genomic point of view the *Zygosaccharomyces* genus is poorly investigated, as only the *Z. rouxii* genome sequence was available until recently. For this reason, the exploitation of species of this genus in the context of food quality improvement, biotechnological and pharmaceutical applications was limited. Recently, sequencing efforts has also made available the genome of *Z. bailii* strain CLIB 213^T (Galeote et al., 2013).

Zygosaccharomyces rouxii is a haploid yeast. The genome of the sequenced strain *Z. rouxii* CBS 732^T consists of seven chromosomes ranging from 1.0Mb to 2.75Mb with a total size of 12.8Mb according to pulsed field gel electrophoresis determination (Solieri et al., 2008a). However, the genome size of CBS 732^T was estimated 9.76Mb as a result of final assembly of genome project

(Souciet et al., 2009). There are approximately 4998 protein encoding genes and 272 tRNA genes, the rDNA locus, an endogenous pSR1 plasmid and 15 mitochondrial genes in *Z. rouxii* (de Montigny et al., 2000).

Z. bailii is a diploid genome content. A total of 5,084 putative protein-coding genes, including 217 pseudogenes, were predicted using Amadea BioPack (ISoft, France) (Galeote et al., 2013). BLASTp comparison revealed that 98.2% of these genes have a homolog in *Z. rouxii*, only 8 are species-specific genes, and 5 were acquired by inter kingdom lateral transfer. A total of 162 tRNAs were identified using tRNAscan-SE v1.3.1.

Besides haploid *Z. rouxii* type strain CBS 732^T, the genus also comprises of diploid, allodiploid and aneuploid strains and species. Diploid *Z. sapae* strains $ABT301^T$ and ABT601; and allodiploid ATCC 42981 respectively have genome sizes 28.1, 39.0, and 21.9Mb (Solieri et al., 2008a). Aneuploid strains OUT7136, CBS 4837 and CBS 4838 have estimated genome sizes of 19.57–21.94Mb (Solieri et al., 2013b), significantly higher than the estimated 9.8-Mb genome of CBS 732^T (Souciet et al., 2009).

While *Z. bailii* is frequently encountered in wines, it has been shown to be the donor of a 17-kb DNA region to *Saccharomyces cerevisiae*. This transferred region was first discovered in the genome of the commercial wine yeast strain *S. cerevisiae* EC1118 and thereafter was widely found in other *S. cerevisiae* wine yeast genomes, in multiple copies and showing different structures and organization (Galeote et al., 2011; Novo et al., 2009). Given the phylogenetic distance between these two species and the large amount of DNA transferred, such eukaryote-to-eukaryote gene transfer is remarkable.

In addition, *Z. rouxii* harbors a natural, non-mitochondrial, 6.2kb replicative plasmid pSR1 containing 3 ORFs (Araki et al., 1985). The plasmid shares no homology with the *S. cerevisiae* 2µ plasmid (Araki et al., 1985). Similar plasmid has been also been reported in *Z. bailii* and *Z. bisporus* (Toh et al., 1982; Toh and Utatsu, 1985). Besides this, centromeric and episomal *Z. rouxii/Escherichia coli* shuttle plasmids with different marker genes (*ScURA3, ZrLEU2, ZrADE2*) that represents homology-driven gene deletion systems make genetic manipulations possible in these species (Pribylova et al., 2007a, 2007b).

1.2. Mechanisms implicated in genome evolution

The evolutionary history of yeast species has been highly impacted by whole genome duplication, polyploidization, and postulated hybridization events. These events resulted in species diversification, altered genome complexities, mosaicism, and acquisition of novel genetic traits (Gordan and Wolfe, 2008). With the availability of genome sequences from an increasing number of yeasts, a marked advancement in understanding evolutionary history and mechanisms has become possible (Dujon, 2004). Analysis of single genomes has limited advantages as it could merely generate hypothesis such as the occurrence of whole-genome duplication or postulated hybridization; however, comparative genomics analysis of close and more distant species represents a powerful tool that demonstrates evolutionary processes and selection pressures acting on genome (Souciet et al., 2009).

1.2.1. Mechanisms implicated in genes duplication

Yeasts genome has shown traces of numerous genomic duplications. These genomic duplications involve individual gene (gene duplication), genomic segments (segmental duplications) or

entire genome (whole genome duplication) and have been proposed as an advantageous path to evolutionary innovation, because paralogous genes can provide raw genetic material for the emergence of new functions over time through mutation and natural selection (Ohno, 1970). The evolutionary consequences of these duplications in creating novel gene functions have attracted considerable attention (Conant and Wolfe, 2008; Innan and Kondrashov, 2010).

In every genome a vast majority of genes are selectively constrained. This has been clearly underlined by comparative genomic studies that demonstrate that coding sequences diverge at slower rates than non-coding regions, largely due to a deficit of mutations at positions where a base change would cause an amino-acid change. Duplication creates a second copy of the gene, which is often free from selective pressure, that is, mutations of it have no deleterious effects to its host organism. Thus, gene duplication provides opportunities to explore this forbidden evolutionary space by generating duplicates of a gene that can move and diverge more freely. Susumu Ohno was the first to comprehensively elucidate the potential of gene duplication, in his book *Evolution by Gene Duplication*, published more than 30 years ago (Ohno, 1970).

Theory suggests three alternative outcomes in the evolution of duplicate genes as depicted in **Figure 1.2.**: (i) one copy may simply become silenced by degenerative mutations (nonfunctionalization); (ii) one copy may acquire a novel, beneficial function and become preserved by natural selection, with the other copy retaining the original function (neofunctionalization); or (iii) both copies may become partially compromised by mutation accumulation to the point at which their total capacity is reduced to the level of the single-copy ancestral gene (subfunctionalization).

Whole genome duplication (WGD)

The ancestral genome of the *Saccharomyces sensu stricto* complex was subjected to whole-genome duplication (WGD) (Wolfe and Shields, 1997; Dietrich et al., 2004; Kellis et al., 2004). WGD is often expected to have a large impact on the evolution of lineages in which it has occurred. It is usually followed by massive gene deletions (returning to singleton state), gene retention and divergence leading to adaptive innovations (Kellis et al., 2004). Massive gene deletions are prone to completely reshuffle genome maps, break down syntenic relationships by forming two imperfect synteny blocks from the original one. However, this does not appear to drive speciation (Dujon, 2010). In *S. cerevisiae*, only ~550 duplicated pairs (ohnologues) still remain (Byrne et al., 2005), and similar or lower figures are observed for other yeasts that emerged from the same duplication event, such as *C. glabrata* (Dujon et al., 2004), *Naumovozyma castellii* (Cliften et al., 2006) and *Vanderwaltozyma polymorpha* (Scannel et al., 2007). The molecular mechanism by which so many individual genes disappear after duplication remains one of the problems to solve (Martin et al., 2007). The remaining pairs of ohnologues shed light on post duplication evolutionary mechanisms. When compared with single copy homologues from protoploid species of Saccharomycetaceae, most ohnologous pairs in duplicated species exhibit strong asymmetry in their evolution rate, which suggests neofunctionalization (Kim and Yi, 2006; Byrne and Wolfe, 2007). The serine/threonine protein kinases encoded by the *NPR1* and *PRR2* ohnologues of *S. cerevisiae* illustrate this post duplication event where the former is slow evolving having role in stabilization of plasma membrane amino acid transporters and the latter is the fast evolving with involvement in the pheromone induced signaling pathway. Interestingly, the fast evolving copy is never an essential gene and its function is generally less well understood than that of the slow evolving copy (Byrne and Wolfe, 2007). The expression of the two gene copies also frequently differs, consistent with the divergence of their *cis* regulatory

elements (Papp et al., 2002; Gu et al., 2005). The *CYC1* gene of *S. cerevisiae,* which encodes isoform 1 of cytochrome *c*, is preferentially expressed in the presence of oxygen, whereas its ohnologue *CYC7*, which encodes isoform 2, is preferentially expressed under hypoxic conditions. Other cases of subfunctionalization involve subcellular localization, often separating mitochondrial from nuclear or cytoplasmic functions. About one third of proteins encoded by pairs of ohnologues localize to distinct subcellular compartments (Marques et al., 2008). Although rare, whole genome duplications have also played an impressive role in the evolution of lifestyle of yeast species. The whole genome duplication in the *Saccharomyces* lineage is thought to have shaped the fermentative lifestyle of these yeasts (Wolfe and Shields, 1997; Piškur, 2001).

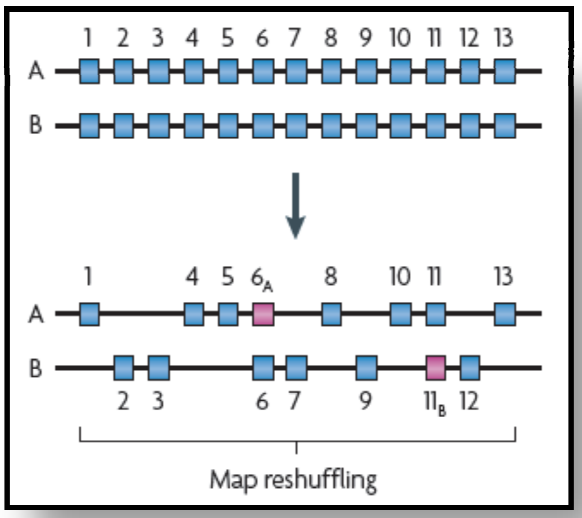


Figure 1.2. Genome duplication followed by massive gene deletions and divergence. "A" and **"B"** represents asymmetrical divergence of ohnologues pairs leading to neo- or sub-functionalization (Source: Dujon 2010).

Single-gene duplications

Differently from the paralogous gene pairs originate from complete genome duplication, these genomic duplications represent dispersed paralogues and tandem gene arrays also exist. It has been observed in relation to the 'protoploid' yeasts of Saccharomycetaceae family that the latter two phenomena occurred so frequently that one third of their protein coding genes are members of paralogous families even in the absence of an ancestral whole genome duplication (Souciet et al., 2009). The numbers of paralogues are even higher in *Yarrowia lipolytica* and in species of the 'CTG group' (so called because its constituent species translate CTG as Serine instead of Leucine).

In the fruitfly, the nematode worm and the fission yeast genomes, most paralogous genes are dispersed. Despite the dispersion of most paralogues in genomes, no molecular mechanisms operating at the chromosomal level have been identified so far that would create single gene duplicates at ectopic locations. The only known examples of such cases in yeasts and other eukaryotes proceed through RNA intermediates, the duplicated gene copy representing a retrogene reinserted into a new chromosomal location. This was directly demonstrated experimentally in *S. cerevisiae*, in which Ty retrotransposons can duplicate part of a gene and integrate the copy into a new chromosomal location (Schacherer et al., 2004). The proposed mechanism involves template switching of the polymerase during the reverse transcription phase, such that a cDNA copy of a cellular mRNA is made and integrated into chromosomes. The general paucity of active retrotransposons in yeast genomes (Neuvéglise et al., 2002) suggests, however, that this mechanism plays only a limited part in their evolution except, possibly, during transient bursts of transposon activity. The scarcity of spliceosomal introns in many yeast species (Bon et al., 2003; Stajich et al., 2007) also suggests that exon shuffling will be limited during yeast evolution.

Segmental duplications

In contrast, duplicated genes in the yeast *S. cerevisiae* (Wolfe and Shields, 1997), the plant Arabidopsis thaliana (Vision et al., 2000), and in humans, Homo sapiens (McLysaght et al., 2002), often occur in large segmental duplications in which members of homologous gene pairs are located in the same order along the two distinct segments and are sporadically interspersed with unique genes. Segmental duplications constitute a major signature of various yeast genomes and are distributed nonrandomly in genomes with clustering near subtelomeric regions (Souciet et al., 2006; Fairhead and Dujon, 2006). Similar clustering in subtelomeric regions and pericentromeric regions also exists in

higher eukaryotes such as Primates where segmental duplications are known to constitute ~5% of the total genome size (Marques-Bonet et al., 2009). Blocks of duplicated sequence totaling about 50 kb retaining a conserved gene order has been found at the sub-telomeric regions of chromosomes I and II in *S. pombe*. Twenty-four genes (in groups of two or four) are 100% identical at the DNA level, and twenty of these are localized in sub-telomeric regions, suggesting frequent exchange of genetic information at these positions. Most of these genes code for proteins belonging to families specific to fission yeast and are predicted to be cell-surface proteins (Wood et al., 2002). Interestingly, in *S. cerevisiae* seven of the sixteen genes (in groups of two, three or four) that are 100% identical at the DNA level are also located in sub-telomeric regions. These gene products include members of the budding-yeast-specific *PAU* and *COS* families, which are also predicted to be cell-surface proteins (Goffeau et al., 1996).

Yeasts, however, have been useful for elucidating the molecular mechanisms involved in the formation of segmental duplications and, given the structural similarities of segmental duplications between yeast and multicellular eukaryotes, it is likely that their varying abundance between genomes reflects different dynamic equilibriums rather than fundamental mechanistic differences (Koszul et al., 2009). In experimental cultures of *S. cerevisiae*, duplications of large chromosomal segments (tens to hundreds of kilobases) containing many genes occur spontaneously at high frequency (~10^{-7} per mitosis) (Koszul et al., 2004; Schacherer et al., 2007; Gresham et al., 2008). Four types of chromosomal structures are formed with different degrees of stability during subsequent generations (Koszul et al., 2006). Intra and interchromosomal duplications are also observed in mammalian genomes. In addition, extrachromosomal copies of a specific chromosome segment were reported (Libuda et al., 2006) and similar structures have been observed in other experiments. This suggests that, after reintegration at ectopic chromosomal locations, extrachromosomal copies may contribute to

genome reshuffling. It is now clear from yeast experiments that large segmental duplications result from untimely DNA synthesis events requiring polymeraseδ(Polδ) (Payen et al., 2008). Dispersed repeated elements in genomes, such as remnants of Ty elements (class i retrotransposons), anchor the duplications by a *RAD52* dependent break induced replication (BiR) mechanism, often resulting in subtelomeric duplications or interchromosomal translocations. But duplications also occur in the absence of such elements, with no preference for genomic locations, as a result of a *RAD52* independent microhomology/microsatellite induced replication (MMiR) mechanism. Owing to the greater technical limitations involved, mechanisms are less precisely described for mammalian genomes, but the presence of sequence microhomology and topoisomerase binding sites at or near junctions also suggests a replication based mechanism (Marques-Bonet et al., 2009). The influence of the segmental duplication mechanism in natural evolution of yeast genomes remains to be better quantified, but this mechanism is prone to leave behind duplicated gene copies and chromosomal rearrangements in yeast cultures that may contribute to the numerous dispersed paralogues observed.

Expansions of tandem gene arrays

Tandem gene arrays (TGAs) can be as a structure of contiguous paralogous gene copies, which are either functional or degenerated (gene relics), allowing for the presence of one heterologous gene inserted between the two homologous copies. Clusters of identical or similar protein coding genes, such as ribosomal DNA genes (**Figure 1.3**), exist in various eukaryotic genomes, but their role in evolution remains largely speculative (Dujon, 2010). In yeasts, tandem gene arrays are generally not conserved except in specific cases, such as B type cyclin genes (Despons et al., 2010; Dujon, 2010). Several examples indicate the role of tandem gene array expansions in rapid adaptive evolution. In *S. cerevisiae*, tandem expansion of the *CUP1* locus occurs when cells are grown under selective pressure

for copper resistance (Fogel and Welch, 1982). Because this cluster is made of identical gene copies (Jhonston et al., 1994), its expansion (or shrinkage) can be accounted for by unequal homologous recombination events. However, in other cases tandem arrays are generally made of sequence diverged gene copies, which is more consistent with functional diversification than with the simple need for copy number increases, although the degree of possible concerted evolution in such tandems remains to be analyzed. In the pathogenic yeast *Candida glabrata*, two large gene clusters are observed that are unique to this species (Dujon et al., 2004). One of them corresponds to the expansion of six additional *YPS* genes (compared with two *YPS* genes in *S. cerevisiae*), which encode extracellular glycosylphosphatidylinositol (GPi) linked aspartyl proteases that are required for virulence (Kaur et al., 2007). Expansion of this cluster corresponds with the roles of these enzymes in processing a GPi linked cell wall adhesin that is necessary for the adherence of *C. glabrata* to mammalian cells. The other cluster is an expansion of the unique *S. cerevisiae MNT3* gene into eight copies in *C. glabrata*. This gene encodes an α1, 3 mannosyltransferase involved in cell wall biogenesis. Again, the sequence divergence between the tandem copies is consistent with functional diversification of the encoded proteins, and the cluster varies in size among clinical isolates of *C. glabrata* (Müller et al., 2009). Another dynamic large tandem gene array called *DUP240*, which is specific to *S. cerevisiae* and closely related species, encodes proteins that facilitate membrane trafficking (Despons et al., 2006). Numerous other species specific arrays are found in other yeasts, but their functional roles are unclear.

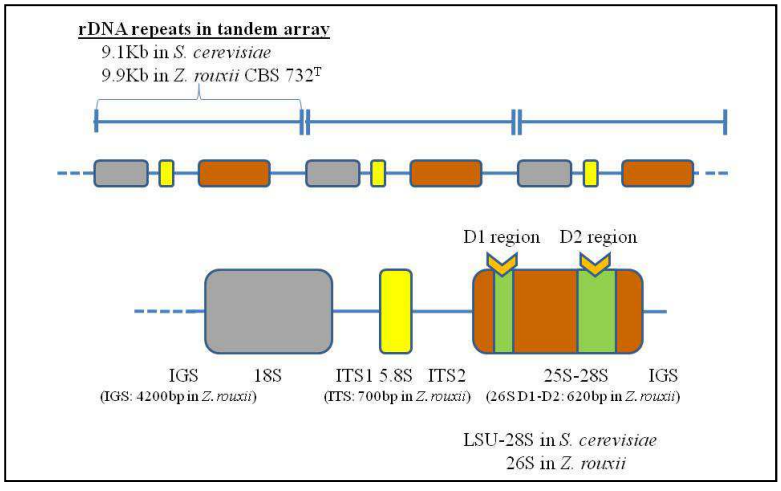


Figure 1.3. Schematic diagram showing yeast ribosomal RNA gene present in tandem array.

1.2.2. Postulated hybridization and polyploidization

Polyploidization and hybridization are closely interrelated processes: allopolyploidy necessarily arises through interspecific hybridization associated with genome doubling. In addition, although autopolyploidy may arise without intraspecific hybridization (i.e. only through genome doubling), many autopolyploid species display higher heterozygosity levels than their diploid counterparts as in yeasts (Albertin et al., 2009), postulating a hybrid origin. Evolution through hybridization, with or without genome doubling, is referred as reticulate evolution or reticulation (Chester et al., 2010) and represents the first step towards polyploidy.

Heterospecific hybridization is defined as the process of the fusion of two gametes differing in their genetic constitution resulting in formation of an alloploid genome consists of copies of the genomes of the postulated hybridizing species. While not very common, a few yeast species have been recognized as being the result of postulated hybridization. Successful experimental endeavor for obtaining postulated interspecific hybrids, by mass mating between different *Saccharomyces* species, were taken several years ago (Marinoni et al., 1999). Recent genomic studies reveal that strains of the lager beer yeast *Saccharomyces pastorianus* are distinct postulated hybrids between *S. cerevisiae* and *Saccharomyces bayanus* (Dunn and Sherlock, 2008; Nakao et al., 2009). Several other beer strains are known to be the postulated hybrids between *S. cerevisiae* and *S. kudriavzevii* (González et al., 2008; Belloch et al., 2009). There are speculations that these postulated hybridization events are accelerated by the stressful conditions, similar to that imposed during alcoholic fermentations (Replansky et al., 2008). Other postulated *Saccharomyces* interspecific hybrids have been detected in different fermentation processes such as those involving wine and cider (Groth et al., 1999; Masneuf et al., 1998; Naumova et al., 2005a). González et al. (2008) reported a considerable fraction (~25%) of the

brewing strains to be *S. cerevisiae* × *S. kudriavzevii* postulated hybrids that contributing to the complexity of *Saccharomyces* diversity in brewing environments. This raises the possibility about the prevalence of postulated hybrids of *Saccharomyces* and of other genera in other fermentative processes representing osmotic stresses such as production of TBV, soy-sauce, miso etc. Pulvirenti et al. (2000) hypothesized that horizontal transfer of genetic material between yeast species and the likelihood of postulated hybrid formation in natural settings, and possibly shed light on the high biodiversity of yeast observed in nature. The formation of postulated interspecific hybrids is not limited to *Saccharomyces sensu strict* yeasts as postulated natural hybrids have been reported in other yeast and fungal genera. The asexual diploid pathogenic yeasts *Candida albicans* and *Candida dubliniensis* form postulated tetraploid hybrids (Pujal et al., 2004), and postulated natural hybrids have also been reported between the Basidiomycota yeasts *Cryptococcus neoformans* and *Cryptococcus gattii* (Bovers et al., 2006). Molecular evidences of postulated natural hybrids in genus *Zygosaccharomyces* has been presented (James et al., 2005) with subsequent identification of two subgenomes (p-subgenome and t-subgenome) in a wild isolate of allodiploid *Zygosaccharomyces rouxii* strain ATCC 42981 (Gordan and Wolfe, 2008). Presence of two subgenomes has been reported in *Pichia sorbitophila* too (Genolevures Consortium, personal communication). Despite the frequent occurrence of postulated hybrids, the role of postulated heterospecific hybridization in the evolution of yeasts remains unclear. In postulated *Saccharomyces sensu stricto* hybrids, the two parental genomes often undergo nonreciprocal exchanges accompanied by extensive modification including loss of genes, chromosomal segments or even complete chromosomes, resulting in production of various alloaneuploids or recombinants and mosaic genomes from which novel lineage might emerge (Greig et al., 2002; Usher et al., 2009).

1.2.3. Aneuploidy and genome instability

Correct transfer of genetic information to daughter cells is essential for successful propagation of any organism. Three processes are involved in maintenance and propagation of genetic information: DNA replication, DNA damage repair and chromosome segregation. Error in any of these processes might result in cell death, or, in another scenario, in survival of cells with altered genetic information. This might be reflected either by single nucleotide changes as well as small insertions and deletions; or it might lead to larger alterations in the structure and number of chromosomes, together called aneuploidy.

What is aneuploidy?

A change in chromosome number that is not the exact multiple of the haploid karyotype is known as aneuploidy. This condition interferes with growth and development of an organism and is a common characteristic of solid tumors. Aneuploidy describes any karyotype that differs from a normal chromosome set (called euploidy) and its multiples (called polyploidy). Aneuploidy can occur either by chromosome gains and losses due to chromosome segregation errors, a so called "whole chromosomal" aneuploidy, or due to rearrangements of chromosomal parts, often accompanied by their deletion and amplification, that is referred to as a "structural" or "segmental" aneuploidy. Frequently, a combination of both structural and numerical chromosomal changes can be found, in particular in cancer cells (composite aneuploidy). Aneuploidy often reflects chromosomal instability (CIN), which is an ongoing defect in faithful transmission of chromosomes. Chromosomally instable cells accumulate new karyotype alterations as they proliferate and they are always aneuploid. In

contrast, not every aneuploidy is linked to CIN, some cells can remain in a stable aneuploid status for multiple generations.

The first aneuploid *Saccharomyces cerevisiae* strains were generated by Mortimer and Hawthorne in 1966 for mapping purposes (Mortimer and Hawthrone, 1966). In 1970, Parry and Cox (1970) generated a series of disomic strains by sporulating triploid yeast strains. They recovered a large number of offspring and identified strains disomic for at least five chromosomes, leading them to suggest that aneuploidy is well tolerated in yeast. Subsequent studies employed disomic and monosomic yeast strains for the purpose of measuring chromosome loss rates (Hartwell et al., 1982), but a systematic characterization of all disomic strains was not conducted until recently. Torres et al. (2008) analyzed 13 of 16 possible $1n + 1$ yeast strains and found them to be impaired in proliferation and sensitive to a number of conditions interfering with protein synthesis and turnover (Torres et al., 2007). The genome size of *Schizosaccharomyces pombe* is similar to that of *S. cerevisiae* but is contained on only three chromosomes. Yanagida and coworkers showed in 1985 that only strains disomic for the smallest chromosome (chromosome III) are viable but severely growth retarded (Niwa and Yanagida, 1985). More recent studies on whole chromosomal aneuploidy as well as segmental aneuploidy indicate that aneuploidy in *S. pombe* hampers cell proliferation (Niwa et al., 2006).

Cause of aneuploidy

The cell division cycle is a highly controlled process that generates two daughter cells of identical genetic makeup. Surveillance mechanisms known as checkpoints ensure that this process occurs with high fidelity. These cell cycle checkpoints delay chromosome segregation until cell has replicated its DNA and sister chromatids have properly aligned on the mitotic plate. These checkpoints ensure fidelity of chromosome segregation. However, these safeguards occasionally fail, resulting in

daughter cells that have gained or lost portions of the missegregated genetic material, i.e., aneuploids. However, despite these surveillance mechanisms, chromosome missegregation occurs once every 5×10^5 cell divisions in yeast (Hartwell et al., 1982) and on the order of once every 10^4 to 10^5 divisions in mammalian cells (Rosenstraus and Chasin, 1978), producing a condition known as aneuploidy.

Sheltzer et al. (2011) measured the segregation fidelity of a yeast artificial chromosome (YAC) containing human DNA and found that the rate of chromosome missegregation was increased in 9 out of 13 disomic strains relative to a euploid control. The increase ranged from 1.7-fold to 3.3-fold, comparable to the fold increase observed in strains lacking the kinetochore components Chl4 or Mcm21. Consistent with chromosome segregation defects, 8 out of 13 disomic strains displayed impaired proliferation on plates containing the microtubule poison benomyl, including a majority of the strains that had increased rates of YAC loss.

Chromosome missegregation can result from defects in chromosome attachment to the mitotic spindle or from problems in DNA replication or repair. Besides this, microtubules also play indispensable role during cell division that include proper segregation of chromosomes and positioning of the nucleus during mitosis. Mutants with defective tubulin folding display specific effect during mitosis such as abnormal nuclear positioning (Lacefield et al., 2006). Defects in any of these processes delay mitosis by stabilizing the anaphase inhibitor Pds1 (securin) (Musacchio et al., 2007). Five out of five disomes (disomes V, VIII, XI, XV, and XVI) exhibited delayed degradation of Pds1 relative to wild type after release from a pheromone-induced G_1 arrest (Sheltzer et al., 2011). Defective chromosome bi-orientation delays anaphase through the mitotic checkpoint component Mad2 (Musacchio et al., 2007). Deletion of *MAD2* had no effect on Pds1 persistence in four disomes, but eliminated this persistence in disome V cells (Sheltzer et al., 2011). Disome V also delayed Pds1 degradation after release from a mitotic arrest induced by the microtubule poison nocodazole, which

demonstrated that this strain exhibits a bi-orientation defect. Disome XVI, which displayed Mad2-independent stabilization of Pds1, recovered from nocodazole with wild-type kinetics (Sheltzer et al., 2011). Thus, Pds1 persistence results predominantly from Mad2-independent defects in genome replication and/or repair (Sheltzer et al., 2011). This study establishes that missegregation of a single chromosome is sufficient to induce the hallmarks of genomic instability, including whole-chromosome instability, mutagenesis, and sensitivity to genotoxic stress (Sheltzer et al., 2011). Genomic instability in the disomes is not correlated with the size of the extra chromosome or the delay in cell cycle progression.

In case of higher eukaryotes, chromosomal missegregation and induction of aneuploidy arise from non-disjunction during cell division as a result of merotelic attachments, supernumerary centrosomes, defects in spindle assembly checkpoint (SAC), and premature loss of chromatid cohesion (Gordon et al. 2012). Perhaps the best-characterized cause of chromosomal instability is weakening of the Spindle Assembly Checkpoint (SAC) (Weaver and Cleveland, 2006; Thompson and Compton, 2010). Most of the cells that undergo chromosomal missegregation die in a p53 dependent cell cycle arrest. Besides this, role of additional signaling circuits such as p38 (MAPK14), Extracellular signal-regulated pathways (ERKs), c-Jun N-terminal Kinase (JNK), and Phosphoinositide 3-kinase PI3 pathways have also been implicated in various ways in context to chromosomal instability (CIN). However, some aneuploid cells show continual reassortment of their genome.

Consequences of aneuploidy

Studies examining the effects of aneuploidy on cell proliferation in *Schizosaccharomyces pombe* (Niwa et al., 2006), *S. cerevisiae* (Torres et al., 2007), *C. elegans* (Hodgkin, 2005), *Drosophila* (Lindsley et al., 1972), mouse (Williams et al., 2008) and the effects of trisomy on cell

proliferation in humans (Hassold et al., 1996) suggest that aneuploidy can also interfere with cell proliferation. To address how aneuploidy affects the proliferation and physiology of normal cells, we generated a set of yeast strains in which each strain bears an extra copy of one or more of almost all of the yeast chromosomes. Their characterization represents a comprehensive analysis of the effects of aneuploidy on cellular physiology. We found that in addition to chromosome-specific phenotypes, aneuploid strains share a number of traits, pointing toward the existence of a general cellular response to aneuploidy.

Genomic instability provides a growth advantage during the experimental evolution of microorganisms and drives the development of tumors. Although aneuploidy confers severe disadvantages to cells by stressing protein homeostasis and altering metabolism (Torres et al., 2007; Torres et al., 2010), our results suggest it may also benefit cells under selective pressure by increasing the likelihood that growth-promoting genetic alterations will develop. The mutagenic effects of aneuploidy that we report here may represent one mechanism by which changes in karyotype influence cancer development and evolution. It has been suggested that aneuploidy increases genomic instability, which could accelerate the acquisition of growth-promoting genetic alterations (Torres et al., 2008; Matzke et al., 2003). However, whereas aneuploidy is a result of genomic instability, there is at present limited evidence as to whether genomic instability can be a consequence of aneuploidy itself. This question has been addressed by assaying chromosome segregation fidelity in 13 haploid strains of *Saccharomyces cerevisiae* that carry additional copies of single yeast chromosomes (Torres et al., 2007). These aneuploid strains (henceforth disomes) display impaired proliferation and sensitivity to conditions that interfere with protein homeostasis (Torres et al., 2007; Torres et al., 2010).

To characterize the effects of aneuploidy on gene expression, a comparative genome-wide gene expression of aneuploid strains relative to the wild-type strain with the use of DNA microarrays can be

performed. An approximate doubling of gene expression was observed along the entire length of the disomic chromosomes, indicating that most of the genes are expressed proportionally to the number of DNA copies in the cell (Sheltzer et al., 2011).

1.2.4. Other mechanisms of genome evolution

Loss of heterozygosity

Interestingly, in addition to mating type switching, yeasts have evolved another mechanism that reduces heterozygosity in diploid cells or postulated hybrids. The asexual diploid yeast *C. albicans* is an interesting model in which homologous chromosome pairs show a mosaic of heterozygous and homozygous regions (Butler et al., 2009). In this case, sequence divergence between alleles is erased by a general chromosomal mechanism, LOH (**Figure 1.4**). By comparing commensal isolates, most LOH events were found to encompass entire chromosomes or large regions extending to telomeres, as predicted by a BiR mechanism (Diogo et al., 2009). The related yeast *C. dubliniensis* also has three homozygous chromosomes, and distinct chromosomal bands for other chromosomes correspond to the rest of the *C. dubliniensis* genome (Jackson et al., 2009). To what extent LOH applies to other yeast evolutionary lineages is presently unclear, although genes involved in LOH have been identified in *S. cerevisiae* (Andersen et al., 2008). In *C. albicans*, long range LOH is increased (relative to *in vitro* cultivation) by passage though a mammalian host (Forche et al., 2009).

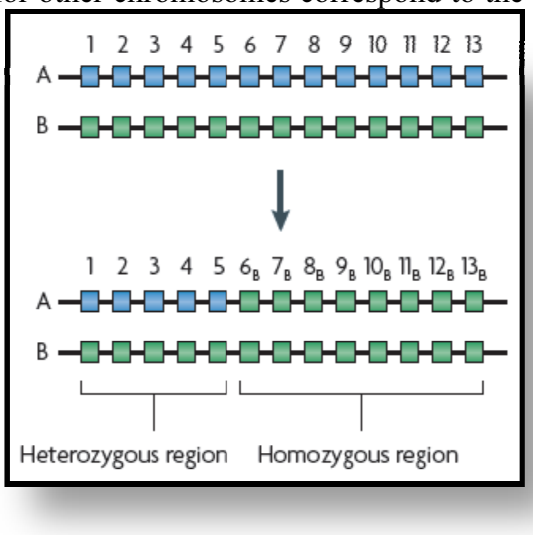


Figure 1.4. Mechanistic basis of loss of heterozygosity. Crosses between two non-isogenic strains (distinct subpopulations) or hybridization between different species that have conserved a substantial degree of synteny creates

diploids with heterozygous chromosomes. Subsequent loss of heterozygosity of large chromosomal segments often occurs (Source Dujon, 2010).

Introgression

The presence in yeast genome sequences of large chromosomal segments that are identical or nearly identical in sequence to other species is puzzling. The genomes of wine strains of *S. cerevisiae* contain DNA fragments from *S. paradoxus*, *S. kudriavzevii*, *Saccharomyces uvarum* and even the distantly related *Zygosaccharomyces bailii* (Liti et al., 2006; Naumova et al., 2005a; Doniget et al., 2008; Muller et al., 2009; Novo et al., 2009), suggesting that very recent introgressions have occurred (as the sequences are nearly identical to those in the donor genomes).

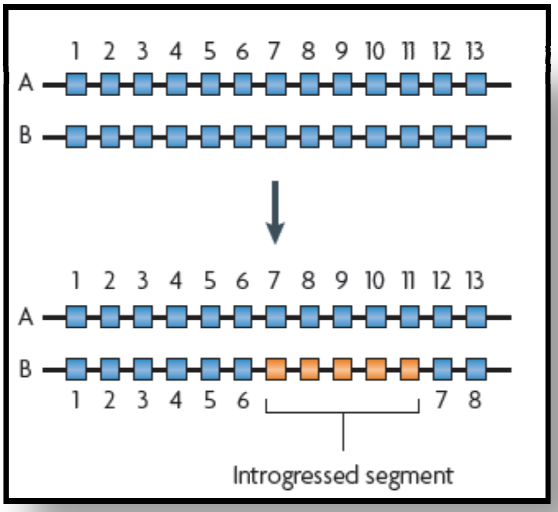


Figure 1.5. Mechanistic basis of introgression. Starting from a yeast with homozygous or heterozygous chromosomes, the acquisition of alien genes locally modifies the genomic map. The alien acquisition may concern a single gene (horizontal transfer, usually from bacteria) or a segment of chromosome from another yeast (introgression) (Source: Dujon, 2010).

Introgression was also reported between varieties of the Basidiomycota yeast *C. neoformans* (Kavanaugh et al., 2006), indicating that this is a general phenomenon in yeast genomes, although its importance for evolution has not yet been determined. The mechanism of introgression remains unknown. Classically in plants, introgression is obtained by hybridization followed by successive backcrossing. It is possible, although not likely, that the same process applies in yeasts. Alternatively, delayed karyogamy in postulated heterospecific hybrids may allow the transfer of chromosomal fragments from one nucleus to the other. In crosses involving *S. cerevisiae* mutants delayed in karyogamy, haploid segregants can inherit intranuclear autonomous elements of the other parent. The

formation of circular chromosomal fragments by segmental duplication mechanisms may provide the source of DNA material transferred to the other nucleus. The ecological proximity and selective pressures to adapt to high sugar, low nitrogen and high ethanol conditions during fermentations may facilitate the phenomenon, explaining the frequent introgressions observed in industrial *S. cerevisiae* strains.

Horizontal gene transfer from bacterial origin

The acquisition of genes from bacteria has long been regarded as a marginal phenomenon in yeasts. However, various cases have been reported, and some of these have contributed to important functional innovations or reacquisitions of genes. A notable example is provided by the acquisition of a bacterial gene encoding a dihydroorotate dehydrogenase, possibly originating from a *Lactococcus*, in the ancestor of the Saccharomycetaceae family followed by its vertical transmission to many extant members of this family to form the *URA1* gene (Gojdovic et al., 2004; Hall et al., 2005). This gene either cohabits with the ancestral eukaryotic enzyme encoded by the *URA9* gene (in protoploids) or has replaced it (in *Saccharomyces sensu stricto* species). Interestingly, the cytosolic bacterial enzyme allows the synthesis of uracil in the absence of oxygen, whereas the mitochondrial ancestral eukaryotic one does not, a property that could have allowed the emergence of facultative anaerobe yeasts. Several other examples of acquisition or reacquisition of metabolic functions in various yeast lineages have been reported (Hall et al., 2007; Wei et al., 2007; Woolfit et al., 2007; Fitzpatrick et al., 2008). Furthermore, a recent systematic analysis of inserted genes in conserved synteny blocks (Rolland et al., 2009) and a systematic computer analysis using phylogenomic criteria suggest that the phenomenon is not rare in yeasts (Marcet-Houben and Gabaldón, 2010). Similarly, *Z. bailii* has been shown to be the donor of a 17-kb DNA region to commercial wine yeast strain *S. cerevisiae* EC1118 and thereafter

was widely found in other *S. cerevisiae* wine yeast genomes, in multiple copies and showing different structures and organization (Galeote et al., 2011; Novo et al., 2009). Given the phylogenetic distance between these two species and the large amount of DNA transferred, such eukaryote-to-eukaryote gene transfer is remarkable. Although the mechanism of transfer has not been elucidated, it is probably facilitated by the fact that most yeast species are terrestrial saprobes living in close association with other organisms. Interestingly, transferred genes have a tendency to duplicate in their new yeast host (Rolland et al., 2009).

De novo gene formation and acquisition of other alien sequences

Every sequenced yeast genome has revealed a small but significant number of specific genes without clearcut homologues in other species, including the most closely related ones. These genes are generally of unknown function and unclear origin. If these genes have not been acquired by horizontal transfer, they may be examples of *de novo* gene formation. In complex eukaryotes, such genetic innovations often result from exonization of mobile sequences or exon shuffling. However, the dearth of spliceosomal introns in Saccharomycotina despite their intron rich ancestors (Stajick et al., 2007) and the inconstant presence of active class I transposons (Neuvéglise et al., 2002) predict that exonization and exon shuffling should have a more limited role in yeast evolution. Nevertheless, these two mechanisms may partly explain the fusions and fissions of protein domains that have been observed in yeast genomes (Durrens et al., 2008). The *de novo* formation of protein coding genes by mutational sequence changes is also possible in yeasts. In *S. cerevisiae*, a short protein coding gene, *BSC4*, that is involved in DNA repair during stationary phase seems to have emerged from a transcribed but noncoding sequence that is present in related species (Cai et al., 2008). Another example is provided by the *S. cerevisiae* specific gene *MDF1*, which seems to have emerged from

translation of the antisense transcript from the ancestral *ADF1* gene, which is conserved across yeast species (Li et al., 2010). The translation product of *MDF1* suppresses mating efficiency of *S. cerevisiae* by binding to the MATa2 protein, hence promoting vegetative growth. The integration of fragments of mitochondrial DNA (nuMTs) into nuclear chromosomes is also a strong mutagenic force that could, theoretically, create novel gene fusions; for example, 5' extension of a protein coding gene was recently reported in *Debaryomyces hansenii* (Sacerdot et al., 2008). The process of transfer of mitochondrial DNA to the nucleus is not entirely understood, except that it must involve the fragmentation of mitochondrial DNA before integration at double strand breaks of chromosomes, as judged from the frequent occurrence of mosaics in both natural and experimental cases (Ricchetti et al., 1999; Sacerdot et al., 2008). Similarly, the presence of nuclear sequences of plasmid and viral origin (nuPAVs) was recognized in about 40% of the sequenced Saccharomycotina genomes (Frank and Wolfe, 2009). Although most correspond to pseudogenes, some active protein coding genes were shown to originate from the non retroviral RNA viruses that are occasionally found in various yeast and fungal species (Frank and Wolfe, 2009; Taylor and Breunn, 2009).

1.3. Major evolutionary forces impacting genomic complexity and diversity

1.3.1. Life cycle and breeding system

Most microbes have complex life cycles with multiple modes of reproduction that differ in their effects on DNA sequence variation. Three types of life cycle patterns are seen among the yeasts: 1) Haplobiontic 2) Diplobiontic, and 3) Haplo-diplobiontic. Besides this, the yeast breeding system provides three alternatives that enable haploids to return to the diploid state that is necessary for meiosis: mating of unrelated haploids (amphimixis), mating between spores from the same tetrad

(intratetrad mating, automixis) and mother daughter mating upon mating type switching (haplo-selfing). Different breeding strategies in yeasts explain how specific breeding strategies impact the structure, composition and evolution of yeasts genome. The frequency of specific mating events affects the level of heterozygosity present in individuals and the genetic diversity of populations. Quantifying the different aspects of the life cycle; however, has remained a daunting task. Recently, population genomic, genetic, and experimental analyses have been used to estimate the relative frequencies of these different modes of reproduction in nature in order to allow accurate measurement of recombinational diversity.

Yeasts reproduce by both sexual and vegetative methods. The primary mode of reproduction is vegetative reproduction that takes place by budding and fission. Sexual reproduction takes place during unfavourable conditions. The ability to undergo meiosis allows the yeast cells to "reshuffle" their genes when conditions are poor, perhaps enabling them to find a combination more suitable for survival in their environment. Yeasts are homothallic or heterothallic. They lack definite sex organs. The somatic cells or ascospores function as copulating gametangia that fuse to form a diploid zygote cell (**Figure 1.6**). The diploid zygote cell will form four haploid spores encased in a protective outer layer called the ascus. Confinement within the ascus is thought to enforce mating between products of the same meiotic division, minimizing outcrossing in this stage of the life cycle. When conditions improve, the spores germinate and are constitutively ready to mate and return to the diploid state. They can mate with another spore from the same tetrad (inbreeding), or they can be released from the ascus and mate with a spore from another tetrad (out-crossing), which may or may not be from the same diploid clone.

If the haploid spores do not mate immediately, they are able to undergo mitoses, during which they repeatedly switch mating types, thus enabling matings between haploid clone mates (haplo-selfing or autodiploidization). This switch is possible because mating type is determined by a system of two cassettes, a and α, both present in the same individual, which alternately insert into the mating type locus (*MAT*) of chromosome III after being copied from their master loci, *HML* and *HMR*, near the two ends of the chromosome (Oliver et al., 1992). In the *Saccharomyces* sensu stricto clade, which includes the well-studied yeast *Saccharomyces cerevisiae*, the number and rough organization of chromosomes are preserved within isolates of the same species. Their genomes are almost collinear and consist of 16 chromosomes (Fischer et al., 2006). This relatively low chromosome variability is probably promoted by regular sexual cycles, which helped in preservation of the genome organization (Delneri et al., 2003). *S. cerevisiae* is a diplontic yeast with highly clonal reproduction (Ezov et al., 2006). *S. cerevisiae* is also homothallic, which confers the possibility of regenerating a diploid cell from a haploid that could be interpreted as a way of genome renewal (Mortimer et al. 1994). Genomic instability, including aneuploidy and changes in chromosome structure, is apparently very low and is results of various mechanisms, such as inactivation of S-phase checkpoints and recombination, mismatch repair and telomere maintenance defects (reviewed in Kolodner et al., 2002).

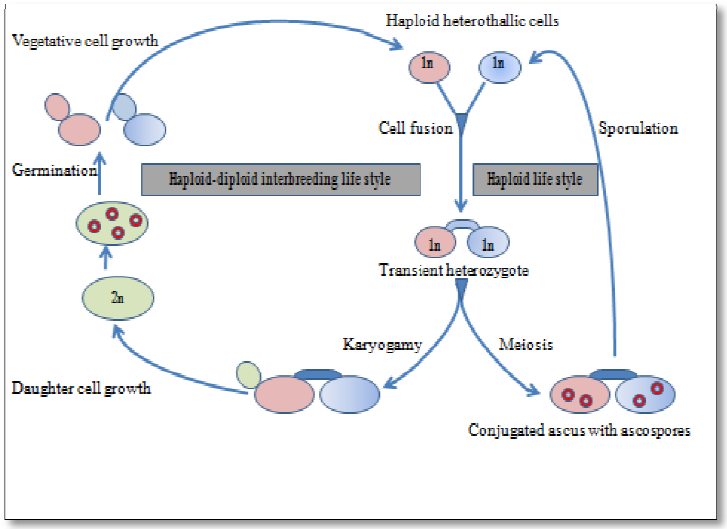


Figure 1.6. Schematic diagram representing different phases of *Zygosaccharomyces* life cycle.

The extent of intra- and interspecific genome variability is not well known for other yeasts, especially among distant relatives of *S. cerevisiae*. Recent works dealt with the analysis of a large number of strains of species other than those of *Saccharomyces* sensus stricto. The most studied ones are the species that belong to the CTG clade. For instance, *Candida albicans* is a diploid heterozygous yeast (Jones et al., 2004). Additionally, halotolerant yeast *Debaryomyces hansenii* was demonstrated to be a complex comprises of cryptic species of which some contain diploid heterozygotes. Recently, evidence was provided that halotolerant and osmotolerant yeast *Pichia sorbitophila* is a diploid heterozygote (Louis et al., 2012). Similarly, another halotolerant yeast *Millerozyma* (*Pichia*) *farinose* was recently characterized as species complex including haploid and heterozygous diploid cryptic species (Mallet et al., 2012). This widespread mosaic nature of these species is supposed to have arisen from genetic exchanges between divergent strains and species at high frequency. Genome structure is known to strictly dependent upon the life cycle and the breeding system in hemiascomycetous yeasts (Knop, 2006; Billiard et al., 2012). A number of studies have shed light on this interesting feature in *Zygosaccharomyces* (Dato et al., 2008; Rodrigues et al., 2003; Solieri et al., 2013b). Natural *S. cerevisiae* strains are often diplontic, homothallic and heterozygous yeasts, which have the possibility of regenerating homozygous diploid cell from a haploid cell through self-fertilization, a mechanism interpreted as a way of genome renewal (Mortimer et al., 1994). However, this is not true in relation to *Zygosaccharomyces* that show complex diploid (as in case of *Z. sapae* ABT301^{T} and ABT601) and aneuploid genomes (as in case of strain CBS 4837, CBS 4838, *Z. rouxii* ATCC 42981). These complex genomes are also referred as ‘mosaic genome’ and are thought to be evolved from mating between divergent strains or species and an comprehensive overview of population genetic perspective of these mosaic strains has been presented by Solieri et al. (2013b). Genome variability generates phenotypic heterogeneity and is of relevance for adaptation to environmental change, but the extent of such

variability in natural populations is still poorly understood. Solieri et al. (2013b) explicitly explained life cycle as the major evolutionary force that brings about genomic diversity. For example, selected *Zygosaccharomyces* strains are variable at the ploidy level, changes in gene copy number and chromosome number. This suggests that genome plasticity provides important genetic diversity which, in turn, can be apparently seen also as corresponding phenotypic divergence in strains of *Z. rouxii* complex.

1.3.2. Mating system: cell differentiation in yeasts

Like the specialized liver, muscle, and nerve cells of animals, the a, α, and a/α cells of yeast are differentiated cells that express different sets of genes. The best studied model is that of the commonly known baker's yeast, *Saccharomyces cerevisiae*. Mating type in *Saccharomyces cerevisiae* is determined by two nonhomologous alleles, *MAT*a and *MAT*α. These sequences encode regulators of the two different haploid mating types and of the diploids formed by their conjugation. Analysis of the *MAT*a1, *MAT*α1, and *MAT*α2 alleles provided one of the earliest models of cell-type specification by transcriptional activators and repressors. Similar to that in most eukaryotic species that reproduce sexually through differentiation of two alternate sexes, yeasts have also devised unique types of sexual differentiation mechanisms, whereby a single yeast cell can switch, by the process of recombination, its mating type between two alternate mating types, namely, a and ***a***.

Genes at the *MAT* locus determine cell type

Mating type is controlled by alleles of a single genetic locus, referred to as *MAT* (**ma**ting **t**ype). Haploid strains possess either the *MAT*a or *MAT*α allele; *MAT*a/α diploids are heterozygous at this locus, carrying both alleles. The MAT locus lies in the middle of the right arm of chromosome III,

~100 kb from both the centromere and the telomere. The two mating-type alleles, *MAT*a and *MAT*α, differ by ~700 bp of sequences. A cell with *MAT*a information at the MAT locus has an **a** mating type; while a cell carrying *MAT*α information has an **α** mating type. Clues regarding how the *MAT* locus determines these different mating types has first emerged from mutant analyses, where sterile mutants of α cells were isolated, one class of mutants had defects at the *MAT* locus. Surprisingly, the *MAT*α mutations fell into two complementation groups, indicating that instead of a single gene at the *MAT*α locus, there are two genes, α1 and α2. Mutations in the α1 and α2 complementation groups have varied molecular effects suggesting different roles of the genes. Strains carrying α1 mutations fail to synthesize pheromone, a protein normally made by α cells; while strains carrying α2 mutations synthesize a pheromone but degrade it before it can be secreted. Only *a* cells express the phenotype of α-pheromone degradation.

Mutant analyses of the *MAT*a locus indicate that it has a slightly different configuration from that of the *MAT*α locus. No sterility-producing mutations have been isolated that map at the *MAT*a locus. Haploid strains carrying *MAT*a mutations mate normally and fall into a single complementation group known as a1. The lack of any detectable haploid phenotype for *MAT*a mutations suggests that the a1 protein performs no function necessary for mating or differentiation into a cell. It is the lack of α1 and α2 functions that makes *a* cells what they are. *a* cells do not express α-specific functions because they lack the α1 activator, and they constitutively express a-specific functions because they lack the α2 repressor. In short, the *a* cell type arises through a default pathway requiring no action on the part of the *MAT*-encoded regulatory proteins.

The properties of these mutations at the mating type locus led to the proposal of the 'α1–α2 model' for control of cell type (Herskowitz and Oshima, 1981). This hypothesis stated: (1) that the *MAT*α1 gene product acts as a positive regulator of α-specific functions presumed to be encoded by α-

specific genes; (2) that the *MAT*α2 gene product is a negative regulator of **a**-specific functions presumed to be encoded by **a**-specific genes; and (3) that sporulation requires both a and α information, presumably the products of the *MAT*a1 and *MAT*α2 genes. This model made several specific predictions, which were subsequently tested and confirmed.

Mating results in fusion of haploid cells, forming *MAT*a/*MAT*α heterozygous diploid, which is no longer capable of mating. The nonmating phenotype has been explained by a model, according to which the a1 and α2 regulators act in concert to perform a series of negative functions. The nonmating phenotype of *MAT*a/*MAT*α diploids results from the action of a very stable corepressor of Mata1 and Matα2 proteins (Jensen et al. 1983; Goutte and Johnson 1988; Strathern 1988; Li et al. 1995; Johnson et al. 1998; Tan and Richmond 1998). This repressor turns off a set of haploid-specific genes (that normally function in both **a** cells and **α** cells but not in **a**/**α**) and allows expression of diploid-specific genes. The a1-α2 repressor turns off transcription of *MAT*α1, the activator of a-specific genes, but allows expression of *MAT*α2, the repressor of a-specific genes. The net result is that a/α diploid cells are sterile but competent to undergo meiosis. The coregulators further repress expression of the *RME1* gene. The presence of the Rme1 protein inhibits the meiotic pathway; its absence, therefore, confers competence for meiosis.

Mating type switching

Mating type is determined, as already mentioned, by two different alleles of the mating-type (*MAT*) locus. Yeast strains differ from each other according to the stability of their mating type. Most laboratory strains have stable haploid mating types such that the progeny of successive mitotic divisions always have the same mating type as their parents. These strains with stable haploid mating types are known as heterothallic strains. By contrast, strains that can switch mating types are known as

homothallic strains. Like many other fungi, budding yeast has acquired the capacity to convert some cells in a colony from one haploid mating type to the other. This process is called homothallism. The subsequent mating of cells to the opposite mating type enables these homothallic organisms to self-diploidize. The diploid state provides yeast with a number of evolutionarily advantageous strategies that are not available to haploids, for instance the ability to undergo meiosis and spore formation under nutritionally limiting conditions. The homothallic phenotype arises from a dominant, functional HO allele at the HO locus on chromosome IV; by comparison, heterothallism results from a recessive loss-of-function allele (ho) at this same locus. Mating-type gene switching in *S. cerevisiae* is a highly choreographed process that has taught us much about many aspects of gene regulation, chromosome structure, and homologous recombination (**Figure 1.7**).

Role of *HML*, *HMR*, and *HO* in mating type switching

The key to understanding the mechanism by which homothallic (HO) strains switch mating type came from the discovery of two additional copies of the *MAT* locus called *HML* and *HMR*. Located near the telomeres on each arm of chromosome III, the *HML* and *HMR* loci are transcriptionally silent. In wild-type strains, *HML* contains *α* information and *HMR* contains *a* information. The mechanism of mating type switching involves a transposition of mating type information from the transcriptionally silent *HML* or *HMR* locus into the transcriptionally active *MAT* locus. Switch ing occurs only in the *a* and α cell types. *a/α* cells do not switch even if they carry HO because the a1 and α2 regulators repress transcription of the HO gene.

A simple way to understand the mechanism of mating type switching in homothallic strains is to imagine an analogy with a cassette tape recorder. A cell can retrieve a cassette of mating type information located in an *HML* or *HMR* storage locus and insert it into the *MAT* playback locus. The

cell retrieves information from the storage locus by a series of steps. The product of the HO gene, a site specific endonuclease, initiates the process. The HO-encoded enzyme makes a double-stranded cut in chromosome III at a specific 18 bp recognition sequence just to the right of a segment known as the Y segment in *MAT*. Next, the double stranded break is repaired through a recombination-like process in which homologous pairing occurs between sequences from one of the storage loci, such as *HML*, and identical sequences surrounding *MAT*. In this example, *HML* DNA then serves as a template to resynthesize DNA at *MAT*. The result is the replacement of *MAT* DNA with DNA from the storage locus. Overall the process is conservative; with the duplication of storage locus DNA balanced by an equal loss of DNA from *MAT*. The information in the storage locus remains unchanged.

The storage loci are silent in two respects. They are not transcribed, and even though they contain the 18 bp HO protein-recognition sequence, they are not susceptible to double-stranded cutting by the HO enzyme. Both these properties result from the presence of "silencer" sequences at the boundaries of the storage loci, which are not present at the boundaries of *MAT*. In a process that may resemble the silencing of genes in higher eukaryotes, the silencer sequences in yeast contribute in a complex way to the configuration of the chromatin structure in the vicinity of each storage locus. RNA polymerases, transcription factors, and the HO endonuclease cannot gain access to these regions. As a result, the HO endonuclease cannot make double-stranded cuts at the storage loci. The silencing effect is general: Genes from other chromosomes that are normally active become transcriptionally silent when inserted into a storage locus.

Are there gene products that function to keep these storage loci silent? Mutational analyses show that the answer is yes. Loss-of-function mutations isolated in four *SIR* (silent information regulator) genes cancel repression and result in the transcriptional activation of the α1, α2 or a1 genes located at *HML* or *HMR*. In haploid cells, transcription of the storage loci results in a phenotype

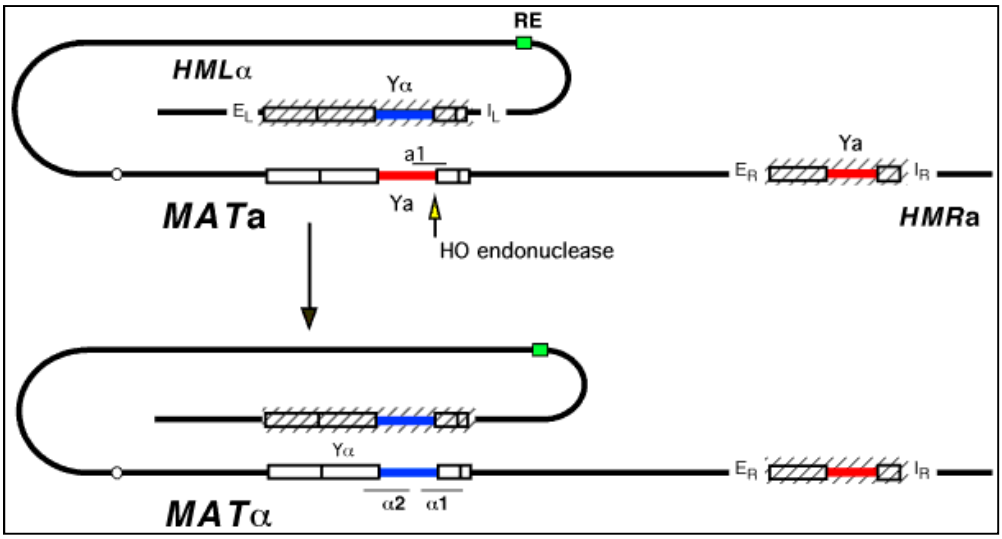


Figure 1.7. Arrangement of *MAT*, *HML* and *HMR* on chromosome III of *S. cerevisiae*. How *MAT***a** to *MAT*α gene conversion takes place has been illustrated. A bidirectional promoter transcribes **a**-and α-regulatory genes at *MAT*. Both *HML* and *HMR* could be transcribed but remain silenced due to interaction of silencing proteins with flanking E and I that represents the *cis*-acting loci that set up heterochromatic chromatin structures at *HML* and *HMR* (hatched lines) The recombination enhancer (RE) located 17kb centromere proximal to *HML* acts to promote the usage of *HML* as a donar in *MAT***a** cells (Source: Haber et al., 1998).

resembling that of a/α an diploid because both *MAT*a and *MAT*α alleles are expressed. Moreover, in strains, the HO enzyme cleaves the chromosome at both *MAT* and the storage loci. Depending on which site is cut in a given cell, HO mediated transposition can occur from *MAT* to one of the storage loci or in the other direction. This observation demonstrates that in normal SIR strains, the silencing of the storage loci ensures that *MAT* serves as recipient and the storage loci as donors in the switching process. As a consequence in the wild-type cells, the mating type information flows in only one direction from the storage loci to *MAT*.

1.4. Eco-physiological, genetic and mechanistic basis of osmotic adjustment

Zygosaccharomyces species represent an appropriate model for studying adaptation to osmotic and salts stresses, with some species being high osmotolerant-low halotolerant (*Z. sapae* and *Z. mellis*),

and other high osmotolerant-high halotolerant (*Z. rouxii*, *Z. bailii* and allodiploid/aneuploid strains of uncertain taxonomical affiliation) (Solieri et al., 2014a) (**Table 1.2**). Upon exposure to high extra-cellular concentration of solutes (sugar or salt), *Zygosaccharomyces* cells experience three major physiological alterations: changing in physical and chemical structure of cell wall and plasma membrane; increase of intracellular solute/ion toxicity; and alterations in osmotic pressure and cell volume. Therefore, three systems mechanistically enable cell to counteract stress challenges have been implicated: a) regulation of morphological and structural properties of cell wall and plasma membrane; b) modulation of transport activity to bring about intracellular solute/cation homeostasis; and c) production, accumulation and retention of metabolically compatible osmolytes.

Of the four classes of macromolecules that form the yeast cell wall as described by Klis et al. (2006), the physico-chemical structure and amount of mannosylated cell wall proteins (also called mannoproteins) and β-D-glucan are known to be affected in response to osmotic upshocks in *Z. rouxii* and *S. cerevisiae* (Hamada et al., 1984). Strain-specific differences in the internal layer of β-D-glucan and cell wall mannans have been noticed. Both macromolecules have been implicated in conferring rigidity and integrity to cell wall. In particular, *Z. rouxii* strains having a more rigid cell wall tend to be less halotolerant than those having a more flexible and elastic cell wall (Pribylova et al., 2007b). Besides this, an orthologous signaling pathway known as cell wall integrity (CWI) pathway, which is principally responsible for orchestrating wall changes in two main model organisms *Schizosaccharmyces pombe* and *S. cerevisiae*, has also been suggested to operate in *Z. rouxii* (Rodicio and Heinisch, 2010).

Category	Species name[1]	D-glucose % (w/v)		NaCl (M)	Food spoilage[2]	References
		50	60			
Moderately osmotolerant and moderately halotolerant	*Saccharomyces cerevisiae*	-	-	< 1.70	Soft drink, fruit juice	Onishi, 1963
	Schizosaccharomyces pombe	-	+§	1.00	Cheese, fruit (rarely)	Lages et al., 1999; Barnett et al., 2000
	Zygosaccharomyces florentinus (*Zygotorulaspora florentina*)	+	-	1.00	Wine	Lages et al., 1999; Barnett et al., 2000
	Candida glabrata	+	-	1.70	Juice concentrates	Pitt and Hocking, 2009
Osmotolerant and moderately halotolerant	*Pichia membranifaciens*	+§	-	3.00	Bread, fermented milk, olive	Lages et al., 1999; Barnett et al., 2000
	Zygosaccharomyces mellis	+	+	1.70	Juice concentrates, honey	James and Stratford, 2003
	Zygosaccharomyces sapae	+	+	2.0	N	Solieri et al., 2013a
	Zygosaccharomyces bailii	+	W	1.0-2.0	Juices,sauces, ciders, wines	Lages et al., 1999; Barnett et al., 2000
	Zygosaccharomyces bisporus	+	+§	1.0-2.0	Soft drink, wine	James and Stratford, 2003
Moderately osmotolerant and halotolerant	*Candida parapsilosis*	+	-	3.0	Dairy food	Pitt and Hocking, 2009
	Candida tropicalis	+	-	1.7-2.0	Fruit juice	Barnett et al., 2000; Deak, 2007; Pitt and Hocking, 2009
	Issatchenkia orientalis (Pichia kudriavzevii)	+§	-	2.0	Olives, pickles and sauces (rarely)	Lages et al., 1999; Barnett et al., 2000
Osmotolerant and halotolerant	*Pichia sorbitophila* (*Millerozyma farinosa)*	+	+	3.0-4.0	Beer, sake, soy sauce, mash of rice vinegar	Lages and Lucas, 1995
	Zygosaccharomyces rouxii (allodiploid strains)	+	+	3.0§	Juice concentrates, honey, jams, confectionery, dried fruits, soy sauce	Lages et al., 1999; Solieri et al., 2013a
	Candida magnolia	+	+	3.0	sugary food	Aguiar and Lucas, 2000; Barnett et al., 2000
	Pichia guillermondii (Meyerozyma guilliermondii)	+§	+§	3.0	Olive, salt meat	Butinar et al., 2005
Extreme halotolerant	*Hortaea werneckii*	+	N	5.50	Salt fish	Butinar et al., 2005; Lenassi et al., 2011
	Debaryomyces hansenii (Candida famata)	+§	+§	3.0-4.0	Olive	Aguiar and Lucas, 2000; Barnett et al., 2000; Lages et al., 1999
	Candida halophila (*Candida versatilis*)	+	+	4.0-5.0	Cheese brines	Barnett et al., 2000; Silva-Graça and Lucas, 2003

Table 1.2. Classification and characteristics of representative osmo- and halotolerant yeast species (Source: Dakal et al., 2014) [[1] species name according to the corresponding reference; current name in bracket; [2] Food spoilage information from Pitt and Hocking, 2009; § variable trait; N, not reported; w, weak.

Zygosaccharomyces species are sodium-excluders and in them Na^+ is comparatively more cytotoxic as compared to K^+. In order to prevent the influx of excessive amount of Na^+, several cation transport systems are present: 1) Na^+/K^+ P-type -ATPase, also called as sodium pump, encoded by *ZrENA1* in *Z. rouxii*; and 2) Na^+/H^+antiporter encoded by *ZrSOD2-22* and other variants in *Z. rouxii* strains (Watanabe et al., 1995) (**Figure 1.8**).

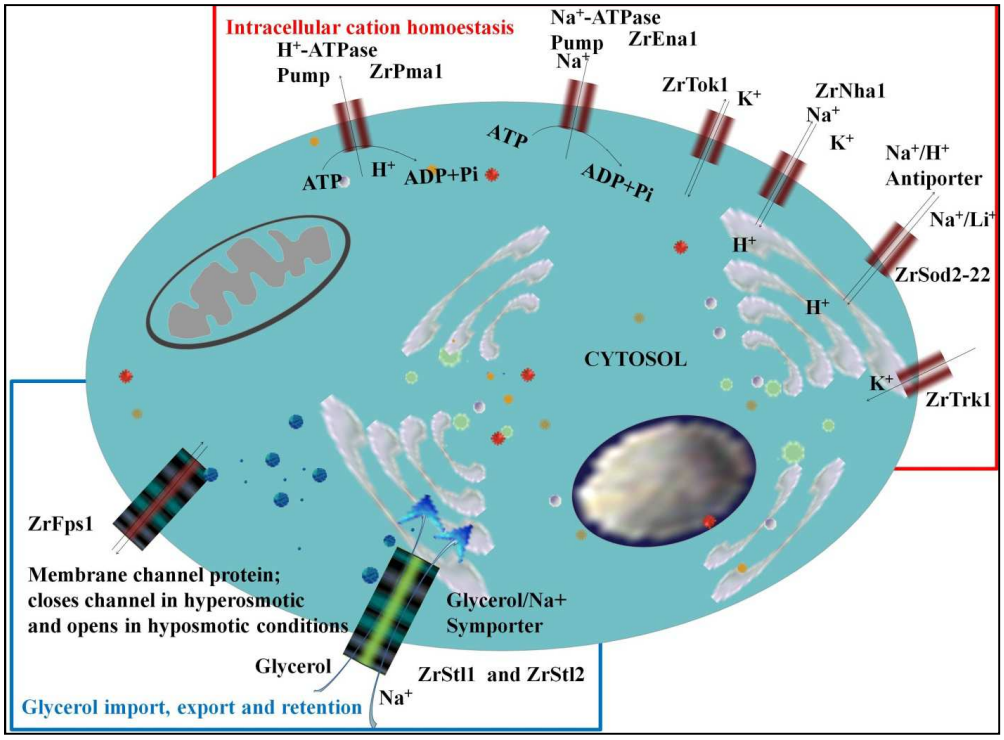


Figure 1.8. Mechanistic basis of cation homeostasis and glycerol uptake/retention.

In sodium excluder yeasts, expulsion of excessive cations (Na^+ or K^+) through the plasma membrane is executed by: 1) Na^+/K^+ P-type -ATPase, also called as sodium pump, encoded by *ENA1-4* genes in *S. cerevisiae*, and *ZrENA1* in *Z. rouxii* (**Figure 1.8**); and 2) Na^+/H^+antiporter encoded by *NHA1* in *S. cerevisiae*, *SpSOD2* and *SpSOD22* in *S. pombe*, *ZrSOD2-22* and other variants in *Z. rouxii* strains (Watanabe et al., 1995; Hahnenberg et al., 1996; Bañuelos et al., 1998; Benito et al., 2002). Functionally, the Na^+/H^+-antiporters can be divided into two subfamilies: those recognizing only Na^+ (and its analogue Li^+) as substrate, such as *S. cerevisiae* Nha1; and those recognizing both K^+ and Na^+

(and their respective analogues Rb^+ and Li^+), such as SpSod2 and ZrSod2 (Kinclová et al., 2001). The former category is more common among halotolerant yeasts, but *Z. rouxii* harbors also an additional gene for the latter one, namely *ZrNHA1* (Pribylova et al., 2008). Tolerance to salts is significantly dependent on the presence of this genenetic repertory. For instance, in *S. pombe* salt-sensitivity is due to presence of two Na^+/H^+ antiporter genes and the lacks of *ENA* gene (Bañuelos et al., 1995; Papouskova and Sychrová, 2007). In *Z. rouxii*, salt tolerance has been attributed to different variants of Na^+/H^+ antiporters and Na^+-ATPase genes (Hahnenberger et al., 1996; Pribylova et al., 2008). Watanabe et al. (1999) cloned the gene homologue of *S. cerevisiae ENA1* gene from *Z. rouxii* strain, named *ZrENA1*. The results pointed out that the Na^+-ATPase encoded by *ZrENA1* has little relevance in *Zygosaccharomyces* salt tolerance (Watanabe et al., 1999).

Under NaCl shock, the major Na^+ pumpout activity relies on Na^+/H^+ antiporter in the highly salt-tolerant *Z. rouxii* (Watanabe et al., 1995, 1999) (**Figure 1.8**) and on Na^+-ATPase in the moderately salt-tolerant *S. cerevisiae* (Prior et al., 1996; Bañuelos et al., 1998). More recently, it was demonstrated that upon initial imposition of NaCl stress, *S. cerevisiae* extrudes intracellular Na^+ primarily by Nha1, whereas the long term salt adaptation is mediated by trascriptional upergulation of *ENA1* gene (Ruiz et al., 2007; Proft and Struhl, 2004; Ke et al., 2013). Other studies reported that both, Na^+-ATPases and Na^+ antiporters, serve almost similar functions, but they are operational at different external pH levels as in sodium includer *D. hansenii* and sodium excluder *S. cerevisiae*. While Na^+ pump exports alkali-metal-cations mainly at high external pH levels, Na^+/H^+-antiporters Nha1 operates at lower external pH levels (Bañuelos et al., 1998). Therefore, regarding Na^+ extrusion in sodium excluder yeast *Z. rouxii*, it can be speculated that extracellular environmental pH could be an important determinant of differential activity of two ion extrusion systems.

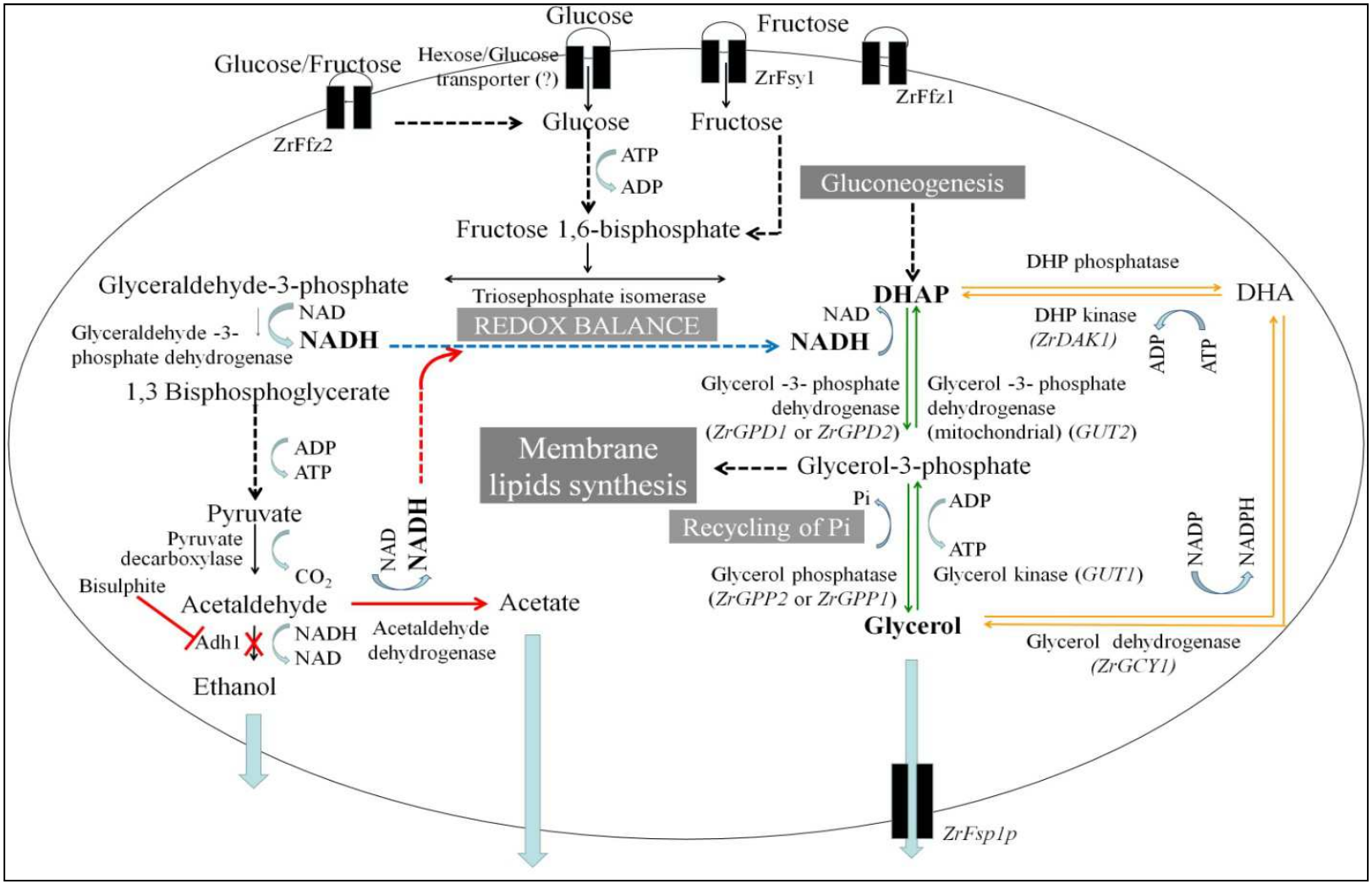


Figure 1.9. Interactions among redox balance, glycolytic pathways, and glycerol production in *Zygosaccharomyces rouxii*. Green and orange lines indicate enzymatic steps in Gpd-Gpp and Gcy-Dak pathways, respectively. Black dotted line indicates omitted steps; blue dotted lines indicate relationships among the pathways under non-stress conditions; and red dotted line, red line, red-cross indicate direction of metabolic flux flow under bisulphite and osmostress conditions so as to provide enough NADH for glycerol production. Roles of glycerol metabolism in Pi recycling, redox balance, gluconeogenesis and fatty acid biosynthesis are reported in gray box.

Glycerol is the main osmolyte that is produced and accumulated intracellularly in response to hyperosmotic stress, and thus the most studied one. Like *S. cerevisiae*, the genome of *Z. rouxii* allopdiploid strain ATCC 42981 harbours two isoforms of genes involved in glycerol synthesis, such as *ZrGPD1* and *ZrGPD2*; *ZrGPP1* and *ZrGPP2*; *ZrGCY1* and *ZrGCY2*; and one isoform of *ZrDAK1* (Iwaki et al., 2001; Wang et al., 2002; Watanabe et al., 2004) (**Figure 1.9**). On the basis of deduced amino-acid sequence, these proteins have close homology with their counterparts in *S. cerevisiae*. The CNV of these genes occurring in *Z. rouxii* genome raises the possibility of their functional differentiation. *ZrGPD1*/*ZrGPD2* and *ZrGCY1*/*ZrGCY2* are constitutively expressed in *Z. rouxii* cells, but their differential roles have not yet studied (Iwaki et al., 1999, 2001; Watanabe et al., 2004). Heterologous expression of *ZrGCY1* but not of *ZrGPD1* restores glycerol production and contributed to salt tolerance in *S. cerevisiae gpd1Δgpd2Δ* mutant strain unable to synthesize glycerol (Watanabe et al., 2004). Moreover, it was found that glycerol production and salt tolerance increase when *ZrGPD1* is expressed along with *ScGPP2* (Watanabe et al., 2004). These evidences support that the main metabolic route leading to glycerol from DHAP in *Z. rouxii* is intermediated by glycerol 3-phosphate (Gpd-Gpp pathway), similarly to that occurs in *S. cerevisiae*.

Other important compatible osmolytes are arabitol and mannitol, but their role in osmoregulation is yet unclear. In *Z. rouxii*, salt and sugar stresses induce the production of glycerol, arabitol or both, depending upon the osmoticum which has been used to lower the a_w. If sugar is used as stress agent in spite of salt, increased amounts of D-arabitol are produced and accumulated, whereas glycerol concentration remains costant (van Zyl and Prior, 1990). Fructose and glucose-containing media have been associated to mannitol production (Tomaszewska et al., 2012), which is inhibited by salt (Onishi and Suzuki, 1968; van Eck et al., 1993; Tomaszewska et al., 2012).

Yeast cells responds to osmo- and salt-stress conditions at system level by up- and down-regulating hundreds genes (Posas et al., 2000; Rep et al., 2000; Causton et al., 2001; Hohmann, 2000; Gasch et al., 2004). Their regulation is mediated mainly via three well-characterized molecular signaling pathways as depicted in **Figure 1.10**: the HOG pathway, one of the five MAPK pathways known in yeasts (de Nadal et al., 2002; Hohmann, 2002); the calcineurin/Crz1p pathway, which is specifically required for adaptation to high-salt conditions (Rusnak and Mertz, 2000); and the Ras-cAMP signalling pathway (Thevelein and de Winde, 1999; Norbeck and Blomberg, 2000).

HOG pathway is a highly structured and integrated MAPK module which is responsible for tranduction of osmosensory signals in yeast species (Brewster et al., 1993). The cascade is comprised of consecutively activated three kinases: MAPKKK, MAPKK and MAPK. The yeast In *S. cerevisiae*, the osmotic stress signals perceived by the Slnl osmosensor is, in turn, transmitted from Ypd1 to Sskl to Ssk2/22 (MAP kinase kinase kinase) to Pbs2 (MAP kinase kinase) and then finally to the MAP kinase Hogl (Posas et al., 1996). On the other hand, Shol (Maeda et al., 1995) regulates the action of Hogl via Stell (MAP kinase kinase kinase) and Pbs2 (Posas and Saito, 1997). Once phosphorylated, P-Hogl translocates into the nuclear compartment and activates transcription factors therein leading to transcription of stress responsive genes, which possess stress regulatory elements (STRE) (Schüller et al., 1994).

The calcineurin/Crz1 signal transduction pathway has an important role in ion regulation (Matsumoto et al., 2002; Ke et al., 2013), but it is relatively less investigated as compared to HOG pathway. Calcineurin, the Ca^{2+}/calmodulin-regulated protein phosphatase 2B, is highly conserved through evolution and is a critical component of Ca^{2+}-regulated signaling in different unicellular (such as yeasts) and higher eukaryotes. Calcineurin is a heterodimer containing a catalytic (A) subunit complexed with an essential regulatory (B) subunit and requires Ca^{2+} and calmodulin for

activity (Cyert et al., 1991). Calcineurin controls Crz1 activity by regulating its subcellular localization (Stathopoulos-Gerontides et al., 1999). When calcineurin-dependent signaling is low, Crz1 is phosphorylated and resides primarily in the cytosol. Upon dephosphorylation by calcineurin, Crz1 enters into nucleus and binds specifically to the Calcineurin Dependent Response Element (CDRE) present in the promoter region of target genes through a *C2H2* zinc finger motif (Stathopoulos and Cyert, 1997; Matheos et al., 1997). Calcineuirn/Crz1 pathway controls expression of the P-type ATPase Ena1 for Na^+ and Li^+ efflux. High cytosolic Ca^{2+} levels activate calcineurin in response to extracellular hyperionic stress. ans increase Na^+ efflux by up-regulating the *ENA* gene (Matsumoto et al., 2002) Ke et al. (2013) demonstrated the restricted role of calcineurin pathway in maintaining a non-toxic level of intracellular Na^+ in the long-term adaptation to NaCl stress, and also showed that the pathway has no role in maintaining Na^+ at low level under high non-ionic stress.

The yeast cAMP-dependent protein kinase (PKA) is the effector kinase of the Ras-cAMP signaling pathway. It is a conserved serine/threonine protein kinase, which thourgh phosphorylation of different targets has pleiotropic effects on cell growth, threalose and glycogen metabolism, dimorphic shift, and stress adaptation (Smith et al., 1998). The influence of PKA on protein expression during exponential growth of *S. cerevisiae* under osmotic stress was studied by proteomic (Boy-Marcotte et al., 1998). Proteins upregulated under NaCl stress could be grouped into three classes with respect to PKA activity: i) proteins PKA-independent proteins (Gpd1, Gpp2 and Dak1); ii) fully PKA-dependent proteins (Tps1 and Gcy1); iii) partly PKA-dependent protein (Eno1, Tdh1, Ald3, and Ctt1) (Boy-Marcotte et al., 1998). Very less has been researched regarding the role of Ras-cAMP signaling pathing in context to *Zygosaccharomyces* yeasts.

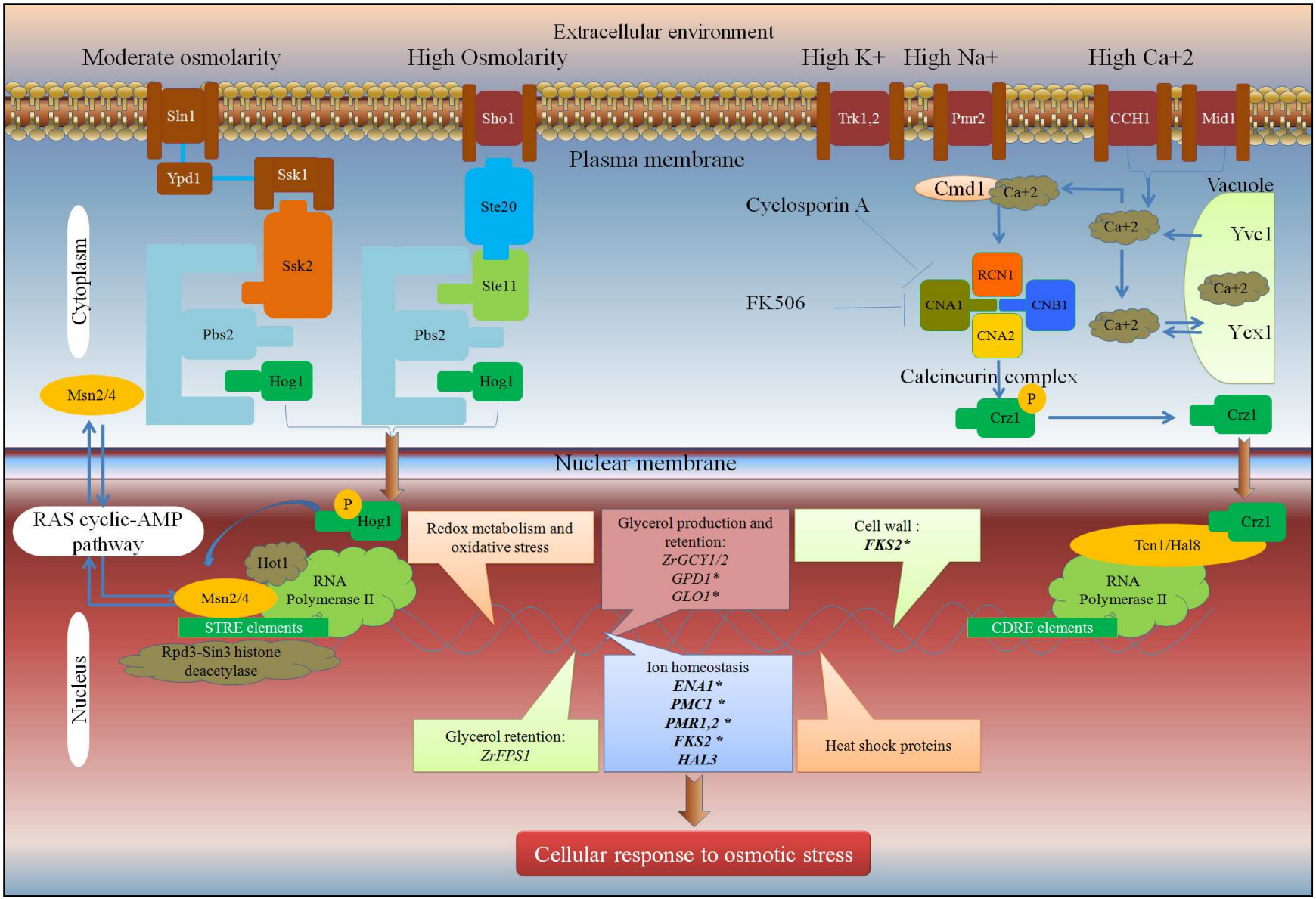


Figure 1.10. Integrated overview of signalling transmission pathways involved in osmodaptative genes regulation. Genes are referred to as *Saccharomyces cerevisiae* genome, with the exception of *ZrFPS1* and *ZrGCY1/2*. Genes regulated by calcineurin pathway are reported in bold.

1.5. Food and biotechnological perspectives

Fundamental knowledge of the physiology and metabolism of yeasts is a prerequisite for the successful use of these organisms in industries and bioprocesses (Giudici et al., 2005). This is true in terms of both for improvements in existing technology, such as production of wine, fermented foods and beverages; and for the development of new technology. Although, *Saccharomyces cerevisiae* was the first yeast to be investigated and successfully implicated for heterologous protein expression in eukaryotes; however, *S. cerevisiae* is not suitable for many biotechnological purposes (Breuer and Harms, 2006). Specific disadvantages encountered in the use of this organism are hyperglycosylation and low secretion efficiency. Another complicating factor in the cultivation of *S. cerevisiae* on sugars is its strong tendency to produce ethanol also in aerobic conditions, the so-called ‘‘Crabtree-effect’’. The close relation between *Z. rouxii* and *S. cerevisiae* could result in interesting biotechnological applications. Osmo- and salt-tolerance in *Z. rouxii* is supposedly due to specific genes whose identification, functional characterization, cloning, and expression in industrial strains of *S. cerevisiae* could improve the growth and fermentation capacities of latter. Having recognized the significance of *Z. rouxii* and *Z. bailii* and other halotolerant species for industrial and technological purposes, over the past several years, more focus has been placed on dissecting the genetic basis of osmostress tolerance in food spoilage yeasts through functional genomics and gene expression studies (Plemenitaš et al., 2008). Surprisingly, *Z. rouxii* strains possess some potent industrial properties which make them more preferable over *S. cerevisiae* strains. For instance, *Z. rouxii* strains are capable of growing at extreme physiological conditions such as very low water activities, where *S. cerevisiae* strains can’t (van Eck et al., 1993). Identification of osmo-responsive genes in *Zygosaccharomyces* and their heterologous expression in industrial strains such as *S. cerevisiae*, could be of a great economical interest to food manufacturing and processing industries.

The ascomycetous yeast *Z. rouxii* and *Z. bailii* are the most problematic spoilage yeasts in food and beverage industries, due to their exceptional tolerance to weak-acid preservatives and the ability to adapt to high osmotic concentrations and high temperatures, vigorously ferment sugar, and grow at low pH. Knowledge of mechanisms regulating osmostress tolerance could increase prevention and prediction of food spoilage by yeasts (Plemenitaš et al., 2008). In this respect, the understanding of genetic determinants underlying osmostress tolerance opens important perspectives. On the other hand, osmo- and salt-tolerance in *Z. rouxii* is due to specific genes whose functional characterization, cloning, and expression in industrial strains of *S. cerevisiae* could improve their growth and fermentation performance. For example, *Z. rouxii ZrSOD2-22* is so efficient for sodium and lithium extrusion at acidic external pH values, that it can effectively used in improving the salt tolerance of *S. cerevisiae* industrial strains (Watanabe et al., 2003). Heterologous expression of *Z. rouxii ZrSOD2*-22, but not of *ZrSOD2* Na^+/H^+ antiporter, increases expulsion of toxic Na^+ and Li^+ cations from *S. cerevisiae* cells (Prior et al., 1996; Kinclová et al., 2001). Moreover, eukaryotic halotolerance is important for alleviating problems such as those caused by soil salinization in agriculture (Gostinčar et al., 2012) or osmotic stress in the production of bioethanol (Zhao and Bai, 2009).

1.5.1. Heterologous protein and metabolite production

Although current commercial uses of halotolerant and osmotolerant yeasts are quite obvious, such as fermentation of miso paste, soy sauces and wine; many novel and unique genetic properties of many of these yeasts have suggested that they have even greater potential for biotechnology, especially in the production of heterologous protein and metabolite production (Breuer and Harms, 2006; Liu et al., 2006). Therefore, changes in the cellular physiology on account of changes in external osmotic pressure have been the subject of several investigations. The osmotic stress conditions affect many

cellular processes such as protein folding, ribosome biogenesis and cell wall formation (Dragosits et al., 2010). This has encouraged researchers to investigate other yeasts in order to gain insights into the wealth of metabolic, biochemical and genetic diversity in order to exploit them industrially. For example, interest was aroused in extremophilic yeasts, including *Z. rouxii*, *D. hansenii*, *P. sorbitophila, C. halophila* that proved to be genetically and biochemically promising for industrial production of heterologous proteins and metabolites. Besides this, nonconventional yeasts, such as *Kluyveromyces lactis*, *Yarrowia lipolytica* and methylotrophic yeasts as *Hansenula polymorpha* and *Pichia pastoris* have also gained increasing interest as attractive hosts for the production of recombinant proteins and metabolite. The ability of an organism to grow at high osmotic, high temperature and acidic condition is of great economical importance to any fermentation and processing industry, as they can simplify the process. Few studies have reported that the osmotic stress could be beneficial in recombinant protein production from bacterial, yeast and mammalian host cells (Blackwell and Horgan, 1991; Shi et al., 2003; Kim et al., 2003). There are some reports where genetic modification and production of heterologous proteins has been demonstrated in *Zygosaccharomyces* (Branduardi, 2002, Branduardi et al., 2004; Vigentini et al., 2005). For instance, *Z. bailii* ability to withstand hyper-osmotic, hyper-thermic and acidic environments renders it suitable for the development of biotechnological processes for production of heterologous protein, such as interleukin-1β (Vigentini et al., 2005). Besides this, production of metabolites has been also demonstrated in some yeasts (Liu et al., 2006). For instance, osmotolerant strains of *M. farinosa* and *P. sorbitophila* that assure high yields of glycerol and xylitol, have presented possiblitity of their valuable use in biotechnology (Mallet et al., 2012). Several osmophilic and osmotolerant yeasts have been employed in large scale fermentation process for the production of important metabolites, of which glycerol is one of the most important one (Liu et al., 2006).

The ability of *Zygosaccharomyces* strains to grow at extreme physiological conditions such as very low a_w gives these strains very attractive as biocatalyzer and as microbial cell factory for the expression of heterologous proteins and metabolite production (Breuer and Harms, 2006; Liu et al., 2006). Genetic modification and production of heterologous proteins have been demonstrated in *Zygosaccharomyces* (Branduardi et al., 2004; Vigentini et al., 2005). *Z. bailii* was used as host for production of interleukin-1β due to its ability to withstand hyper-osmotic, hyper-thermic and acidic environments (Vigentini et al., 2005). Taken advantage of the ability to produce polyols, osmotolerant and halotolerant yeasts, such as *M. farinosa* (formely *P. sorbitophila*) and *C. tropicalis*, have been employed in large scale fermentation process for the production of glycerol and xylitol (Liu et al., 2006; Mallet et al., 2012). *Z. rouxii* strains have been exploited as cell factories for producing enzymes such as glutaminase (Kashyap et al., 2002; Iyer and Singhal, 2008), D-arbitol (Saha et al., 2007), chiral compounds (Hauck et al., 2003), and heterologous proteins (Ogawa et al., 1990).

Besides production of heterologous protein and metabolites in yeasts there exist many other ways in which yeast species including species of genus *Zygosaccharomyces* can be exploited industrially. Wegner (1983) listed *Z. bailii* as possible species for the production of single cell protein (SCP). *Z. bailii* and *Z. bisporus* are known to contain 2 μm-like episomal plasmids (Toh et al., 1984), which may aid transformation in these species.

1.5.2. Fermentation of food and beverages

Z. rouxii has been regarded as the most important yeast for fermentative elaboration of Japanese soy-sauce and miso (Suezawa et al., 2008) as it produce a variety of flavor compounds. Other evidences show that some novel *Zygosaccharomyces* species, such as *Z. sapae* and *Z. gambellarensis*, isolated from Italian traditional balsamic vinegar (Solieri et al., 2007; Solieri et al., 2013a) and highly

sugary wine (Torriani et al., 2011) respectively, could have possible role in fermentation and sensorial attribution (Solieri and Giudici, 2008).

1.5.3. Enhancement of sensorial attributes

Zygosaccharomyces species have found large application as biocatalyzer of peculiar low food and beverages. The osmotolerant and halotolerant *Z. rouxii* plays a central role in the flavor formation during soy sauce and brine fermentation (Sluis et al., 2001; Cao et al., 2009; Wei et al., 2013). Certain flavor components, such as 4-hydroxyfuranone derivative, 4-hydroxy-2 (or 5) -ethyl-5 (or 2) -methyl-3 (2H)-furanone (HEMF), are produced from D-fructose-1,6-bisphosphate by *Z. rouxii* and positively affect flavor and quality of soy sauce ((Hauck et al., 2003; Kataoka, 2005). A highly salt-tolerant *Z. rouxii* mutant constructed by whole genome shuffling strategy, showed increased growth rate in soy sauce fermentation compared to wild-type strain and improved ability to produce aroma compounds such as ethyl acetate, HEMF and 4-ethylguaiacol in high-salt liquid fermentation (Cao et al., 2009).

1.6. Molecular tools and techniques for genotypic and phenotypic characterization

1.6.1. Molecular typing for yeast characterization

Recent progress in molecular biology has led to the development of new techniques for yeast identification based on similarity or dissimilarity of DNA, RNA or proteins. These include allozyme patterns, DNA-DNA hybridization, electrophoretic karyotyping, microsatellite analysis, nested-PCR, random amplified polymorphic DNA (RAPD) analysis, RFLP of chromosomal DNA or RFLP of mitochondria1 DNA (as reviewed by Alcoba-Flórez et al., 2007).

From the above mentioned techniques, in the last decade, molecular genotyping techniques that directly rely on sequence differences have been used extensively. Several methods are now available to study relatedness between yeast strains and to infer DNA sequence similarity. Electrophoretic karyotyping is based upon the variable migration of large DNA restriction fragments in an electrical field of alternating polarity and is also referred as pulse-field gel electrophoresis (PFGE) (Gorman and Adley, 2006). Randomly amplified polymorphic DNA (RAPD) (Lieckfeldt et al., 1993; Baleiras Couto et al., 1994; Tornai-Lehoczki et al., 2000) is based on the amplification of random genomic DNA fragments by arbitrarily selected polymerase chain reaction (PCR) primers such as M13. Amplified fragment length polymorphism (Vos et al., 1995; Azumi et al., 2001) is based on selective PCR amplification of restriction fragments from a total digest of genomic DNA, and restriction fragment length polymorphism (RFLP) of specific DNA regions (ITS, NTS, etc.) (Baleiras Couto et al., 1996; Jespersen et al., 2000). These techniques can be applied to quantify genetic variation among yeast strains (Tornai-Lehoczki et al., 2000; Azumi et al., 2001; Baleiras Couto et al., 1996; de Barros Lopes et al., 1999). However, they scan only a fraction of the genome and do not yield information on the chromosomal location of polymorphisms or detect the presence, absence or modification of specific genes. In context to *Zygosaccharomyces* strains or isolates, several molecular and typing methods have previously been used to type and distinguish isolates of this yeast at the molecular level in order to determine source of isolation, source of contamination of food and beverages, and selection of starter cultures (Soleiri et al., 2008a; Wrent et al., 2010).

Electrophoretic karyotyping

Electrophoretic karyotyping or pulse-field gel electrophoresis is a highly discriminative molecular typing technique that is used in microbial genetics studies worldwide. The advent of electrophoretic techniques for separating the intact chromosomal DNA molecules of lower eukaryotes,

such as yeasts, has provided during last two decades means of characterizing the chromosome sets of these organisms. These techniques have provided fundamental new information about the basic organization of the genomes of many species of yeasts. These approaches are based on the electrophoretic separation of undigested genome DNA or in the comparison of genome macro-restriction patterns obtained by genome digestion with low frequency restriction endonucleases (Lopez-Ribot et al, 2000; Shin et al, 2004; Chen et al, 2005).

Electrophoretic karyotyping is based upon the variable migration of large DNA restriction fragments in an electrical field of alternating polarity and is also referred as pulse-field gel electrophoresis (PFGE) (Johnston, 1994; Gorman and Adley, 2006) (**Figure 1.11**). By comparing the fingerprints of any two isolates, one can investigate if they belong to the same species/strain (i.e. the two isolates are clonal) or if they are genetically unrelated. In practice, standardized cell suspensions of investigated isolates are embedded in agar to enhance DNA stability during the procedure. Then, yeast cells are lysed and large fragments of the chromosome are generated by macro-restriction analysis. DNA fragments are separated during electrophoresis using an electrical field of alternating polarity, after which the fingerprints are visualized and documented. Comparison of such profiles has been a great advance in the species differentiation within the genera *Candida*, *Saccharomyces*, *Kluyveromyces* and *Zygosaccharomyces*, in the study of anamorph teleomorph relationships between *Candida*, *Kluyveromyces, Pichia* and *Saccharomyces* species, as well as for synonyms verification (Belloch et al, 1997; Solieri et al., 2008a). This technique has also been successfully used to detect polymorphism in industrial *S. cerevisiae* strains (Bidenne et al., 1992; Casey et al., 1988; Hayford and Jespersen, 1999; Codón and Korhola, 1998) as well as in laboratory strains (Casaregola et al., 1998).

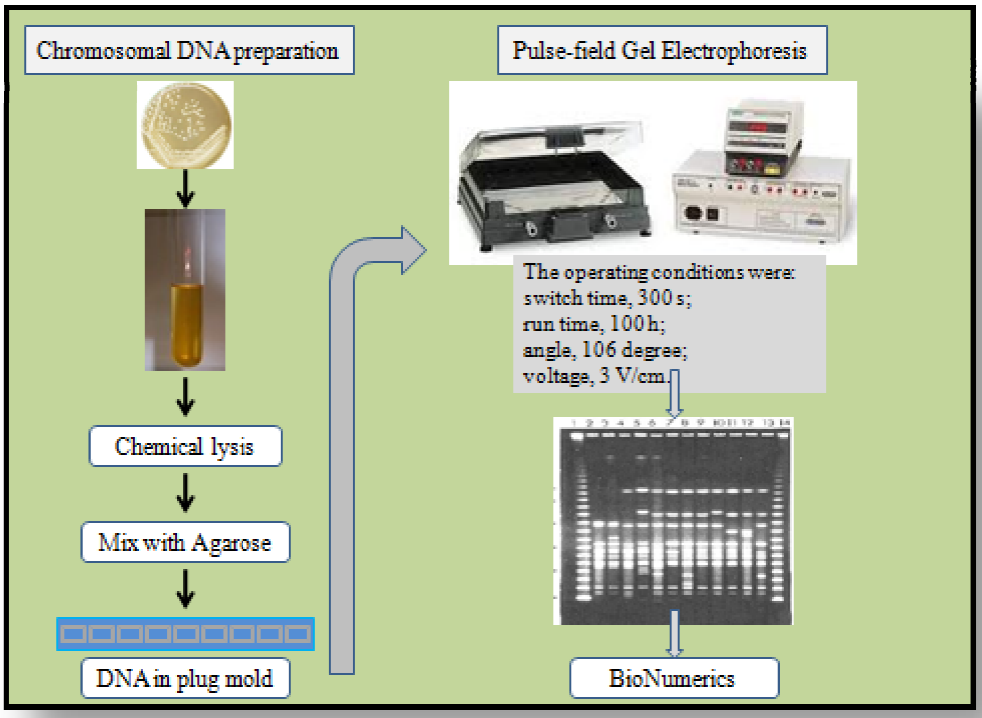


Figure 1.11. **Schematic workflow for karyotypic analysis, genome size and chromosome number estimation.**

Electrophoretic karyotype analysis has frequently been used and has been demonstrated to have a high degree of discrimination. It gives information about number of chromosomes and size of each chromosome and entire genome. However, it does not give any information regarding the ploidy status of yeast, which could be determined separately by another molecular technique, flow cytometric analysis. Thus, the method lacks the ability to accurately determine the degree of genetic relatedness between strains. Moreover, the technique is generally performed in two steps: plug preparation and PFGE. For complete karyotypic analysis it takes two-three days.

Mitochondrial DNA polymorphism (mtDNA)

Yeasts are eukaryotic microorganisms that show a wide range of variability in the mitochondrial DNA size, ranging from 6 to 25 μm of length. In most yeast species, mitochondrial

genome has circular topology. The use of mtDNA in yeast taxonomy has several advantages as: (a) small size, (b) high number of mitochondrial DNA molecules by cell, and (c) one single mitochondrial karyotype in each wild dikaryotic isolate (Belloch et al, 1977). During long time, the main limitation of this technique was the difficult isolation of mithocondrial DNA. In 1990, Querol et al designed a rapid method to overcome this problem and later, in 2001, this new protocol was slightly modified by López et al. (2001). The modified protocol comprises the standard miniprep isolation of yeast total DNA, and the use of restriction endonucleases that recognise a large number of sites in yeast nuclear DNA, but few sites in the mitochondrial DNA. The method permits to analyse the mtDNA without previous isolation and purification requirements. The technique is based on the GC content differences between the nuclear DNA (nDNA) and the mtDNA, being the GC% around 40% in the former but about 20% in the later. This difference brings that when total fungal DNA is digested with restriction enzymes that only recognize GC rich regions, as for example *Msp*I, *Hae*III and *Cfo*I, all with 50% GC target site, the nuclear DNA is over-digested giving rise to a high number of short fragments, that are not detected by conventional agarose gel electrophoresis. This characteristic permits to assume that when total DNA digested with these endonucleases is subjected to agarose electrophoresis only the mtDNA fragments will be observed. These fragments will be ordered by size constituting species specific patterns (Fernandez-Espinar et al, 2000; Rycoyska et al, 2004) or even strain specific ones (Sabate et al, 1998; 2002).

Restriction fragments length polymorphism (RFLP)

This technique is based on the differentiation between microorganisms by the comparison of the restriction patterns obtained by digestion of a chosen target DNA with restriction endonucleases. In general, yeast phylogenetic markers DNA sequences, such as 5.8S ITS genes and D1-D2 domain of 26S rDNA region, encompassed in its rDNA are first amplified using standard PCR reaction and later

for restriction digestion with suitable restriction endonucleases. The degree of similarity of the generated patterns allows establishing correlations between species, whilst the existence of unique patterns permits their use as phylogenetic markers. This method has been successfully used to differentiate between the species of genera *Candida*, *Cryptococcus* etc. (Esteve-Zarzoso et al, 1999; Sabate et al, 2002; Deak et al, 2004; Pinto et al, 2004). This technique has been succesfully applied using the ribosomal DNA region including the intergenic spacers ITS1 and ITS2, and the 5.8S rRNA enoding gene (Kurtzman 1994; Esteve-Zarzoso, 1999).

The 5.8S ITS gene has a highly conserved sequence showing a low intraespecific variability which is not enough to delimitate between conspecific strains (Solieri et al., 2013a). However, the ITS regions, which are non coding hypervariable ones, could permit depending on the case the identification at the intra or interespecific levels (Solieri et al., 2013b). In several studies, this technique is used in combination with PCR, what is called PCR-RFLP. In this combinative method, firstly specific DNA fragments are amplified by PCR and then, these amplicons are digested with restriction endonucleases to obtain specific patterns (Dendis et al, 2003; Llanos-Frutos et al, 2004).

Random amplified polymorphic DNA (RAPD)

This typing system is based on the PCR amplification of genomic DNA on the presence of a single short primer, often 10 nucleotides of length. Due to the use of a low annealing temperature (35-39 ºC), the primer binds to unspecific target sites, sites that are randomly distributed along the genome, what finally permits the obtention of DNA polymorphic amplicons. The amplified products are separated and visualized by gel electrophoresis. The use of RAPD permits to obtain the so called fingerprints which are combinations of different numbers of amplicons with different sizes, generating a pattern which is species or even strain specific (Orbera, 2004; Ergon and Gulay, 2005). RAPDs have

been used to develop genetic markers within several species and to discriminate between varieties of pathogenic yeasts. By means of this technique, the different *Candida* spp. serotypes have been distinguished (Alonso-Vargas et al 2000).

The number of fragments constituting the RAPD, the amplification intensity and the reproducibility of results depend on the amplification conditions, the components of the reaction mixture and the thermal cycler. These limitations make obvious the need of subjecting this technique to an optimization process, by which are defined the conditions for obtaining reproducible and reliable patterns. However, although these restrictions could make RAPD to seem a very limited approach, it constitutes a highly discriminative typing technique yielding feasible results comparable within a laboratory. But, patterns should not be compared between different laboratories although conclusions can be correctly extrapolated.

These techniques are methodologically easier, less time-consuming, and more cost-effective as compared to other genomic typing strategies, in particular, karyotyping using pulse-field gel-electrophoresis. However, a limitation of RAPD and microsatellite primed PCR has been the observation that the low-stringency conditions of these PCR-based methodologies may result in poor reproducibility of typing strain. While RAPDs are generally available to many researchers, to date, there are no standardized sets of primers, isolates, or amplification conditions; and a more serious problem is the relative instability of RAPD profiles.

Microsatellite-primed PCR analysis.

This molecular approach is based on the PCR amplification of fragments using oligonucleotides complementary to single repetitive sequences present in the target DNA. These repetitive sequences are called microsatellites. In this approach oligonucleotides complementary to

microsatellites, that is, tandem repeats of short sequence motifs, serve as single primers. Some of the most frequently used are $(GTC)_5$, $(GTG)_5$, $(GACA)_4$ etc. If repeated microsatellites are located within an amplifiable distance from one another, the inter-repeat sequences are amplified and a specific banding profile can be seen after electrophoretic run. The banding profile varies from species-to-species making this tool highly valuable tool for yeast characterization and genotyping. Sometimes, intraspecific strain genomic variability can also be seen. This technique, until recently, has been used successfully to amplify hypervariable repetitive DNA sequences in a wide range of animal, plant, and fungal species (Meyer et al. 1993; Meyer and Mitchell 1995; Weising et al. 1995). Microsatellite-primed PCR presents several advantages as it does not require prior sequence information. This technique differs from RAPD in the use of a higher annealing temperature (55ºC) in microsatellite analysis instead of 37ºC in RAPD. The application of a higher annealing temperature drives more specific primer hybridization that in turn consequently render a higher reproducibility (Botterel et al, 2001; Stephan et al, 2002; Dalle et al, 2003).

1.6.2. Emerging interdisciplinary techniques

BioNumerics: for fingerprint comparison and creating cluster analysis

The BioNumerics software offers an integrated platform for the analysis of PFGE, RAPD, microsatellite-primed PCR and RFLP fingerprints discussed in **section 1.6.1**. BioNumerics utilizes a database engines that allow you to store all your gel images, and generated clustering and comparison results in one database. Dendrograms of up to 20,000 database entries can be calculated for any type of experiment, based on pairwise similarity or distance values helps you to calculate similarity matrices and dendrograms in a few mouse clicks. For creating comparison and performing clustering analysis, BioNumerics offers convenient wizards that help defining new fingerprint types, choose the optimal

settings for normalization, resolution, background subtraction, smoothing, and band finding (**Figure 1.12**). BioNumerics accept gel images (file type: TIFF) of fingerprints that are first imported in and then processed for creating comparison and clustering. Processing and normalization of gel image files is done in 4 easy steps: Strips, Curves, Normalization, and Bands (**Figure 1.12**).

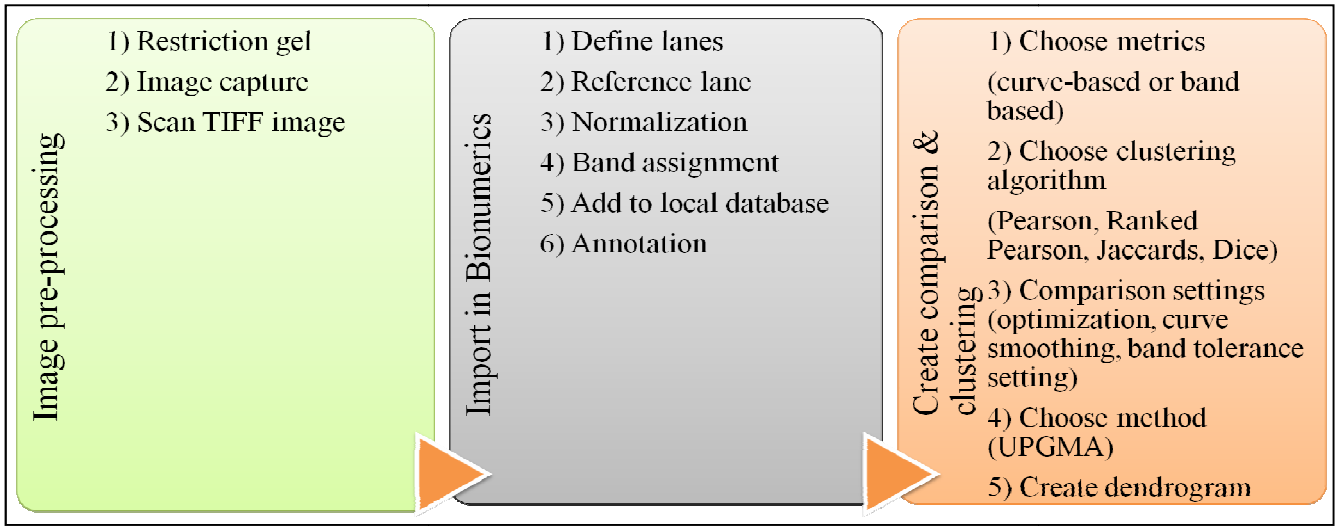


Figure 1.12. A fully automated processing workflow of BioNumerics. It showing step-wise the set-up scheme for image processing, import and creation of comparison and dendrograms.

The entire gel image pre-processing is contained in a powerful tab-based window, allowing easy access to re-edit the processing at any stage without losing any editing in another step. For creating comparison and clustering analysis, various similarity and distance coefficients are available for different data types. In general, two types of metrics are available as depicted in **Figure 1.12**: band-based: Jaccard, Dice, Jeffrey's X, Ochiai; and curve-based: Pearson Product-moment Correlation Coefficient, Ranked Pearson Coefficient.

Various cluster analysis methods can be chosen: UPGMA, Ward, Neighbor Joining, Single Linkage, and Complete Linkage. In addition to standard methods such as bootstrap analysis or cophenetic correlation, BioNumerics employs proprietary technology to assess the reliability of clusters for any clustering algorithm and data set. The method is based on resampling/permutation

techniques operating at the data level or at the similarity level and is designed as a framework encompassing all available clustering algorithms in BioNumerics. Generally, the significance of created clusters is generally assessed by calculating error flags and cophenetic correlations.

Phenomics and phenotypic microarray

Phenotype MicroArray (PM) is a high-throughput technology and represents the third major technology, alongside DNA Microarrays and Proteomic Technologies that is needed in the genomic era of research and drug development. Just as DNA Microarrays and Proteomic Technologies have made it possible to assay the level of thousands of genes or proteins all at once, Phenotype MicroArrays make it possible to quantitate simultaneously a large number of cellular and metabolic phenotypes all at once (~ 2000 phenotypes in a single experiments). Through comprehensive and precise quantitation of phenotypes, researchers are able to obtain an unbiased perspective of the effect on cells of genetic differences, environmental change, and exposure to drugs and chemicals.

Important Applications of PM technology fall into three broad categories. 1) PM could be used for testing cell lines exposed to drugs or other chemicals. Potential applications include evaluation of new drug candidates and toxicological studies (Zhang and Biswas, 2009). 2) Similarly, PM could also be used for testing cell lines with genetic differences especially in functional genomics studies and in cell line improvements (Bochner et al., 2001; Ito et al., 2005). 3) Direct testing of cell lines for optimal growth, sporulation, germination conditions as well as for optimal enzymatic activity and optimal production of secondary metabolites (Borglin et al., 2012; Lei and Bochner, 2013; Vital-Lopez et al., 2013).

Flow cytometry: for cell cycle analysis and ploidy estimation

The genome size (C value) of unicellular eukaryote can be estimated by biochemical assays, which determine the phosphate content of DNA isolated from a defined number of cells (Shapiro, 1970); analysis of reassociation kinetics, using hydroxyl apatite chromatography to separate single-stranded and double-stranded DNA (dsDNA) (Britten and Kohne, 1986); pulsed-field gel electrophoresis (PFGE), in which the lengths of whole chromosomes are measured directly (Johnston and Mortimer, 1986; Vaughan-Martini et al., 1993); more recently by Real-time PCR (Wilhelm et al., 2003). Ploidy level can be obtained comparing the haplotype size generated by PFGE with data obtained by microfluorimetry or by flow cytometry-based methods (FCM). Nowadays, FCM are widely used in cell biology and are known for providing high-resolution and multi-parametric quantitative measurement of light scatter and fluorescence emission properties of hundreds of thousands of individual cells in each analyzed sample.

FCM is used routinely both in research labs to study cells with normal and abnormal morphological, cytological and physiological characteristics and in clinical labs to diagnose and monitor human disease as well as response to therapy and vaccination. In a typical FCM analysis, cells are stained with fluorochrome-conjugated antibodies that bind to the cell surface and intracellular molecules. Within the flow cytometer, cells are passed sequentially through laser beams that excite the fluorochromes. The emitted light, which is proportional to the antigen density, is then measured. The latest flow cytometers can analyze 20 different characteristics for individual cells in complex mixtures (Baumgarth et al., 2000), and recently developed mass spectrophotometry–based cytometers could dramatically increase this number (Tanner et al., 2008; Bendall et al., 2012; Newell et al., 2012). Recently FCM-based methods were successfully employed to determine the genome size and ploidy

level of multibudding yeast *Paracoccoioides brasiliensis* (Almeida et al., 2007), spoilage yeast *Z. bailii* (Rodrigues et al., 2003; Dato et al., 2008), and *Z. rouxii* (Solieri et al., 2008a).

Chapter 2:

Phylogenetic analysis of *Zygosaccharomyces* strains and delineation of a novel yeast species, *Zygosaccharomyces sapae*

2.1. Introduction

Zygosaccharomyces is a group of ascomycetous yeasts that belong to family Saccharomycetaceae. For reasons that it buds (pullulates), fosters alcoholic fermentation, and sporulates, *Zygosaccharomyces* yeasts were firstly assigned to *Saccharomyces*. However, in *Zygosaccharomyces* yeasts sporulation precedes immediately after conjugation, typical of this genus. Since, no known *Saccharomyces* yeast display this character, Barker (1901) proposed a separate genus, so called as *Zygo-saccharmyces*, for these species of yeast. However, some authors argued taxonomic separation of *Zygosaccharomyces* from genus *Saccharomyces* and consequently referred *Zygosaccharomyces* as a subgenus of *Saccharomyces* (Stelling-Dekker, 1931). As a result, misnomers such as *Saccharomyces rouxii* still exist in literature. Finally, Yarrow (1984) revived the concept of this genus and *Zygosaccharomyces* was reclassified to its current name by Barnett et al. (1983). The type material of *Zygosaccharomyces barkeri*, which was designated as the type species of *Zygosaccharomyces* (Barker, 1901), is no longer exists. However, it has been believed to be the same species as *Z. rouxii* (Yarrow, 1984). In view of the absence of an extant type species, *Zygosaccharomyces rouxii* (Boutroux) Yarrow is proposed as the neotype species of the genus *Zygosaccharomyces* with type material represented by the culture CBS 732^T (NRRL Y-229), isolated from grape must in Italy.

In the fourth edition of yeast taxonomy (Kurtzman and Fell, 1998), genus *Zygosaccharomyces* accommodates nine species viz. *Z. bailii, Z. bisporus*, *Z. cidri*, *Z. fermentati*, *Z. florentinus*, *Z. mellis*, *Z. microellipsoides*, *Z. mrakii* and *Z. rouxii* (Kurtzman, 1998). Previous studies on some species of *Zygosaccharomyces* revealed that *Z. microellipsoides and Z. mrakii* are remarkably similar to *Torulaspora delbrueckii* in terms of phylogeny, both displaying 99.8% 18S rRNA sequence similarity (James et al., 1994). Based upon phylogenetic studies, 18S ribosomal RNA gene sequencing and after comparative sequence analysis of gene sequences, species of genus are grouped in different subdivisions. *Z. bailii*, *Z. bisporus*, *Z. rouxii* and *Z. mellis* were grouped in one division; *Z. cidri* and *Z. fermentati* formed a distinct species pair, as did *Z. microellipsoides* and *Z. mrakii*; and *Z. florentinus* did not display any specific relationship with other species was put into a separate line (James et al., 1994).

The genus *Zygosaccharomyces* (Barker, 1901) was a polyphyletic group of ascomycetous yeasts that was recently divided into four phylogenetically circumscribed genera, *Zygosaccharomyces*, *Zygotorulaspora*, *Torulaspora* and *Lachancea* (Kurtzman and Robnett, 2003; Kurtzman, 2003). Currently the genus comprises six species, viz. *Z. bailii*, *Z. bisporus*, *Z. kombuchaensis*, *Z. lentus*, *Z. mellis* and *Z. rouxii* (James and Stratford, 2011). Other recently identified *Zygosaccharomyces* species includes *Z. machadoi* (Rosa and Lachance, 2005), *Z. gambellarensis* (Torriani et al., 2011), *Z. siamensis* (Saksinchai et al., 2012), *Z. sapae* (Solieri et al., 2013a), *Z. parabailii* and *Z. pseudobailii* (Suh et al., 2013).

These species differ from that those of genus of *Saccharomyces* in terms of their vegetative growth, in which sexual conjugation between cells or a cell and its bud precedes sporulation. Species of the genus *Zyosaccharomyces* are often associated with spoilage and food-grade fermentation (Deak and Beuchat, 1996). Their ability to withstand high sugar and high osmotic conditions enables them

to ferment sugars present in food and beverage at low pH conditions to promote spoilage (Snowdon and Cliver, 1996; Loureiro and Malfeito-Ferreira, 2003). As compared to species of the genus *Saccharomyces*, these species exhibit tolerance to various osmotic and salt stresses significantly different (Dakal et al., 2014). In particular, *Z. rouxii* is one of the important spoilage yeast related to sweet foods and beverages. The most common example is fermentative spoilage of honey by *Z. rouxii* (Snowdon and Cliver, 1996). In contrast, the beneficial role of species of genus *Zygosaccharomyces* has been constantly implicated in few fermented food such as soy sauces, miso paste, wine and traditional balsamic vinegar (Suezawa et al., 2008; Solieri and Giudici, 2008). Traditional balsamic vinegar (TBV) represents such an osmotic niche. TBV is a condiment typical of the Italian Northern provinces of Modena and Reggio Emilia and obtained from very sweet, cooked grape must through a two-step process of spontaneous alcoholic fermentation and subsequent acetic oxidation (Solieri and Giudici, 2008). Different yeasts with diverse phylogenetic affiliations have been recovered from TBV, yet this ecosystem represents a highly unexplored habitat for yeast species (Solieri et al., 2006). During a survey of yeasts associated with alcoholic fermentation of TBV, fourteen *Zygosaccharomyces* isolates were detected based on their unique nucleotide sequences in the 26S rRNA gene D1/D2 domains (Solieri et al., 2006). In this report, a novel species is proposed to accommodate these strains. Two TBV samples (named as A and B) were collected from different factories located in the provinces of Modena and Reggio Emilia, Italy, in May and June 2004. Details of the sample processing, isolation and maintenance of the yeasts have been described previously by Solieri et al. (2006). During a survey of yeasts associated with alcoholic fermentation of TBV, fourteen *Zygosaccharomyces* isolates were identified from two fermenting TBV samples based on their unique nucleotide sequences in the 26S rDNA D1-D2 domain (Solieri et al., 2006). In this work, we carried out phylogenetic analysis (based on 5.8S ITS gene and 26S D1-D2 domain), morphological characterization and metabolic and physiological screening of a number of

Zygosaccharomyces strains. Out of the strains analyzed, the fourteen *Zygosaccharomyces* isolates from TBV samples were found to represent a novel yeast species. We propose to accommodate these isolated strains as a novel species, *Zygosaccharomyces sapae* sp. nov. with strain ABT301^T (=CBS 12607^T =MUCL 54092^T) as the type strain. In order to assign proper taxonomic position to the isolated strains, a polyphasic taxonomic approach comprising a number of genotypic, Phylogenetic, morphological and physiological screening studies were performed.

2.2. Materials and methods

2.2.1. Strains, medium and culture conditions

The strains were maintained on solid media (YPDA) containing 1% yeast extract, 2% peptone, 2% glucose and 2% agar on slants in test tubes. The strains were routinely grown in rich medium (YPD) containing 1% yeast extract, 2% peptone and 2% glucose at 28°C.

Table 2.1. Strains used in this study, their source of isolation, mating behavior and genetic status. * postulated hybrid status

Species	Strain Code(s)	Source of Isolation/Country	Mating type	Genetic status	References
CBS 732^T	ATCC 2623^T NRRL Y-229^T	Grape must, Italy	*MATa*/homothallic	Haploid	Yarrow, 1984
ATCC 42981	-	Miso, Japan	-	Allodiploid	Kiuchi et al. (1980)
CBS 4837	NCYC 1682	Miso, Japan	*MATa*/heterothallic	Hybrid*	Wickerham and Burton, 1960; James et al. (2005)
CBS 4838	-	Miso, Japan	*MATα*/heterothallic	Hybrid*	Wickerham and Burton, 1960
ABT301^T-306; ABT601-608	CBS 12607^T MUCL 54092^T	TBV, Italy		Diploid	Solieri et al., 2007
NCYC3042	-	Soft drink, UK		Hybrid*	James et al., 2005
CBS 736^T	NBRC 1615^T MUCL 39114^T	Honey, USA		-	
OUT7136	-			-	Provided by Prof. Y. Kaneko
OUT7140	-			-	
DBVPG 6920	DBVPG 6920	Salad dressing, USA		-	
NBRC0505		Shoyu yeast, Japan		-	Suezawa et al., 2008
NBRC0506		Shoyu yeast, Japan		-	Suezawa et al., 2008
NBRC0521		Shoyu-moromi, Japan		-	Suezawa et al., 2008
NBRC0525		Miso, Japan		-	Suezawa et al., 2008
NBRC0845		Mash of Tamari soya, Japan		-	Suezawa et al., 2008

2.2.2. Genomic DNA extraction and DNA manipulation

Genomic DNA (gDNA) isolation procedure utilizes physical disruption of the yeast cell wall using small glass beads and Phenol-Chloroform based extraction protocol as described by Hoffman and Winston (1987). For extraction of gDNA overnight liquid cultures (in YPD medium) of the strains were used. The cells were collected by centrifugation for 5 mins at 8,000 rpm at 4°C and resuspended in 200µL Lysis buffer, 400µL Phenol-Chloroform-Isoamyl alcohol (25:24:1) and 200µL TE buffer. Genomic DNA isolation procedure utilizes physical disruption of the yeast cell wall using small glass beads (600mg) by vortexing cells for 4 min twice with a 2 min pause during which the tubes were placed in ice. Tubes were centrifuged at 14,000 rpm for 5 min at 4°C to allow separation of aqueous and organic phase. The DNA in the aqueous phase (~500µL) is subsequently purified using a mixture of 500 µL of phenol-chloroform (24:1) solution by centrifugation at 14,000 rpm for 5 min at 4°C. The aqueous phase is pipette out and the DNA in the aqueous phase is precipitated with same volume of Isopropanol for 30 min at -20°C. The tubes were kept at room temperature for 10-15 min and thereafter spun in centrifuge at 14,000 rpm for 15-20 min to obtain DNA pellets. gDNA pellets obtained after extraction procedure were washed with 70% Ethanol, air dried for 20 min, resuspended in 35µL sterile double distilled water containing 1.5µL of RNAse (10mg/mL) and incubated at 37°C for 1hr in water bath.

The quantity of extracted DNA was measured using Nanodrop spectrophotometer (ND-1000). Approximately, 1500-3500 ng/µg of genomic DNA can be extracted from the 4mL of an overnight culture. The ratio of absorbance at 260nm and 280nm is used to assess the purity of DNA. 260/230 ratio is used as a secondary measure of nucleic acid purity. If the ratio is appreciably lower than expected, it indicates the presence of contaminants, such as EDTA, carbohydrates and phenol, all having the absorbance near 230nm. The extracted DNA is separated on 1% Agarose gel with

electrophoresis run at 120V for 45 min. Finally, the extracted DNAs were stored at -20°C. Resultant gDNAs after appropriate dilution were used for performing genotyping and in standard 5.8S ITS and 26S D1-D2 PCR amplification reactions.

2.2.3. Strain genotyping

Microsatellite-primed PCR (MSP-PCR) fingerprinting and random amplified of polymorphic DNA (RAPD) analysis were carried out with primers $(GTG)_5$ (5'-GTGGTGGTGGTGGTG-3') and M13 (5'-GAGGGTGGCGGTTCT-3'), respectively (Baleiras Couto et al., 1996). The reproducibility of the $(GTG)_5$-PCR and RAPD fingerprinting patterns were evaluated by three independent DNA preparations from six randomly selected strains (data not shown). The PCR products were separated on 1.8% agarose gels with electrophoretic run for 150 min at 80V, visualized under UV, and photographed using BioDoc Analyser (Biometra, Göttingen, Germany). Gel images were processed using the BioNumerics (version 6.0; Applied Maths). The dendrogram was generated using the Pearson's correlation similarity coefficient and the unweighted-pair group method using the arithmetic means (UPGMA) clustering method, with 1% optimization.

2.2.4. PCR amplification of 5.8S ITS region and D1-D2 domain of 26S rDNA

The entire internal transcribed spacer (ITS) domains (ITS 1, ITS 2 and the intervening 5.8S rRNA gene) were amplified with primers ITS1 and ITS4 (White et al., 1990). For the amplification of 5.8S ITS region a pair of universal primers ITS1 and ITS4 (where ITS1 (5'-TCC GTA GGT GAA CCT GCG G-3'), which hybridizes at the end of 18S rDNA, and ITS4 (5'-TCC TCC GCT TAT TGA TAT GC-3'), which hybridizes at the beginning of 28S rDNA) were used. The 25µL PCR mixture contained 2.5µL of PCR buffer (10X), 2µL of 2.5mM of deoxynucleoside triphosphates mixture,

1.5µL of 5pmol of each primer, 0.125µL of 5U/µL of r*Taq* DNA polymerase (TaKaRa Bio. Inc., Japan) and 1µL of 250ng/µL template genomic DNA. Reactions involved 1 cycle at 95°C for 5 min, followed by 35 cycles with a denaturation step at 95°C for 30 s, an annealing step at 55°C for 1 min, and an extension step at 72°C for 1 min, followed by 1 cycle at 72°C for 10 min.

Domains 1 and 2 of the 26S rRNA gene (D1-D2) were amplified using primers NL1 and NL4 (O'Donnell, 1993). For the amplification of D1-D2 domain of 26S rDNA a pair of universal primers NL1 and NL4 (NL1 5'-GCATATCAATAAGCGGAGGAAAAG-3' and NL4 5'-GGTCCGTGTTTCAAGACGG-3') were used. The 25µL PCR mixture contained 2.5µL of PCR buffer (10X), 2µL of 2.5mM of deoxynucleoside triphosphates mixture, 1.5µL of 5pmol of each primer, 0.125µL of 5U/µL of r*Taq* DNA polymerase (TaKaRa Bio. Inc., Japan) and 1µL of 250ng/µL template genomic DNA. Reactions involved 1 cycle at 95°C for 5 min, followed by 35 cycles with a denaturation step at 95°C for 30 sec, an annealing step at 55°C for 1 min, and an extension step at 72°C for 1 min, followed by 1 cycle at 72°C for 10 min.

PCR products were purified by the DNA Clean and Concentrate TM-500 Kit (Zymo Research), following the manufacturer's instructions, and delivered to MWG Sequencing Service (Germany) for sequencing. Aliquots (5µL) of each amplified product were electrophoretically separated in a 0.8% agarose gel in 0.5X Trisborate-EDTA buffer and visualized using ethidium bromide under UV illumination. GenRuler 100bp Plus molecular weight marker was included in each run (MBI Fermentas).

2.2.5. Sequencing and phylogenetic analysis

To confirm the identification of *Zygosaccharomyces* strains at species level, the 5.8S ITS gene and D1-D2 domain of 26S rDNA was send for sequencing to MWG Biotech, Germany after

purification and quantification by Nanodrop of the amplicons generated by PCR amplification of the cloned genes. The concentration of the amplicons send for sequencing was in accordance with the MWG guidelines.

Sequences of 5.8S ITS gene and D1-D2 domain of 26S rDNA were edited and assembled using SeqMan software (DNASTAR) and BLASTed against the GenBank database to retrieve sequences of the closest relatives. The sequences were aligned using the Clustal-X software package (Thompson et al., 1997) and phylogenetic trees were inferred from trimmed alignments by using the neighbour-joining (NJ) method with 1000 bootstrap iterations (Saitou and Nei, 1987). Sequences from *Wickerhamomyces anomalus* CBS 5759^{T} (GenBank accession no. U74592) and *Z. bailii* CBS 680^{T} (AY046191) were used as an outgroup in the D1-D2 and ITS NJ-trees, respectively.

2.2.6. Morphological characterization

Cell morphology was examined after incubation on YPD (2.0% yeast extract, 2.0% peptone, 2.0% glucose, 1.5% agar) medium for three days at 27°C by using a phase-contrast microscope (Axioskop 40; Carl Zeiss) connected to a digital camera (Optech). Asci and ascospore production were examined by growing the cultures alone or mixed with either the tester mating types a and a, CBS 4837 and CBS 4838, respectively, of *Z. rouxii* (mating type designation adopted from Wickerham and Burton, 1960) at 27°C on YPD and 5% ME (Oxoid, Milan) with and without 1.5% agar, as well as on Gorodkowa agar (Kurtzman et al., 2011). Ascospore formation was observed at regular intervals for up to one month.

2.2.7. Physiological and metabolic screening

Physiological characterization of isolates was performed according to standard methods (Kurtzman et al., 2011). For physiological and metabolic screening seven fermentation, nineteen

assimilation and seven biochemical tests were performed along with growth in presence of antibiotic (0.1% Cycloheximide), tolerance to 1% acetic acid, growth at different temperature (8, 16, 28 and 37°C), and growth on different osmotic medium (60% Glucose-0.5% Yeast Extract; 16% NaCl-5% Glucose; Yeast Nitrogen Base-3% Glycerol-1% NaCl).

Fermentation of glucose, fructose, maltose, lactose, galactose, sucrose and trehalose was tested. Assimilation of glucose, fructose, maltolse, galactose, sucrose, trehalose, D-xylose, ribose, D-arabinose, L-rhamnose, melibiose, raffinose, D-sorbitol, D-mannitol, inositol, erythritol, ethanol, glycerol was tested. Biochemical tests performed include growth in presence of vitamin-free, sodium nitrate, potassium nitrate, L-lysine, ethylamine-hydrochloride, cadaverine and ammonium sulphate were also performed. For morphological characterization of vegetative and sexual stages such as production of pseudohyphae, bud formation and production of ascospores, culture were grown in 5% malt extract and Gorodkowa medium.

For fermentation tests, yeast extract concentrate (1.5-fold concentrated) and 6% concentrated solution of different sugars to be tested were prepared, except for Raffinose which was prepared at 12%. Final fermentation test medium was prepared in the test tube (dimension 16x180mm) containing Durham insert by adding 5.0mL of yeast extract concentrate (1.5-fold concentrated) and 2.5mL of any one sugar solution (concentration 6%) to be tested. The resulting 7.5mL of fermentation medium was inoculated with a loopful of the cell suspension from a 24-48 hour YPD liquid pre-culture of different strains mentioned in **Table 2.1**. The tubes are incubated at 28°C. The growth from these tubes was routinely monitored after inoculation at 3rd day, after 1 week, 2 weeks, and 3 weeks. The readings were recorded as per guidelines described by Yarrow (1998).

Growth medium for carbon assimilation tests in liquid medium using Yeast Nitrogen Base. A 10-fold concentrated solution of yeast nitrogen base medium is prepared by dissolving 6.7 g of Bacto

Yeast Nitrogen Base in 100 ml of demineralized water and the amount of the carbon compound equivalent to 5 g of glucose (with warming if necessary). Tubes of assimilation test medium were prepared by aseptically pipetting 0.5 ml of the YNb-sugar concentrate into 4.5 ml of sterile water in a 16x180mm test tube. Tubes with assimilation medium were inoculated with a loopful of the cell suspension from a 24-48 hour YPD liquid pre-culture, and the tubes are incubated at 28°C. The growth from these tubes was routinely monitored after inoculation at 3rd day, after 1 week, 2 weeks, and 3 weeks. The readings were recorded as per guidelines described by Yarrow (1998).

The ability to grow at 8, 16, 28 and 37°C was tested on YPD medium both in tubes under static conditions and in Erlenmayer flasks under shaking conditions (180 r.p.m.) for one week and the growth was routinely monitored spectrophotometrically at 600 nm (Jasco V-550 UV/VIS spectrophotometer) (1 OD_{600} = 2.8 x 10^7 c.f.u. per ml).

The ability to grow in the presence of different osmotic medium (60% Glucose-0.5% Yeast Extract; 16% NaCl-5% Glucose; Yeast Nitrogen Base-3% Glycerol-1% NaCl) was tested in Erlenmayer flasks under shaking conditions (180 r.p.m.) for one week and the growth was routinely monitored spectrophotometrically at 600 nm (Jasco V-550 UV/VIS spectrophotometer) (1 OD_{600} = 2.8 x 10^7 c.f.u. per ml).

2.3. Results and discussion

2.3.1. Genotyping of fourteen TBV isolates

Initially, MSP-PCR fingerprinting with primer $(GTG)_5$ was used to identify clonal relationships among 14 TBV isolates and to select strains for species delineation. Microsatellite primer $(GTG)_5$ generated complex banding patterns, with eight electrophoretic bands, ranging in size

from 300 to 2810 bp (**refer Table 6.6, Chapter** 6). Fourteen TBV isolates possessed similar band patterns and grouped into a homogeneous cluster separated from those of the reference strains. At a similarity threshold of 94 %, the TBV isolates could be further subdivided into two subclusters, namely A and B (**Figure 2.1**). This grouping coincided with their source of isolation: subcluster A included 6 isolates, all collected from sample A; while B comprised 9 isolates retrieved from sample B. One strain from each subcluster, respectively, ABT301^{T} and ABT601, was selected for further studies.

Figure 2.1. UPGMA dendrogram based on Pearson's coefficient for MSP-PCR fingerprints of 14 TBV isolates.

2.3.2. Sequence analysis of phylogenetic markers

Sequencing of the D1/D2 domains produced a unique sequence for both the TBV strains. Comparison of the similarity of D1/D2 domains of strains ABT301^{T} and ABT601 with *Zygosaccharomyces* type strains from the GenBank database revealed that the novel strains differed

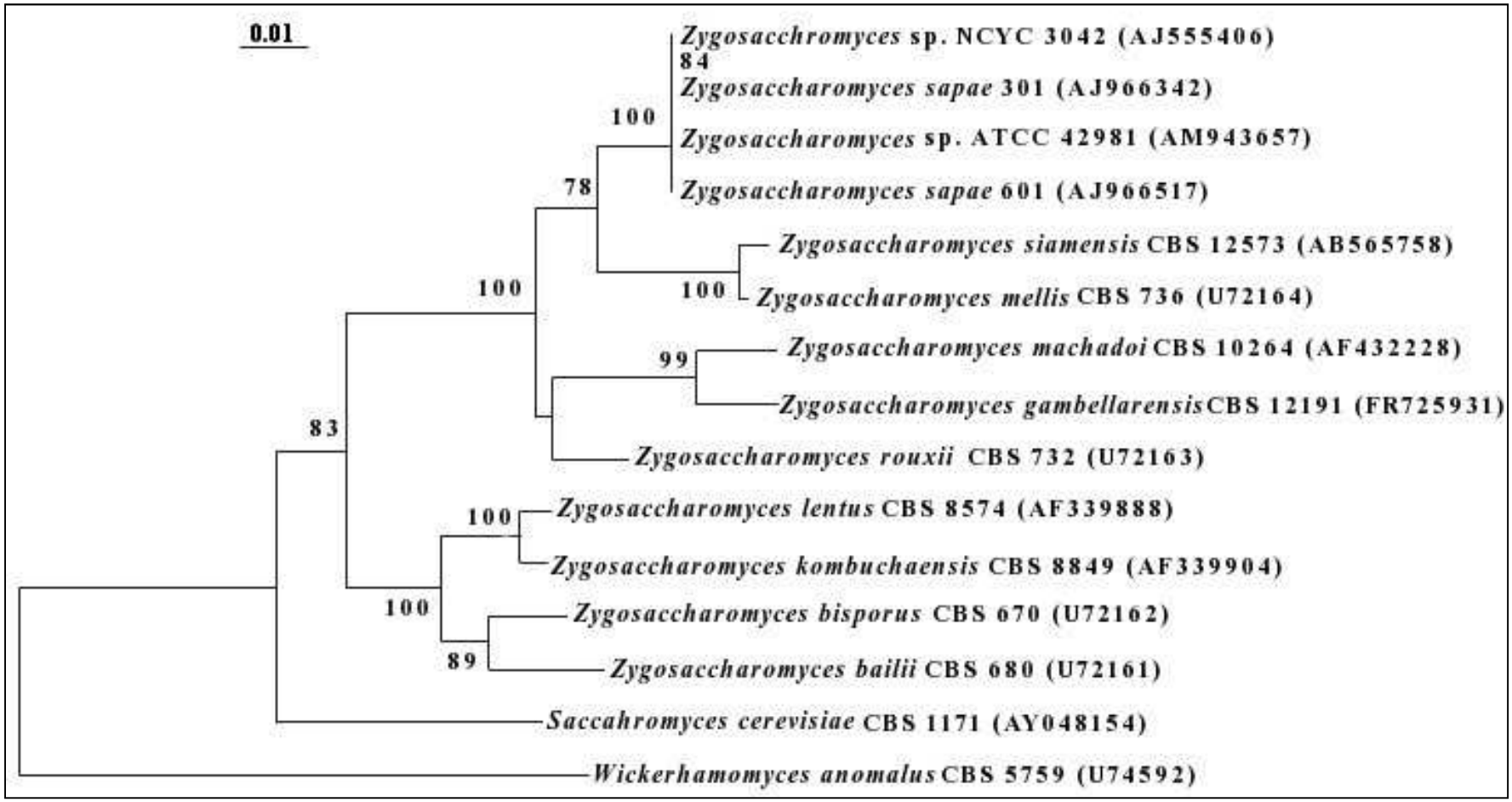


Figure 2.2. NJ tree based on the 26S D1-D2 sequences of *Z. sapae* sp. nov. and other *Zygosaccharomyces* species. Percentage bootstrap values were obtained from 1000 iterations (values ,50% not shown). The tree was rooted by inclusion of the sequences for *Wickerhamomyces anomalus* as the outgroup. Sequences were retrieved from the DDBJ/ GenBank/EMBL databases under the indicated accession numbers. Bar, 0.01 substitutions per nucleotide.

from *Z. rouxii* by 2.5% (14 mismatches and one indel in 567 nt) and from *Z. mellis* by 4.4% (19 mismatches and six indels in 572 nt). Phylogenetic analysis of the D1-D2 domain sequences placed strains ABT301^T and ABT601 in a distinct and supported clade (78% bootstrap) within the genus *Zygosaccharomyces*, with *Z. mellis*, *Z. rouxii* and three recently proposed species, *Z. siamensis* (Saksinchai et al., 2012), *Z. machadoi* (Rosa and Lachance, 2005) and *Z. gambellarensis* (Torriani et al., 2011), being the closest relatives (**Figure 2.2**). No differences were detected between the NJ and maximum-parsimony methods, particularly as regards the position of strains ABT301^T and ABT601 (data not shown). Considering the guidelines given by Kurtzman and Robnett (1998), i.e. distinct species usually exhibit 1% or more sequence differences in the D1-D2 domains, the data allowed the designation of strains ABT301^T and ABT601 as representing a novel species in the *Zygosaccharomyces clade*, for which the name

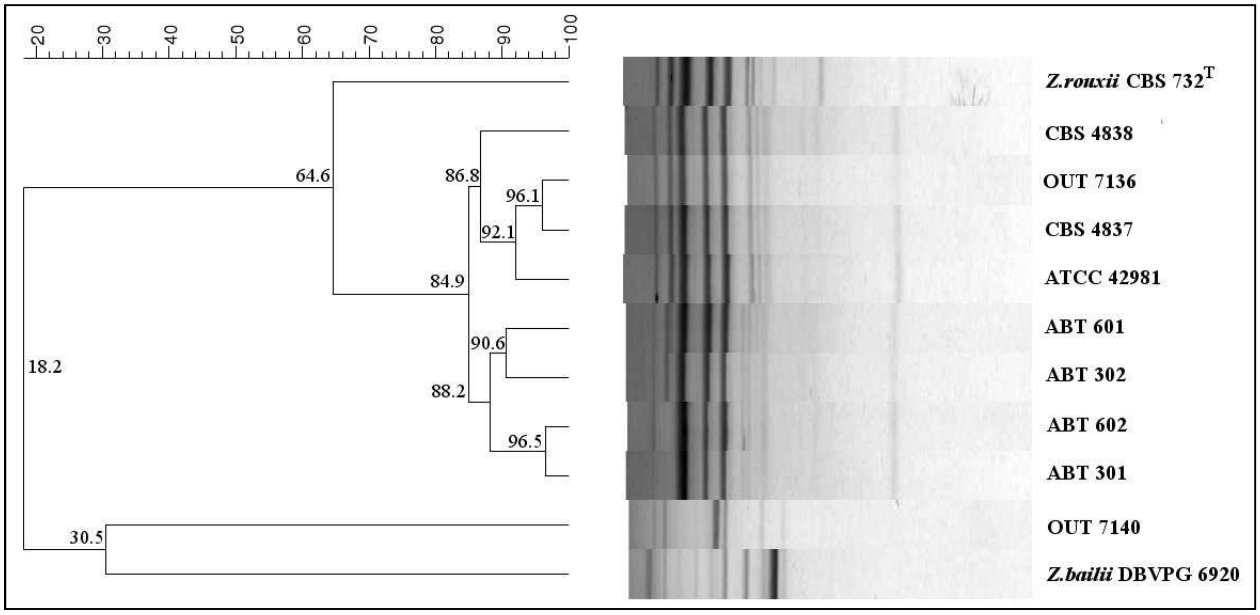


Figure 2.3. Pearson's coefficient based UPGMA dendrogram of ABT301^T and phylogenetic neighbors (MSP-PCR). (bar, expressed as a percentage similarity value).

Zygosaccharomyces sapae sp. nov. is proposed. A BLAST search of the GenBank database showed that *Z. sapae* sp. nov. was identical to strains of uncertain taxonomical affiliation with regard to 26S rRNA gene D1-D2 domains (**Figure 2.2**). In particular, strain NCYC 3042 (isolated from sugar) and has been deposited in culture collection as a postulated hybrid and recently suggested as belonging to a novel species with the provisional name *Zygosaccharomyces pseudorouxii* James et al. (2005) nom. inval. The name is not validly published as the description of the species was based on a single isolate, without a Latin diagnosis. Based on multi-gene genealogy, James et al. (2005) were actually unable to clarify the taxonomic status of strain NCYC 3042. Strains ATCC 42981 and CBS 4837 were isolated from Japanese salty miso paste and described as allodiploid yeasts with a p-subgenome, supposedly retrieved from *Z. pseudorouxii* nom. inval., and a t-subgenome identical to *Z. rouxii* (James et al., 2005; Gordon and Wolfe, 2008; Suezawa et al., 2008). On the basis of the D1/D2 sequences, strain NCYC 3042 may be considered to be conspecific with *Z. sapae* sp. nov. To elucidate the relationships between *Z. sapae* sp. nov. and the *Zygosaccharomyces* sp. strains NCYC 3042, ATCC 42981 and CBS 4837, these strains were included in this study.

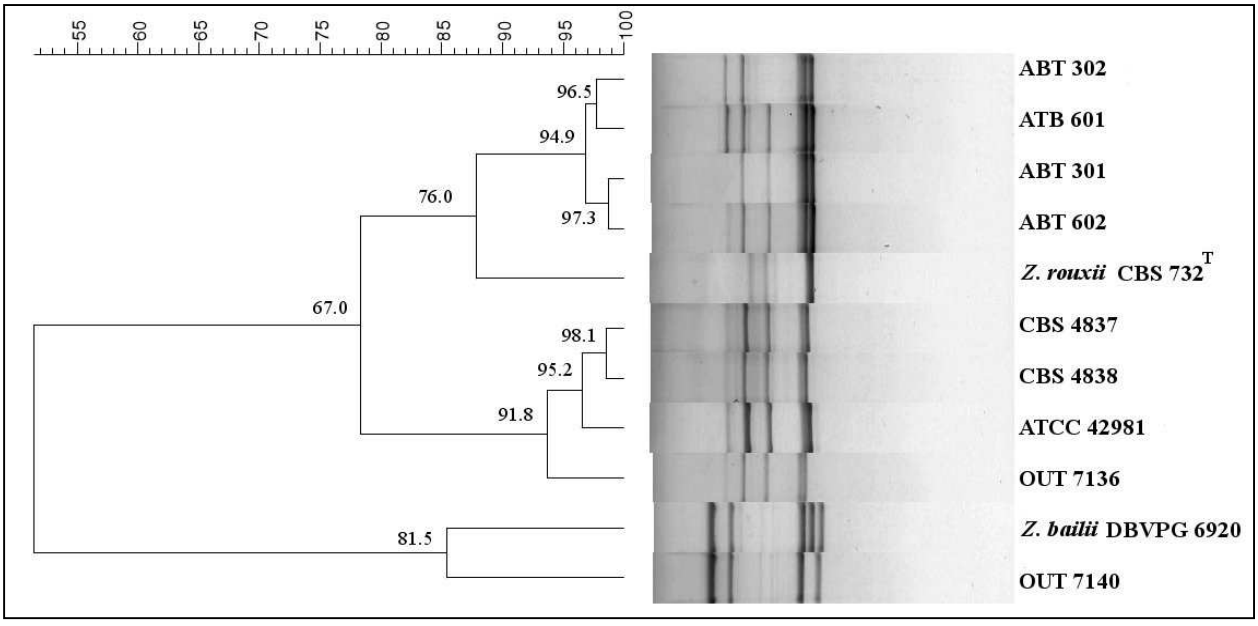


Figure 2.4. Pearson's coefficient based UPGMA dendrogram of ABT301^T and phylogenetic neighbors (M13-RAPD-PCR) (bar, expressed as a percentage similarity value).

The $(GTG)_5$-fingerprint analysis showed that the allodiploid strains ATCC 42981 and CBS 4837 were genotypically separated from strains of *Z. sapae* sp. nov. ($r \leq 86$ %), as well as from NCYC 3042 ($r \leq 78$ %) (**Figure 2.3**). $(GTG)_5$-fingerprints of strains ATCC 42981 and CBS 4837 were highly similar to each other and close to that of *Z. rouxii* CBS 732^T. These results were consistent with those obtained by M13-RAPD, which grouped the novel strains and the postulated hybrids ATCC 42981 and CBS 4837 into two homogeneous and distinct clusters (**Figure 2.4**). The position of strain NCYC 3042 was incongruent in $(GTG)_5$ and M13-RAPD dendrograms. *Zygosaccharomyces* sp. NCYC 3042 stood apart based on $(GTG)_5$-PCR profiling, whereas it was placed in a cluster with the postulated hybrid strains ATCC 42981 and CBS 4837, based on M13-RAPD patterns. The ITS domain sequencing of strains $ABT301^T$ and ABT601 revealed that three non-orthologous ITS sequences (arbitrarily referred to as copies 1–3) coexisted within the genome of a single strain (Solieri et al., 2007), whereas two ITS sequences were found from strain ATCC 42981 (Gordon and Wolfe, 2008). An increasing number of examples have been recently reported on intragenomic polymorphism in ITS sequences within the same strain of fungal and yeast species (Wang and Yao, 2005; Fell et al., 2007; Kageyama et al., 2007; Jacques et al., 2010; Connell et al., 2010; Saksinchai et al., 2012). *Z. sapae* sp. nov. strain $ABT301^T$ differed from *Zygosaccharomyces* sp. strain NCYC 3042 in the restriction profiles of the 5.8S ITS region (**refer Table 6.2; Chapter 6**). Moreover, cloning and sequencing of the ITS regions identified a unique sequence within the NCYC 3042 genome. This sequence showed 98.3% similarity to that of *Z. sapae* sp. nov. ITS copy 2 and to the p-haplotype of strain ATCC 42981. Sequence differences included one transversion and two indels in 647 nt located in the 5.8S rRNA gene. The NCYC 3042 ITS sequence differed significantly from *Z. sapae* sp. nov. ITS copy 3 by 19.5% in 690 nt: 38 transitions, 44 transversions and seven indels were mainly located in the ITS2 domain. The greatest difference (34.9 %) was found between strain NCYC 3042 and *Z. sapae* sp. nov. ITS copy 1, including 37 transitions, 32 transversions and 20 indels in 703 nt. The

majority of these differences were scattered throughout ITS1 and ITS2 domains. Considering only the mismatching positions, the differences between strain NCYC 3042 and *Z. sapae* sp. nov. ITS copies 1, 2 and 3 were 0.2 %, 12.2% and 14.0 %, respectively. Phylogenetic analysis of the ITS alignment showed that ITS haplotypes were structured into two very divergent groups: one with low diversity, including ITS haplotypes more similar to that of *Z. rouxii*; one including highly divergent ITS haplotypes, which did not match ITS sequences from any recognized species of the genus *Zygosaccharomyces* (**refer Figure 3.5; Chapter 3**). Data acquired from PCR genotyping and ITS-based phylogenesis indicate that, albeit conspecific based on D1-D2 sequences, strain NCYC 3042 and allodiploid strains ATCC 42981 and CBS 4837 differed from *Z. sapae* sp. nov. in relevant genome-level properties. These results were also confirmed by Phylogenetic divergences among these strains, previously demonstrated using the mitochondrial cytochrome oxidase II gene (*COX2*) and two nuclear housekeeping markers (imidazole-glycerol-phosphate dehydratase, *HIS3*, and Na^+/H^+-antiporter, *ZrSOD2*, respectively) (Solieri et al., 2007).

2.3.3. Morphological and physiological features

Novel strains ABT301^T and ABT601 exhibited similar morphological and physiological characteristics. They reproduced asexually by multilateral budding with mother and daughter cells remaining attached to each other, resulting in chain or star-shaped multicellular clumps (**Figure 2.5b**). Pseudomycelia could form after 20 days both on YPD and Gorodkowa media at 27°C. Single colonies of *Z. sapae* sp. nov. were able to produce persistent asci with 1–4 spheroid to ovoid ascospores when growing on YPD medium after 15–20 days at 27°C (**Figure 2.5d**). The asci were conjugated.

Strains NCYC 3042, ATCC 42981 and CBS 4837 were similar to *Z. sapae* sp. nov. strains ABT301^T and ABT601 with regard to morphological properties, but they did not form pseudomycelia

in all the test conditions. According to its heterothallic status, strain CBS 4837 (mating type a) formed conjugated asci along with strain CBS 4838 (mating type α). Strain NCYC 3042 formed rare conjugated asci (including two round to oval spores per ascus) in individual culture after 5 days on 5% ME medium, whereas strain ATCC 42981 did not form ascospores in single culture, either when mixed with strain CBS 4837 or with strain CBS 4838.

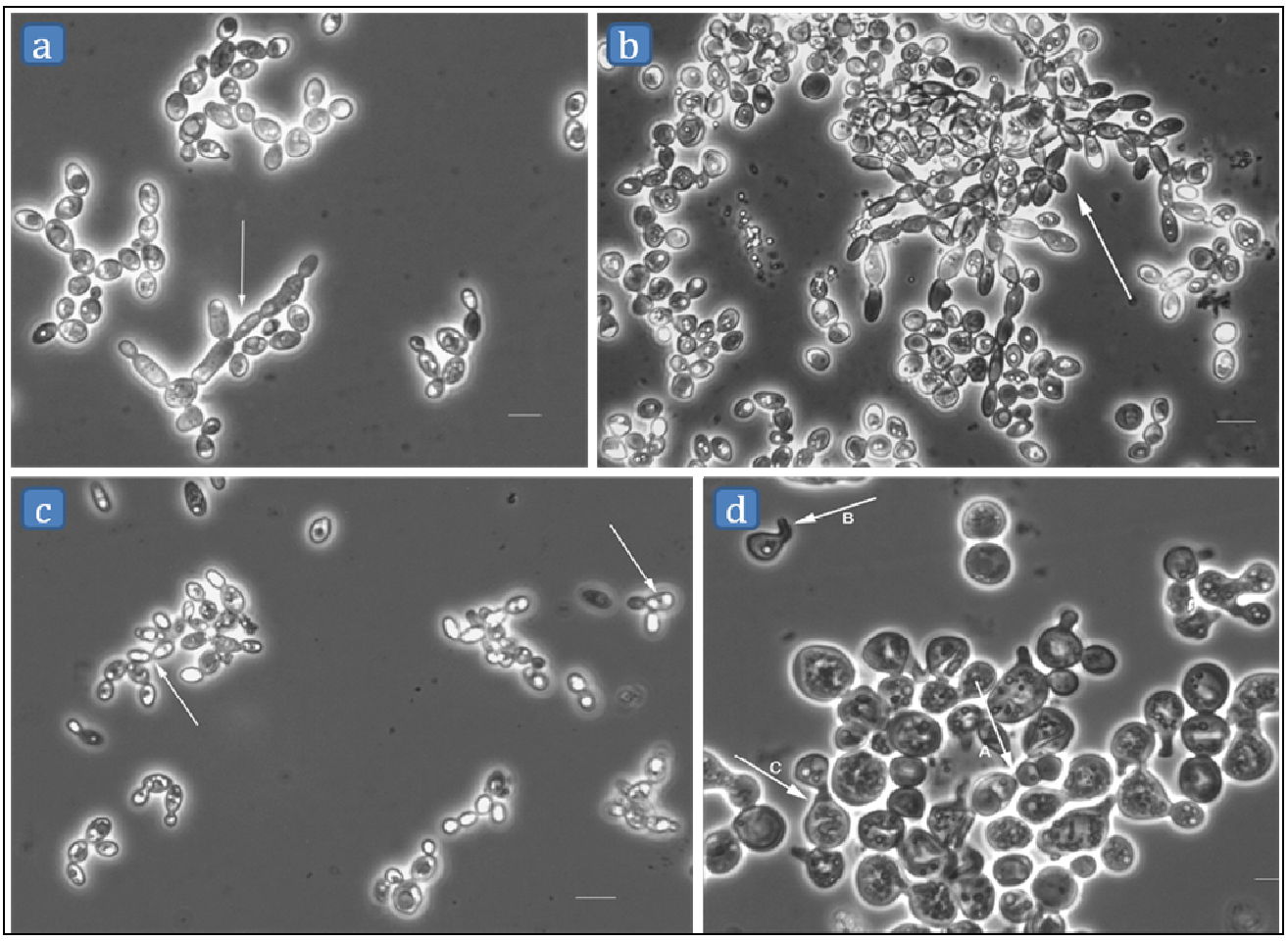


Figure 2.5. Phase-contrast micrograph. Budding cells of *Zygosaccharomyces sapae* sp. nov. ABT301^T (5CBS 12607T) grown on YPD for 6 days at 27°C show formation of pseudohyphae (a) and star-shaped dispositions (b) indicated with arrows. Phase-contrast micrographs of asci with ascospores of *Zygosaccharomyces sapae* sp. nov. ABT301^T (=CBS 12607^T) (c) and ABT601 (=CBS 12608) (d) on YPD after 20 and 15 days at 27°C where. conjugated asci (A), cell with abortive conjugation tube (B) and dumbbell-shaped configuration between cells (C) are indicated with arrows. Bars, 10.82 mm (a, b, c); 4.40 mm (d).

The physiological characteristics of *Z. sapae* sp. nov. are presented in **Table 2.2**. The novel species can be distinguished from *Zygosaccharomyces* sp. strains NCYC 3042, ATCC 42981, and CBS 4837, as well as from recognized species of the genus *Zygosaccharomyces* by numerous physiological features. *Zygosaccharomyces sapae* sp. nov. differed from *Z. mellis* by the ability to grow on 16% (w/v) NaCl plus 5% D-glucose. The novel species and strain NCYC 3042 differed in their fermentation of maltose, in their assimilation of D-mannitol, trehalose and glycerol, and, finally, in their ability to grow on 16% NaCl plus 5% D glucose. *Zygosaccharomyces sapae* sp. nov. can also be separated from *Z. rouxii* based on maltose fermentation and D-mannitol assimilation, which are positive for *Z. rouxii* but negative for the novel species (**Table 2.2**). The novel species gave a negative result in tests for the assimilation of trehalose, whereas postulated hybrids ATCC 42981 and CBS 4837 were able to use this sugar source. Finally, *Z. sapae* sp. nov. could be distinguished from *Z. siamensis* based on the ability of the latter species to assimilate erythritol (Saksinchai et al., 2012). When cultured on YPD medium under shaking conditions at 27°C, *Z. sapae* sp. nov. showed a slow growth rate compared with *Z. rouxii* CBS 732^{T}, *Z. bailii* DBVPG 6920, and *Zygosaccharomyces* sp. strains NCYC 3042, ATCC 42981 and CBS 4837 (**Figure 2.6**). The novel species showed a weak growth at 37 °C under static conditions, whilst it was fast growing at 37 °C under shaking conditions (data not shown).

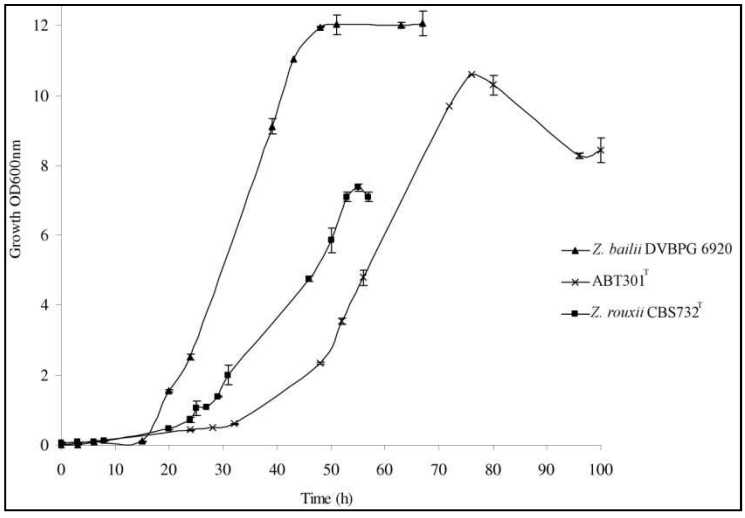


Figure 2.6. Growth curves of the ABT301^{T} and the relatives *Z. bailii* (DBVPG 6920) and *Z. rouxii* (CBS732^{T}). Growth in YPD medium in shaking conditions (180 rpm) at 28 °C. Points represent mean values out of two independent cultures and error bars (where visible) represent the standard deviation.

Fermentation of	CBS 732	ATCC 42981	CBS 4837	CBS 4838	OUT 7136	OUT 7140
Glucose	+	+	+	+	+	+
Fructose	+	+	+	+	+	+
Maltose	+	+	+	+	+	-
Lactose	-	-	-	-	-	-
Galactose	-	-	-	-	-	-
Sucrose	-	-	-	-	-	+
Trehalose	-	-	-	-	-	-
Growth on						
Glucose	+	+	+	+	+	+
Fructose	+	+	+	+	+	+
Maltose	+	+	+	+	+	+
Lactose	-	-	-	-	-	-
Galactose	+	+	+	+	+	+
Sucrose	w/l	w/l	+	+	+	+
Trehalose	+	+	+	+	l	+
Xylose	-	-	-	-	-	-
Ribose	-	-	-	-	-	-
Arabinose	-	-	-	-	-	-
Rhamnose	-	-	-	-	-	-
Melebiose	-	-	-	-	-	-
Raffinose	-	-	-	-	-	w
Sorbitol	+	w	w	w	w	+
Mannitol	+	+	+	+	+	+
Inositol	w	-	-	-	-	-
Erythritol	-	-	-	-	-	w
Ethanol	-	-	-	-	-	w
Glycerol	w	+	+	+	+	+
NH4SO4	+	+	+	+	+	+
KNO3	-	-	-	-	-	-
NaNO2	-	-	-	-	-	-
Cadaverine	+	+	+	+	+	+
L-lysine	+	+	+	+	+	+
Ethylamine-HCL	+	+	+	+	+	+
Vitamin-free	-	-	-	-	-	-
Growth on/at						
50% Dextrose	+	+	+	+	+	+
60% Dextrose	+	+	+	+	+	+
Cycloeximide resistance	-	-	-	-	-	-
1% Acetic acid tolerance	-	-	-	-	-	+
16% NaCl/5% Dextrose	+	+	+	+	+	-
3%Gly_1M NaCl	w/+	+	+	+	+	-
8°C	-	-	-	-	-	+
16°C	+	+	+	+	-	
37°	+	+	+	+	w	+

Fermentation of	ABT 301	ABT 302	ABT 601	ABT 602	NCYC 3042	ABT 401
Glucose	+	+	+	+	+	+
Fructose	+	+	+	+	+	+
Maltose	-	-	-	-	+	+
Lactose	-	w/+	-	-	-	-
Galactose	-	w/+	-	-	-	-
Sucrose	-	w/+	-	-	-	-
Trehalose	-	-	-	-	-	-
Growth on						
Glucose	+	+	+	+	+	+
Fructose	+	+	+	+	+	+
Maltose	+	+	+	+	+	+
Lactose	-	w/+	-	-	-	l
Galactose	+	+	+	+	+	+
Sucrose	w	w/l	w/l	w/l	+	+
Trehalose	-	-	-	-	-	+
Xylose	-	-	-	-	-	-
Ribose	-	w/+	-	w	-	w/+
Arabinose	-	w	-	w/+	-	-
Rhamnose	-	-	-	w/+	-	w/+
Melebiose	-	w/+	-	-	-	w
Raffinose	-	-	-	-	-	w
Sorbitol	+	w/+	w/+	w/+	w	w
Mannitol	+	+	+	+	w	+
Inositol	-	-	-	-	-	w
Erythritol	-	-	-	-	-	w
Ethanol	-	-	-	-	-	-
Glycerol	+	w/+	w/+	w/+	w	+
NH4SO4	+	+	+	+	+	+
KNO3	-	-	-	-	-	-
NaNO2	-	-	-	-	-	-
Cadaverine	+	+	+	+	+	+
L-lysine	+	+	+	+	+	+
Ethylamine-HCL	+	w	w	w	w	+
Vitamin-free	-	-	-	-	-	-
Growth on/at						
50% Dextrose	+	+	+	+	+	+
60% Dextrose	+	+	+	+	+	+
Cycloeximide resistance	-	-	-	-	-	-
1% Acetic acid tolerance	-	-	-	-	-	-
16% NaCl/5% Dextrose	+	w	w	w	-	+
3%Gly_1M NaCl	w/+	w/l	w/l	w/l	-	+
8°C	-	-	-	-	nd	-
16°C	-	-	-	-	nd	-
37°	w	w	w	w	nd	w

Fermentation of	DBVPG 6920	NBRC0505	NBRC0506	NBRC0521	NBRC0525	NBRC0845
Glucose	+	+	+	+	+	+
Fructose	+	+	+	+	+	+
Maltose	-	-	-	+	+	+
Lactose	-	-	-	-	-	-
Galactose	-	-	-	-	-	-
Sucrose	-	-	-	-	-	-
Trehalose	-	-	-	-	-	-
Growth on						
Glucose	+	+	+	+	+	+
Fructose	+	+	+	+	+	+
Maltose	+	+	+	+	+	+
Lactose	l	l	l	l	l	l
Galactose	+	+	+	+	+	+
Sucrose	+	-	-	-	-	-
Trehalose	+	+	+	+	+	+
Xylose	-	w/+	w/+	w/+	w/+	w/+
Ribose	-	w/+	w/+	w/+	w/+	w/+
Arabinose	-	-	-	-	-	-
Rhamnose	l	-	-	-	-	-
Melebiose	+	-	-	-	-	-
Raffinose	w	-	-	-	-	-
Sorbitol	+	w	w	w	w	w
Mannitol	+	+	+	+	+	+
Inositol	-	-	-	-	-	-
Erythritol	w	-	-	-	-	-
Ethanol	+	-	-	-	-	-
Glycerol	+	+	w	w	w	w
NH4SO4	+	+	+	+	+	+
KNO3	-	-	-	-	-	-
NaNO2	-	-	-	-	-	-
Cadaverine	+	+	+	+	+	+
L-lysine	+	+	+	+	+	+
Ethylamine-HCL	+	+	+	w	w	w
Vitamin-free	-	-	-	-	-	-
Growth on/at						
50% Dextrose	+	+	+	+	+	+
60% Dextrose	+	nd	nd	nd	nd	nd
Cycloeximide resistance	-	-	-	-	-	-
1% Acetic acid tolerance	+	-	-	-	-	-
16% NaCl/5% Dextrose	-	l	l	l	l	l
3%Gly_1M NaCl	-	l	l	l	l	l
8°C	+	nd	nd	nd	nd	nd
16°C		nd	nd	nd	nd	nd
37°	+	nd	nd	nd	nd	nd

Table 2.2. Physiological growth responses obtained from strains of *Z. bailli*, *Z. mellis*, *Z. rouxii*, and *Z. sapae*.

Above test responses were recorded routinely for period spanning 3 days to one month (**Supplementary Table S2.1**).

2.4. Conclusion: evolutionary implication

Yeasts belonging to the genus *Zygosaccharomyces* inhabit selective food-grade niches, including sugary and salty foods and beverages. Based on the molecular and other taxonomic characteristics mentioned above, it is concluded that the 14 yeast isolates recovered from spontaneously fermenting sweet TBV samples, represent a distinct species, namely *Z. sapae* nov. sp., described formally below. Phylogenetic analysis based on D1-D2 domains suggested that *Zygosaccharomyces* sp. strain NCYC 3042 could be conspecific with the novel strains, but PCR-genotyping, ITS phylogenetics and physiological profiling highlighted marked differences resulting in incongruence with regard to the taxonomical status of the latter strain. For instance, *Z. sapae* nov. sp. and *Zygosaccharomyces* sp. NCYC 3042 shared the same D1-D2 domain, but differed both in genotypic features and physiological characteristics. Incongruity reflects difficulties in reconciling various species concepts (SC) in hemiascomycetes taxonomy, such as the morphological, the biological or the genealogical SC. Species boundaries can vary considerably depending on which concept is applied. Moreover, incongruity among different phenotypic, genotypic and phylogenetic information may be indicative of genetic exchange or reticulation among individuals within a species or among closely related species (Taylor et al., 2000). Postulated hybrid species are very common in plants, animals (Mallet, 2007) and fungi (Albertin and Marullo, 2012). In the *Saccharomyces* genus, the postulated hybrid species *Saccharomyces pastorianus* is extensively exploited as lager-style brewing starter. Like *Saccharomyces pastorianus*, *Z. sapae* sp. nov. could be postulated as a hybrid species. This hypothesis is partially supported by sequence polymorphisms in phylogenetic (ITS regions) and housekeeping (*ZrSOD2* and *HIS3*) markers, but not by the sequence of D1-D2 domains, in which no polymorphisms were detected. This peculiar mosaic structure may suggest that the novel species have gained polymorphic alleles by postulated hybridization, followed by gene rearrangement

and loss. Based on the similarity of D1-D2 and ITS domains, a putative parent could be a strain resembling *Zygosaccharomyces* sp. NCYC 3042.

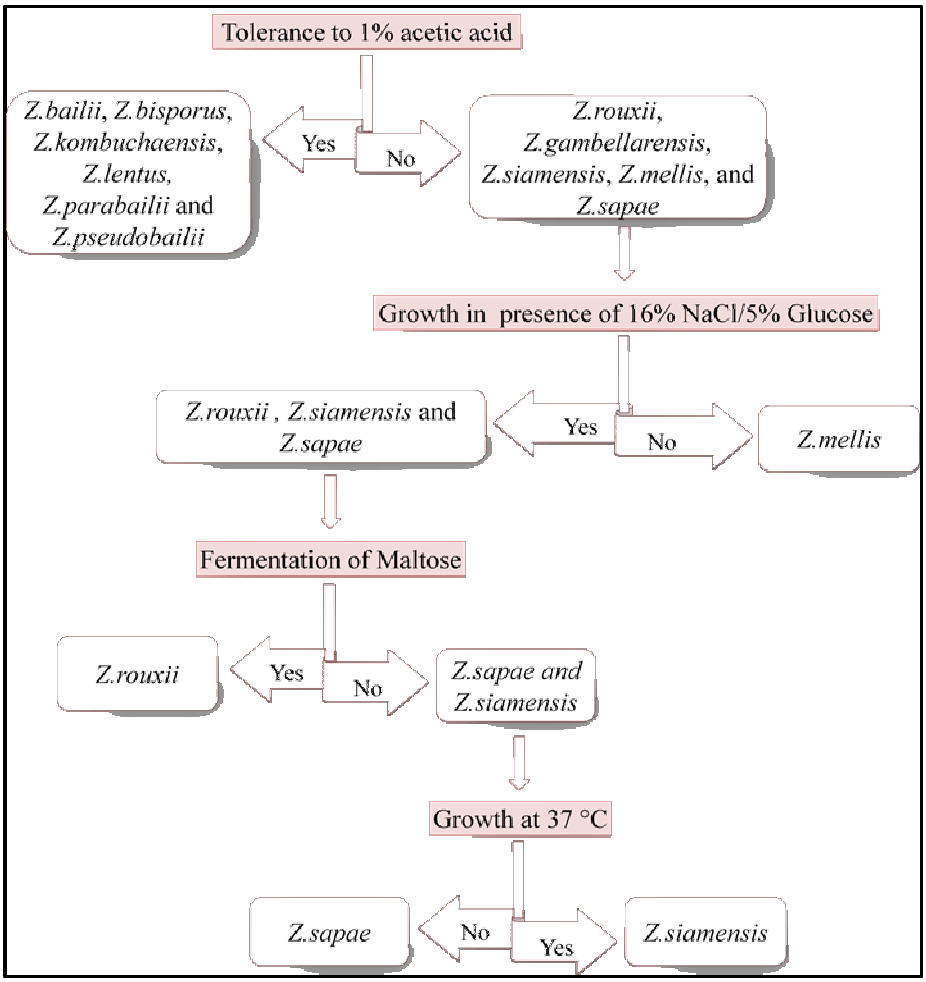


Figure 2.7. PSR (Phylogenetic species recognition) species diagnostic tree for discrimination of *Z. sapae* from other *Zygosaccharomyces* species.

Alternatively, we may hypothesize that *Z. sapae* sp. nov. and *Zygosaccharomyces* sp. NCYC 3042 shared an ancestral progenitor carrying multiple copies of several loci or 'double genome', which were lost in different ways across lineages (Gogarten and Townsend, 2005). This latter scenario envisages that gene loss events lead to the divergence of different clusters, some of them resembling a postulated 'hybrid genome', others showing a 'patchy' structure, with double copies of some genes and single copies of others. The differences highlighted here between *Z. sapae* sp. nov.

and *Zygosaccharomyces* sp. NCYC 3042 suggest that the position of the latter in relation to *Z. sapae* sp. nov. should be re-examined in the future, in order to provide relevant clues on the taxonomy of the genus *Zygosaccharomyces* and the evolutionary forces shaping diversity within this clade.

	CBS 732[T]					ATCC 42981					CBS 4837					CBS 4838				
Fermentation of	**3[§]**	**7**	**14**	**21**	**Final profile**	**3**	**7**	**14**	**21**	**Final profile**	**3**	**7**	**14**	**21**	**Final profile**	**3**	**7**	**14**	**21**	**Final profile**
Glucose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Fructose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Maltose	-	+	+	+	+	-	-	+	+	+	-	+	+	+	+	-	-	+	+	+
Lactose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Galactose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sucrose	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-
Trehalose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Assimilation on																				
Glucose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Fructose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Maltose	-	+	+	+	+	-	+	+	+	+	-	+	+	+	+	-	+	+	+	+
Lactose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Galactose	-	-	+	+	+	-	-	+	+	+	-	-	+	+	+	-	-	+	+	+
Sucrose	-	-	-	w	**w/l**	-	-	-	w	**w/l**	-	-	+	+	+	-	-	+	+	+
Trehalose	-	-	+	+	+	-	+	+	+	+	-	+	+	+	+	w	+	+	+	+
Xylose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ribose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Arabinose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Rhamnose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Melebiose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Raffinose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sorbitol	-	w	+	+	+	-	-	w	w	**w**	-	-	w	w	**w**	-	-	w	w	**w**
Mannitol	-	+	+	+	+	-	+	+	+	+	-	-	-	-	-	-	-	-	-	-
Inositol	-	-	w	w	**w**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Erythritol	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethanol	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Glycerol	-	w	w	w	**w**	-	+	+	+	+	-	+	+	+	+	-	+	+	+	+
Growth on																				
50% Dextrose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
60% Dextrose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Cycloeximide resistance	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1% Acetic Acid tolerance	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16% NaCl/5% Dextrose	-	+	+	+	+	w	+	+	+	+	-	+	+	+	+	-	+	+	+	+
3%Gly_1M NaCl	-	-	w	+	**w/+**	-	+	+	+	+	-	+	+	+	+	-	+	+	+	+

	ABT301^T					ABT302					ABT601					ABT602				
Fermentation of	**3**	**7**	**14**	**21**	**Final profile**	**3**	**7**	**14**	**21**	**Final profile**	**3**	**7**	**14**	**21**	**Final profile**	**3**	**7**	**14**	**21**	**Final profile**
Glucose	w	+	+	+	+	w	+	+	+	+	w	+	+	+	+	w	+	+	+	+
Fructose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Maltose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Lactose	-	-	-	-	-	-	-	-	w	w/+	-	-	-	-	-	-	-	-	-	-
Galactose	-	-	-	-	-	-	-	-	w	w/+	-	-	-	-	-	-	-	-	-	-
Sucrose	-	-	-	-	-	-	-	-	w	w/+	-	-	-	-	-	-	-	-	-	-
Trehalose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Growth on																				
Glucose	w	+	+	+	+	w	+	+	+	+	w	+	+	+	+	w	+	+	+	+
Fructose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Maltose	-	-	+	+	+	-	-	+	+	+	-	-	+	+	+	-	-	+	+	+
Lactose	-	-	-	-	-	-	-	-	w	w/+	-	-	-	-	-	-	-	-	-	-
Galactose	-	-	+	+	+	-	-	+	+	+	-	-	w	+	+	-	-	w	+	+
Sucrose	-	-	w	w	w	-	-	-	w	w/l	-	-	-	w	w/l	-	-	-	w	w/l
Trehalose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Xylose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ribose	-	-	-	-	-	-	-	-	w	w/+	-	-	-	-	-	-	-	w	w	w
Arabinose	-	-	-	-	-	-	-	w	w	w	-	-	-	-	-	-	-	-	w	w/+
Rhamnose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	w/+
Melebiose	-	-	-	-	-	-	-	-	w	w/+	-	-	-	-	-	-	-	-	-	-
Raffinose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sorbitol	-	w	w	+	+	-	-	-	w	w/+	-	-	w	+	w/+	-	-	w	+	w/+
Mannitol	-	w	+	+	+	-	-	w	+	+	-	w	+	+	+	-	-	w	+	+
Inositol	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Erythritol	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethanol	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Glycerol	-	w	w	+	+	-	-	w	+	w/+	-	-	w	+	w/+	-	-	-	w	w/+
Growth on																				
50% Dextrose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
60% Dextrose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Cycloeximide resistance	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1% Acetic Acid tolerance	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16% NaCl/5% Dextrose	-	-	w	+	+	-	-	+	+	w	-	-	w	w	w	-	-	+	+	w
3%Gly_1M NaCl	-	-	w	+	w/+	-	-	+	+	w/l	-	-	-	w	w/l	-	w	+	+	w/l

	OUT7136					OUT7140					ABT 401					DBVPG 6920				
Fermentation of	**3**	**7**	**14**	**21**	**Final profile**	**3d**	**7**	**14**	**21**	**Final profile**	**3**	**7**	**14**	**21**	**Final profile**	**3d**	**7**	**14**	**21**	**Final profile**
Glucose	+	+	+	+	+	+	+	+	+	+	w	+	+	+	+	+	+	+	+	+
Fructose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Maltose	-	+	+	+	+	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-
Lactose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Galactose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sucrose	-	-	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-
Trehalose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Growth on																				
Glucose	+	+	+	+	+	+	+	+	+	+	w	+	+	+	+	+	+	+	+	+
Fructose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Maltose	-	+	+	+	+	w	+	+	+	+	-	+	+	+	+	-	+	+	+	+
Lactose	-	-	-	-	-	-	-	-	-	-	-	-	-	+	**l**	-	-	-	+	**l**
Galactose	-	-	-	+	+	-	-	+	+	+	-	-	+	+	+	-	-	+	+	+
Sucrose	-	-	+	+	+	w	+	+	+	+	-	-	+	+	+	-	-	+	+	+
Trehalose	-	-	-	+	**l**	-	-	+	+	+	-	-	+	+	+	-	-	+	+	+
Xylose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ribose	-	-	-	-	-	-	-	-	-	-	-	-	-	w	**w/+**	-	-	-	-	-
Arabinose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Rhamnose	-	-	-	-	-	-	-	-	-	-	-	-	-	w	**w/+**	-	w	w	+	**l**
Melebiose	-	-	-	-	-	-	-	-	-	-	-	-	w	w	**w**	-	w	+	+	+
Raffinose	-	-	-	-	-	-	-	w	w	**w**	-	-	w	w	**w**	-	-	w	w	**w**
Sorbitol	-	-	w	w	**w**	-	+	+	+	+	-	w	+	+	**w**	-	w	+	+	+
Mannitol	-	-	w	w	**w**	-	+	+	+	+	-	-	+	+	+	-	+	+	+	+
Inositol	-	-	-	-	-	-	-	-	-	-	-	-	w	w	**w**	-	-	-	-	-
Erythritol	-	-	-	-	-	-	-	w	w	**w**	-	-	w	w	**w**	-	w	w	+	**w**
Ethanol	-	-	-	-	-	-	-	-	-	**w**	-	-	-	-	-	-	w	+	+	+
Glycerol	-	+	+	+	+	-	w	+	+	+	-	+	+	+	+	-	w	+	+	+
Growth on																				
50% Dextrose	+	+	+	+	+	w	+	+	+	+	+	+	+	+	+	w	+	+	+	+
60% Dextrose	+	+	+	+	+	w	w	+	+	+	+	+	+	+	+	w	w	+	+	+
Cycloeximide resistance	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1% Acetic Acid tolerance	-	-	-	-	-	+	+	+	+	+	-	-	-	-	-	+	+	+	+	+
16% NaCl/5% Dextrose	+	+	+	+	+	-	-	-	-	-	+	+	+	+	+	-	-	-	-	-
3%Gly_1M NaCl	-	+	+	+	+	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-

	NBRC0505					NBRC0506					NBRC0521 and NBRC0525					NBRC0845					NBRC0845				
Fermentation of	**3**	**7**	**14**	**21**	**Final profile**	**3d**	**7**	**14**	**21**	**Final profile**	**3**	**7**	**14**	**21**	**Final profile**	**3d**	**7**	**14**	**21**	**Final profile**	**3d**	**7**	**14**	**21**	**Final profile**
Glucose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Fructose	-	-	+	+	+	-	-	+	+	+	-	-	+	+	+	-	-	+	+	+	-	-	+	+	+
Maltose	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	-	+	+	+	-	-	+	+	+
Lactose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Galactose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sucrose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Trehalose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Growth on																									
Glucose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Fructose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Maltose	-	-	+	+	+	-	-	+	+	+	-	-	+	+	+	-	-	+	+	+	-	-	+	+	+
Lactose	-	-	-	+	l	-	-	-	+	l	-	-	-	+	l	-	-	-	+	l	-	-	-	+	l
Galactose	-	-	+	+	+	-	-	+	+	+	-	-	+	+	+	-	-	+	+	+	-	-	+	+	+
Sucrose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Trehalose	-	-	+	+	+	-	-	+	+	+	-	-	+	+	+	-	-	+	+	+	-	-	+	+	+
Xylose	-	-	-	w	w/+	-	-	-	w	w/+	-	-	-	w	w/+	-	-	-	w	w/+	-	-	-	w	w/+
Ribose	-	-	-	w	w/+	-	-	-	w	w/+	-	-	-	w	w/+	-	-	-	w	w/+	-	-	-	w	w/+
Arabinose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Rhamnose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Melebiose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Raffinose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sorbitol	-	-	w	w	w	-	-	w	w	w	-	-	w	w	w	-	-	w	w	w	-	-	w	w	w
Mannitol	-	-	w	w	w	-	-	w	w	w	-	-	w	w	w	-	-	w	w	w	-	-	-	-	-
Inositol	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Erythritol	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethanol	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	w	w
Glycerol	-	-	+	+	+	-	-	w	w	w	-	-	w	w	w	-	-	w	w	w					
Growth on																					nd	nd	nd	nd	+
50% Dextrose	nd	nd	nd	nd	+	nd	nd	nd	nd	+	nd	nd	nd	nd	+	nd	nd	nd	nd	+	nd	nd	nd	nd	nd
60% Dextrose	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	-	-	-	-	-
Cycloeximide resistance	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1% Acetic Acid tolerance	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	w	w	+	l
16% NaCl/5% Dextrose	w	w	w	+	l	w	w	w	+	l	w	w	w	+	l	w	w	w	+	l	-	-	w	+	l
3%Gly_1M NaCl	-	-	w	+	l	-	-	w	+	l	-	-	w	+	l	-	-	w	+	l					

Supplementary Table S2.1. **Physiological growth response recorded routinely over a period (from 3 days to 1 month).**

Chapter 3:

Life cycle as the major evolutionary force for genetic complexity and diversity in *Zygosaccharomyces rouxii* complex

3.1. Introduction

Recent advancements in genome sequencing techniques have made a large number of sequence data available. Much of our understanding regarding genome complexity and diversity has come from these genome sequence data. Approximately one-half of the protein-coding gene families are preserved in all of the yeasts sequenced to date. However, there is a large variation in the gene order and configuration of chromosomes among different yeast species that affect the genome structure in whole. In this regard, meiotic recombination is of great importance because it promotes genetic diversity by creating new and potentially beneficial genetic combinations, and purges the harmful mutations. Therefore, in yeast context, genome structure has been frequently mentioned in literature to be strictly dependent upon the life cycle and the breeding system of hemiascomycetous yeasts (Knop, 2006; Billiard et al., 2012). Different mechanisms of breeding (mating) are known for individual heterozygosity and genetic diversity among yeast species (Knop, 2006).

In the genus *Saccharomyces*, which includes the well-studied yeast *Saccharomyces cerevisiae*, the gene content and the rough organization of chromosomes are mostly preserved within isolates of the same species and among species (Kellis et al., 2003). This low variability is probably promoted by regular sexual cycles which help to preserve the genome organization (Fischer et al., 2000; Delneri et al., 2003). Natural *S. cerevisiae* strains are often diplontic, homothallic and heterozygous yeasts, which have the possibility of regenerating homozygous diploid cell from a haploid cell through self

fertilization, a mechanism interpreted as a way of genome renewal (Mortimer et al., 1994). With the important exception of mating-type interconversion, the genome seems to be quite stable: Other DNA rearrangements appear to be infrequent and of peripheral significance to the yeast life cycle. Genes are tightly packed, with little "spacer" DNA and strikingly few pseudogenes. There is no extensive clustering of functionally related or coordinately regulated genes. Genetic distance is largely proportional to physical distance, at least when averaged over tens or hundreds of kilobase pairs. Finally, the *cis*-acting sequences required for chromosome function—telomeres, centromeres, and replication origins—are only a few hundred base pairs long and function in almost any context.

The extent of intra- and interspecific genome variability is not well known for other hemiascomycetes. Recently, genome sequencing efforts have revealed an increasing number of yeast diploid mosaic species. In the GTG clade (so called because its constituent species translate CTG as Serine instead of Leucine), *Candida albicans* (Jones et al., 2004), *Debaryomyces hansenii* (Jacques et al., 2010), *Pichia sorbitophila* (Louis et al., 2012) and *Millerozyma (Pichia) farinosa* (Mallet et al., 2012) showed frequently complex diploid genomes, resulting from putative genetic exchanges between divergent strains and species at high frequency and thus referred to as 'mosaic genomes' (Zhaxybayeva et al., 2004).

Zygosaccharomyces clade encompasses industrially important halotolerant and osmotolerant yeasts that participate both in the elaboration and spoilage of foodstuff. Genome sequencing demonstrated that the type species of the genus, namely *Zygosaccharomyces rouxii*, is a protoploid lineage that did not undergo the whole genome duplication (WGD) leading to the genus *Saccharomyces* (Dujon et al., 2004). The genome of *Z. rouxii* strain CBS 732^{T} has a low number of duplicated genes (Souciet et al., 2009) and a haploid DNA content (Solieri et al., 2008a). Given the tree topology of Kurtzman (2003) and the fact that the *Vanderwaltozyma* clade emerged after WGD

(Scannell et al., 2007), *Zygosaccharomyces* clade represents the closest relatives to the putative ancestral genome of *S. cerevisiae* (Souciet et al., 2009).

Zygosaccharomyces rouxii and its close relatives display an important phylogenetic position, yet the extent of genome variability and the life cycle of species in this clade have not been studied extensively. These yeasts were generally regarded as spending their vegetative life in the haploid phase and meiosis follows immediately after mating of two cells (Wickerham and Burton, 1960). Some genetic surveys have reported that *Z. rouxii* strains had unusual hypervariable karyotypes (de Jonge et al., 1986; Török et al., 1993; Oda and Tonomura, 1995). Other works have also suggested that *Z. rouxii* and phylogenetically closely related species exhibit mosaic genome structure with respect to some nuclear and mitochondrial phylogenetic markers (James et al., 2005; Solieri et al., 2007; Gordon and Wolfe, 2008; Suezawa et al., 2008; James and Stratford, 2011). This allowed us to refer the group of these species as *Z. rouxii* complex. For instance, a new diploid species distinct from haploid *Z. rouxii*, namely *Zygosaccharomyces sapae*, has been recently delineated on the basis of multi-gene phylogeny and physiological evidence (Solieri et al., 2008a, 2013a). Intragenomic heterogeneity in the internal transcribed spacer region and the intervening 5.8S rRNA gene and the distribution of housekeeping gene markers partially supported the hypothesis of reticulate speciation (Solieri et al., 2013a). Additionally, strains extensively used in *Z. rouxii* genetics and molecular biology, such as ATCC 42981 and CBS 4837, harbor mosaic genomes with two copies of many genes. Sequence analysis of individual genes indicated that the parental strains contributing to the mosaic genome closely resemble *Z. rouxii* type strain CBS 732^{T} (t-subgenome) and *Zygosaccharomyces pseudorouxii* nom. inval. (p-subgenome) (James et al., 2005; Gordon and Wolfe, 2008). However, strain CBS 4837 was previously described as being a *Z. rouxii* haploid and heterothallic strain with the mating type MTa (Wickerham and Burton, 1960), a condition incongruent with the hypothesis of allodiploidization (James et al., 2005). Furthermore, strain CBS

4837 is able to mate with the sibling strain CBS 4838 (mating type MTα) (Wickerham and Burton, 1960), but the genomic features of the latter strain have not yet been investigated.

To cover these gaps and to unravel the genetic structure of species of *Z. rouxii* complex, we compared *Z. rouxii* CBS 732^T with six relatives on the basis of a polyphasic analysis. These strains included two that were retrieved from sweet food-grade and recently ascribed to *Z. sapae*; the remaining four, isolated from salt-rich foodstuff, are currently classified as *Z. rouxii* or of doubt assignment. The results provided evidence that the strains considered in this study are a complex of haploid, aneuploid and diploid mosaic lineages. The life cycle of *Zygosaccharomyces* strains was discussed to explain the high diversity within *Z. rouxii* complex.

3.2. Material and methods

3.2.1. Strains, medium and culture conditions

Strains	Other collections	Source	Taxonomical positions	Mating type[a]/thallism	Sporulation	References
CBS 732^T	NCYC 568, NRRL Y-229	Grape must	*Z. rouxii*	MTa/homothallic	-	Sacchatti, 1932
CBS 4837	NCYC 1682, NRRL Y2547	Miso	*Z. rouxii*/hybrid*	MTa/heterothallic	+	Mori and Onishi, 1967; James et al. 2005
CBS 4838	NRRL Y2548	Miso	*Z. rouxii*	MTα/heterothallic	+	Mori and Onishi, 1967
ATCC 42981	-	Miso	*Z. rouxii*/allodiploid	Nd	Nd	Kiuchi et al., 1980
OUT7136	-	Soy moromi	*Z. rouxii*	Nd		Provided by Prof. Y. Kaneko
ABT301^T	-	TBV	*Z. sapae*	Nd	+	Solieri et al. 2006, 2007
ABT601	-	TBV	*Z. sapae*	Nd	+	Solieri et al., 2006, 2007

[a] defined as mating behavior; nd: not determined; * postulated hybrid status

Table 3.1. Strain used in this study.

The strains used in this study are listed above in **Table 3.1**. Strains were obtained as lyophilized stocks from CBS (Centralbureau voor Schimmelcultures, Delft, the Netherlands) and ATCC (American Type Culture Collection). Strain OUT7136 was kindly provided by Prof. Y.

Kaneko (Osaka University, Japan). The strains were routinely grown in rich medium (YPD) containing 1% yeast extract, 2% peptone and 2% glucose at 28°C. The strains were maintained on solid media (YPDA) containing 1% yeast extract, 2% peptone, 2% glucose and 2% agar on slants in test tubes plugged with cotton. Alternatively, 1-mL stocks containing 25% glycerol were stored at 80 °C. For ploidy estimation and flow cytometry (FCM) analysis, yeast nitrogen base (Difco) medium supplemented with 5% glucose (YNB-5%Glucose) was used.

3.2.2 Genomic DNA extraction

For extraction of genomic DNA overnight liquid cultures of the strains were used. Genomic DNA isolation procedure utilizes physical disruption of the yeast cell and relies on Phenol-Chloroform extraction methods as discussed in previous chapter (**refer section 2.2.2, Chapter 2**). The extracted DNA is analyzed on 1% Agarose gel with electrophoresis run at 120V for 30 mins and stored at -20°C. Resultant DNAs after appropriate dilution were used in standard PCR reactions.

3.2.3. Genotyping and sequence analysis of phylogenetic markers

Randomly amplified polymorphic DNA (RAPD) with M13 primer and microsatellite primed-PCR with synthetic primer $(GTG)_5$ [$(GTG)_5$-PCR] were performed and analyzed, as previously reported (Solieri et al., 2013a) (**refer section 2.2.3, Chapter 2**). The *HIS3* and *ZrSOD2* genes, encoding imidazole-glycerol-phosphate dehydrate gene and plasma-membrane Na^+/H^+-antiporter gene, respectively, were amplified, as previously reported (Solieri et al., 2007; Harrison et al., 2011). The mitochondrial cytochrome C oxidase II gene (*COX2*) was amplified with primers COII-5 and COII-3 as reported by Belloch et al. (2000).

3.2.4. PCR amplification of 5.8S ITS D1-D2 domain of 26S rDNA region

The entire internal transcribed spacer (ITS) region (ITS 1, ITS 2 and the intervening 5.8S rRNA gene) for cloning purpose was amplified using primers ITS1 and ITS4 (White et al., 1990) as described in previous chapter (**refer section 2.2.4, Chapter 2**). Domains 1 and 2 of the 26S rRNA region (D1-D2), for restriction digestion with *Ava*I and cloning purpose, were amplified using primers NL1 and NL4 (O'Donnell, 1993) as described in previous chapter (**refer section 2.2.4, Chapter 2**). Aliquots (5µL) of each amplified product were electrophoretically separated in a 0.8% agarose gel (containing Ethidium bromide) in 0.5X Trisborate-EDTA buffer and visualized under UV illumination (BioDoc Analyze, Biometra, Göttingen, Germany). GenRuler 100bp plus molecular weight marker was included in each run (MBI Fermentas).

3.2.5. Purification of PCR amplicons and sequencing

To confirm the identification of *Zygosaccharomyces* strains at species level, the 5.8S ITS region and D1-D2 domain of 26S rDNA was send for sequencing to MWG Biotech, Germany after purification and quantification by Nanodrop of the amplicons generated by PCR amplification of the cloned genes. PCR products were purified by the DNA Clean and Concentrate TM-500 Kit (Zymo Research), following the manufacturer's instructions. The concentration of the amplicons send for sequencing to commercial sequencing service provider (MWG, Germany) for sequencing.

3.2.6. Cloning of 5.8S ITS region and 26S D1/D2 domain

Ligation

Ligation of PCR amplicons (8µg) of amplified 5.8S ITS genes and D1-D2 domain of 26S rDNA in PGEM vector (1µL) was carried out with 1µL of T4 DNA ligase in 1X ligase buffer [50

mM Tris-Cl (pH 7.8), 10 mM $MgCl_2$, 10 mM dithiothreitol, 25 mg/ml BSA and 1mM rATP] overnight at 4°C. 4-6mL of the ligation mixture was used for transforming *E. coli* competent cells (JM 109).

Transformation

The ligated DNA was used to transform *E. coli* JM 109 (competent cells) according to manufacturer's instructions. Detection of successful ligations was done by blue-white screening technique wherein competent *E. coli* cells were spread on LB agar plates containing 100µg/ml ampicillin, 0.5mM IPTG (Cat.# V3955) and 40µg/ml X-Gal (Cat.# V3941) and incubated overnight at 37°C. Successful transformation of plasmid containing insert DNA in the multiple cloning site (MCS) disrupting the *LacZα* gene resulting in non-production of functional β-galactosidase and produces white colony. Whereas transformants containing plasmid without insert DNA possess functional *LacZα* gene (uninterrupted) produces blue colonies. Subsequently, white colonies were picked aseptically and inoculated in LB tubes which thereafter are incubated overnight at 37°C. The following day, the plasmid DNA from different clones was extracted using Qiagen Miniprep kit according to manufacturer's instructions.

3.2.7. PCR-Restriction fragment length polymorphism (PCR-RFLP)

The cloned genes for 5.8S ITS region and D1-D2 domain of the 26S rDNA is amplified using standard PCR procedure and the PCR products (8µL) were digested with 1U of restriction endonucleases (*Hae*III, *Hinf* I and *Cfo*I for 5.8S ITS region and *Ava*I and *Cfo*I for D1-D2 domain of 26S rDNA) (Roche Diagnostics, Mannheim, Germany) in eppendorf tube containing 2µL of Tango buffer (10X) in final reaction volume of 20µL at 37°C for 2 hrs. PCR products and their restriction fragments were analyzed on 2% agarose gel in 0.5×TBE (89 mM Tris-borate, 2 mM EDTA pH 8)

buffer. Fragment lengths were estimated by comparing them with a GenRuler 100bp Plus DNA ladder as size marker (MBI Fermentas).

3.2.8. Flow cytometric analysis

For flow cytometry analysis, 1 X 10^7 control cells and Quinoline treated cells harvested during culture (at OD_{600nm}~0.5) were fixed in Ethanol. For Ethanol fixation, cells were collected by centrifugation at 8,000 rpm for 5 min at 4°C, washed with 50mM Sodium citrate, fixed in 1mL 70% Ethanol and kept at 4°C until used. Prior to flow cytometric analysis cells were stained with Propidium iodide (PI staining). For PI staining, 500µL of the Ethanol-fixed cells were recovered by centrifugation at 8,000 rpm for 5 mins at 4°C. The supernatant was discarded and the cells were washed with 500µL of 50mM Sodium citrate buffer (pH-7) followed by resuspension in same buffer with 12.5µL RNAse stock solution (10mg/mL) and incubation at 50°C for 2 hrs. After incubation, 25µL of Proteinase K (20mg/mL) was added and 2.5 X 10^6 cells were stained overnight at 4°C with Propidium iodide at a final concentration of 4µg/mL. These stained cells were used for flow cytometry analysis. Prior to analysis, in order to remove clumps and excessive debris, two consecutive ultrasound pulses were applied at 70% of the total output for 20 secs with an interval of 1-2 secs between the two pulses using Microson Ultrasonic Cell Disruptor XL (Misonix Inc., New Highway Farmingdale, NY, USA). Flow cytometric measurements were performed by EPICs XL (Coulter, Instrumentation Laboratory) flow cytometer equipped with an argon-ion laser emitting a 488 nm beam at 15mW, as previously reported (Solieri et al., 2008a).

3.2.9. Karyotying

Cell cycle analysis

Zygosaccharomyces rouxii and *Saccharomyces cerevisiae* yeast strains were grown in YPD and YNB5%Glucose medium batch culture to mid-log phase (OD600nm = 0.5-0.6). Cell samples in YNB5% Glucose medium were inducted to arrest in G1 phase by treatment with 8-hydroxyquinoline at a final concentration 100 mg/l for 24 h, as previously described (Johnston and Singer, 1978; Stöver et al., 1998). Cells were harvested by centrifugation (8000 g for 5 min at 4°C), fixed overnight with 70% ethanol (v/v) at 4°C and subjected to cell cycle analysis as previously described (Fortuna et al., 2000) and modified as follows. After fixation, cells were harvested, washed, and suspended at 5 X 10^6 cells/ml in 500 l of sodium citrate buffer (50 mM; pH 7.0). Cells suspension was treated for 120 min at 50°C with RNAse A (0.25 mg/ml) and/or proteinase K (2 mg/ml). A cell suspension of 2.5 X 10^6 cell/ml was stained overnight at 4°C with propidium iodide (PI; Sigma) at a final concentration of 4 µg/ml. To remove of clumps and excessive debris, two consecutive ultrasound pulses were applied before FMC analysis at 70% of total output for 20 s with an interval of 1–2 s between the two pulses by using Microson Ultrasonic Cell Disruptor XL (Misonix Inc., New Highway Farmingdale, NY, USA).

Estimation of genome size

All FCM experiments were performed in triplicates on an EPICs XL (Coulter, Instrumentation Laboratory, Milano, Italia) flow cytometer equipped with an argon-ion laser emitting a 488 nm beam at 15 mW. A minimum of 20,000 cells per sample were acquired at low flow rate and an acquisition protocol was defined to measure forward scatter (FS LOG) and side scatter (SS LOG) on a four-decade logarithmic scale and red Fluorescence (FL3) on a linear scale. For DNA content analysis, doublets and larger aggregates were discarded by electronic analysis of integral- and height-signals from the particles analyzed. Offline data were analyzed with the Windows Multiple Document Interface for Flow Cytometry 2.8 (WinMDI 2.8).

Plaque preparation

Yeast chromosomal DNA was prepared in plugs and preserved in 70% Ethanol until used. Plugs are then used in pulsed-field gel electrophoresis (PFGE) by a Chef Mapper apparatus (Bio-Rad), as previously reported (Solieri et al., 2008a).

Pulse-field gel electrophoresis (PFGE)

Pulse-field gels were run on a Bio-Rad CHEF (contour-clamped homogeneous electric field) apparatus. Chromosomal DNA was isolated as described by Carle and Olson (1985), with the exception of 5 h incubation with Lyticase (Sigma), in spite of an incubation of 2 h at 37°C with Zymoliase (Pribylova et al., 2007a). The 1% agarose gels were prepared with Seakem GTG agarose (FMC Bioproducts) and run in 0.5X TBE buffer at 13°C. The running conditions were: switch time 300 s, run time 100/120 h, angle 106°, voltage 3 V/cm. The chromosomal DNA size marker of *S. cerevisiae* S288 C (Bio-Rad Laboratories) was used to estimate the chromosome size. Pictures of the gels were digitally captured using the BioDoc Analyze gel imaging and analysis system (Biometra, Göttingen, Germany). The DNA banding patterns were analyzed with Bio-Numerics version 3.0 (Applied Maths, Belgium). Similarity matrix was calculated using Pearson's product moment correlation coefficient. Cluster analysis of similarity matrix was performed by the Unweighted Pair Group Method with Arithmetic mean (UPGMA) algorithm.

3.2.10. Multiple sequence alignment and phylogenetic analysis

Alignment of sequences obtained for the clones and reference sequences retrieved from the public database was performed with Clustal-X (Thompson et al. 1997). Phylogenetic analysis was conducted with MEGA version 4 (Tamura et al. 2007), with maximum likelihood estimation of

distances, neighbour joining (NJ) as the algorithm for tree reconstruction, and complete deletion of positions of ambiguous alignment.

3.2.11. Nucleotide sequence accession numbers

The DNA sequences described here were deposited in the GenBank data library under the accession numbers HE687308 and HE687309, and the accession numbers ranging from HE664087 to HE664107.

3.3. Results

3.3.1. Structuring *Z. rouxii* complex in two lineages

We applied two typing methods, such as M13-RAPD and $(GTG)_5$-PCR, to detect clonality and to assess the genetic relatedness among *Z. rouxii* CBS732^T and related strains listed in **Table 3.1**. Both methods generated complex fingerprinting patterns, with PCR products (number of bands 7 to 15 for M13-RAPD and 8 to 16 for GTG_5-PCR) ranging from 370 to 2920 bp in M13-RAPD and from 300 to 2830 bp in $(GTG)_5$-PCR profiles (**refer Table 6.2 and Table 6.3**). The strains fell into two groups, as identified by the cluster analyses of M13-RAPD patterns (**Figure 3.1**). $(GTG)_5$-PCR analysis also placed the strains into two clusters, with the exception of strain CBS 732^T, which did not fall into any of the clusters (**Figure 3.1**). In both cases, one group was delineated that includes *Z. sapae* strains isolated from high sugary foodstuff, whereas the other one grouped strains from salt-rich food, including strains previously described as being allodiploid (James et al., 2005; Gordon and Wolfe, 2008). For this reason we referred to the latter cluster as the mosaic lineage.

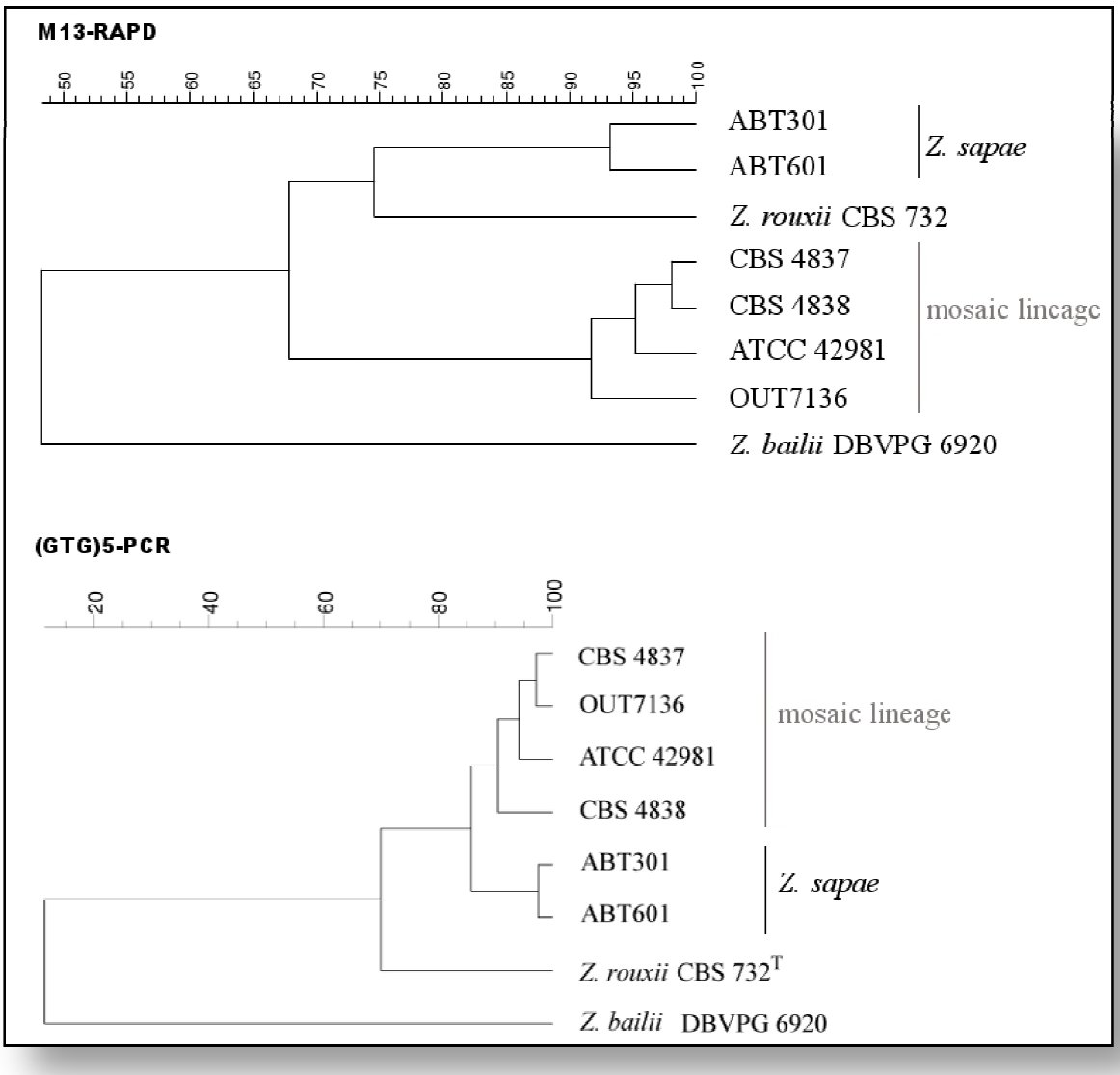


Figure 3.1. UPGMA dendrograms of *Z. rouxii* CBS 732^T and six relatives using M13-RAPD (a) and MSP-PCR(b). Pearson's similarity scale is shown at the top.

3.3.2. Analysis of mitochondrial DNA (mtDNA)

To address the question of the mtDNA inheritance within the mosaic lineage, sequence polymorphism of mitochondrial *COX2* marker was investigated in strains ATCC 42981, CBS 4838 and OUT7136. Phylogenetic analysis based on *COX2* sequences showed that strains of the mosaic lineage shared the *COX2* allele identical to that of *Z. rouxii* CBS 732^T, consistent with the hypothesis that these allodiploids derived from crosses between clades, one of which could be *Z. rouxii*.

Moreover, the mosaic lineage differed from *Z. sapae* strains, which carried a unique allele at the mitochondrial gene *COX2*, divergent both from *Z. rouxii* (2.1% divergence) and *Zygosaccharomyces mellis* (2.5% divergence), a close relative of *Z. rouxii* (**Figure 3.2**). These results support the differentiation of strains into two groups obtained by PCR fingerprinting.

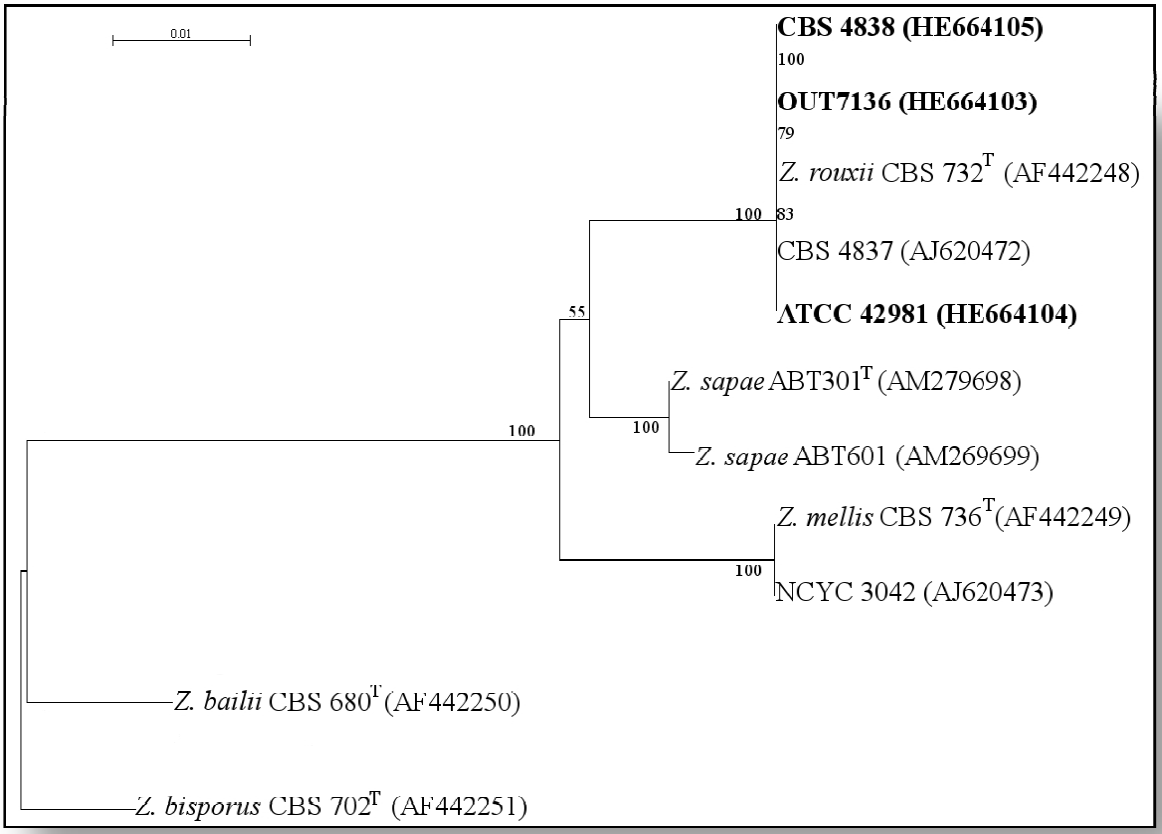


Figure 3.2. Phylogenetic analysis of mitochondrial *COX2* gene from *Z. rouxii* complex strains.

3.3.3. Sequence variations within intragenomic rRNA gene arrays

The phylogenetic markers 26S D1-D2 domains and ITS region are members of the structural RNA components, which are highly repetitive and conserved owing to homogenization among unit copies through unequal crossing-over and gene conversion (Dover, 1982). Their analysis is described as a useful tool for studying genome origin and stability (James et al., 2009), so these markers were

amplified and sequenced in strains CBS 4837, CBS 4838 and OUT7136 and compared with those available for *Z. sapae* ABT301^{T} and ABT601 and allodiploid strain ATCC 42981. Sequencing of 26S D1-D2 domains revealed intra-individual heterogeneity in strains CBS 4838 and CBS 4837. Mixed populations of PCR products were generated, which need to be cloned before sequencing. The results showed that both strains harbor at least two types of 26S D1-D2 domains, suggesting that they could be either diploid heterozygotes (variation between rRNA gene clusters located on two homeologous/homologous chromosomes) or heterogeneous for this marker (variation harbored by repeating units arranged in the same rRNA gene cluster). In both strains, we arbitrarily referred to the two D1-D2 types as copy 1 (100% *Z. rouxii* similarity in both strains) and copy 2 (100% and 99.7% *Z. sapae* similarity for CBS 4837 and CBS 4838, respectively). The latter copy was slightly different in strains CBS 4837 and CBS 4838 at five variable sites (divergence 0.87% in 572 nt analyzed). The 26S D1-D2 copies 1 and 2 of strains CBS 4837 differed from each other in one single indel and 15 variable sites, including 10 transitions and five transversions, resulting in an overall divergence of 2.8%. The occurrence of two divergent copies of D1-D2 domain in CBS 4837 partially disagreed with that previously reported by another study, which described only one 26S D1-D2 sequence occurring in this strain (James et al., 2005). Peculiar PCR conditions have been demonstrated to favor preferential amplification of a few repeat types (Fenton et al., 1998) and this could explain the homogeneity in CBS 4837 26S D1-D2 domains previously observed. The difference between the 26S D1-D2 copies 1 and 2 of strain CBS 4838 was of 1.9% and consisted of one single indel and 10 variable sites, including five transitions and five transversions. Similarly, sequence variation within intra-individual 26S D1-D2 genes was previously described in strain ATCC 42981 (Gordon and Wolfe, 2008). Although a member of the mosaic lineage, strain OUT7136 showed a single 26S D1-D2 domain sequence, 99% similar to that of *Z. sapae*. Phylogenetic relationships based on alignment of variable 26S D1-D2 sequences both among and within individual yeast 26S D1-D2 domains are

shown in **Figure 3.3**. This analysis separated the strains into two groups, namely *Z. rouxii* and *Z. sapae*. The *Z. sapae* clade consists of six strains, of which *Z. sapae* ABT 301^{T} and ABTz601 and strain OUT7136 share a unique copy of 26S D1-D2 domain. The remaining three, ATCC 42981, CBS 4838 and CBS 4837, harbor two copies, one resembling that of *Z. sapae* and the other that of *Z. rouxii*. In addition to the analysis of 26S D1-D2 domain variation, we applied a similar strategy to examine the patterns of variations in the ITS regions among and within strains CBS 4837, CBS 4838 and OUT7136. We detected a high degree of heterogeneity in strains CBS 4838 and CBS 4837 but not in OUT7136. ITS PCR products from strains CBS 4837 and CBS 4838 failed in sequencing and were subcloned. Approximately 12 resulting plasmids from each subcloning experiment were verified for the insert sequence. Comparisons of sequence variation patterns from conserved (5.8S) and less constrained (ITS 1 and ITS 2) regions showed that CBS 4837 and CBS 4838 harbor three divergent ITS sequences (arbitrarily referred to as copy 1–3). The set of ITS variants was only partially identical between the two strains. Strain CBS 4838 showed an ITS copy 1 identical to that of CBS 732T and very similar to ITS copy 1 of *Z. sapae* ABT301^{T} and t-subgenome copy of ATCC 42981 (**Figure 3.4**). The remaining copies 2 and 3 from strain CBS 4838 showed an identity of 88.4% and 87.2% compared with *Z. rouxii* CBS 732^{T}, respectively. Strain CBS 4837 displayed three variants in ITS regions, but all of them were significantly divergent from *Z. rouxii* (identity ranging from 91.7% to 87.2%). Conversely, sequencing of ITS domain from strain OUT7136 resulted in a single sequence identical to that of *Z. rouxii* CBS732^{T} (**Figure 3.4**). This result disagreed with that reported for 26S rDNA

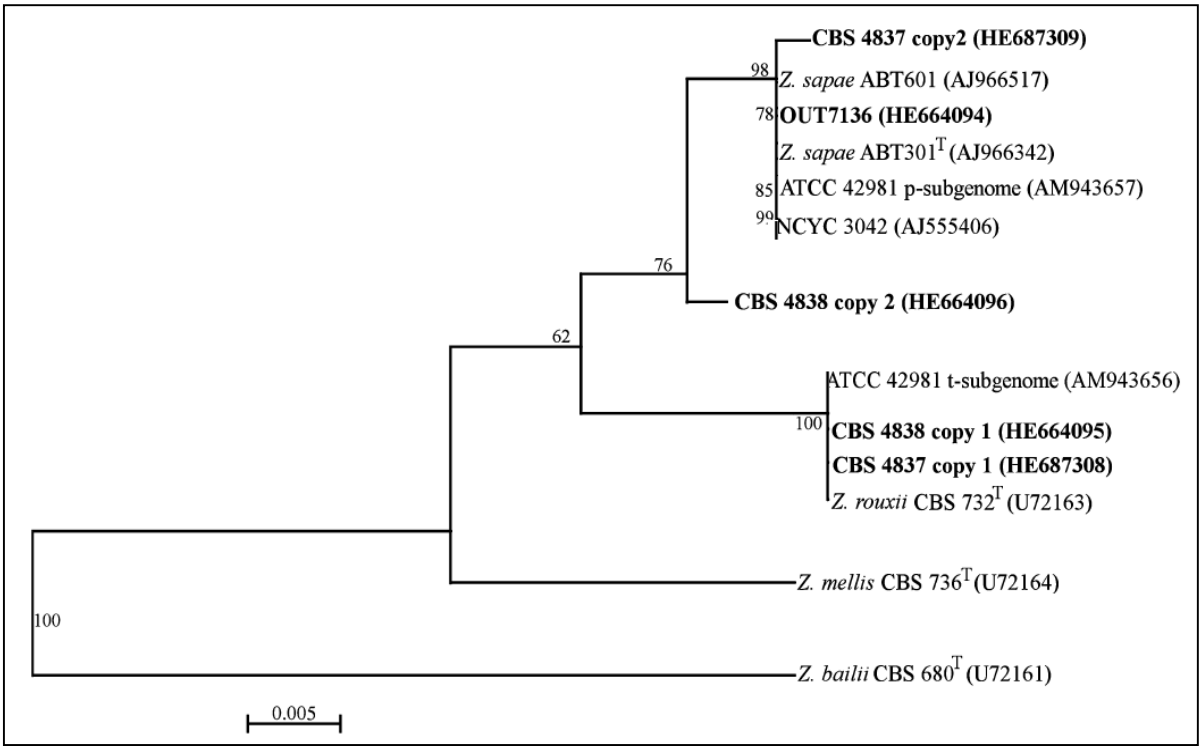
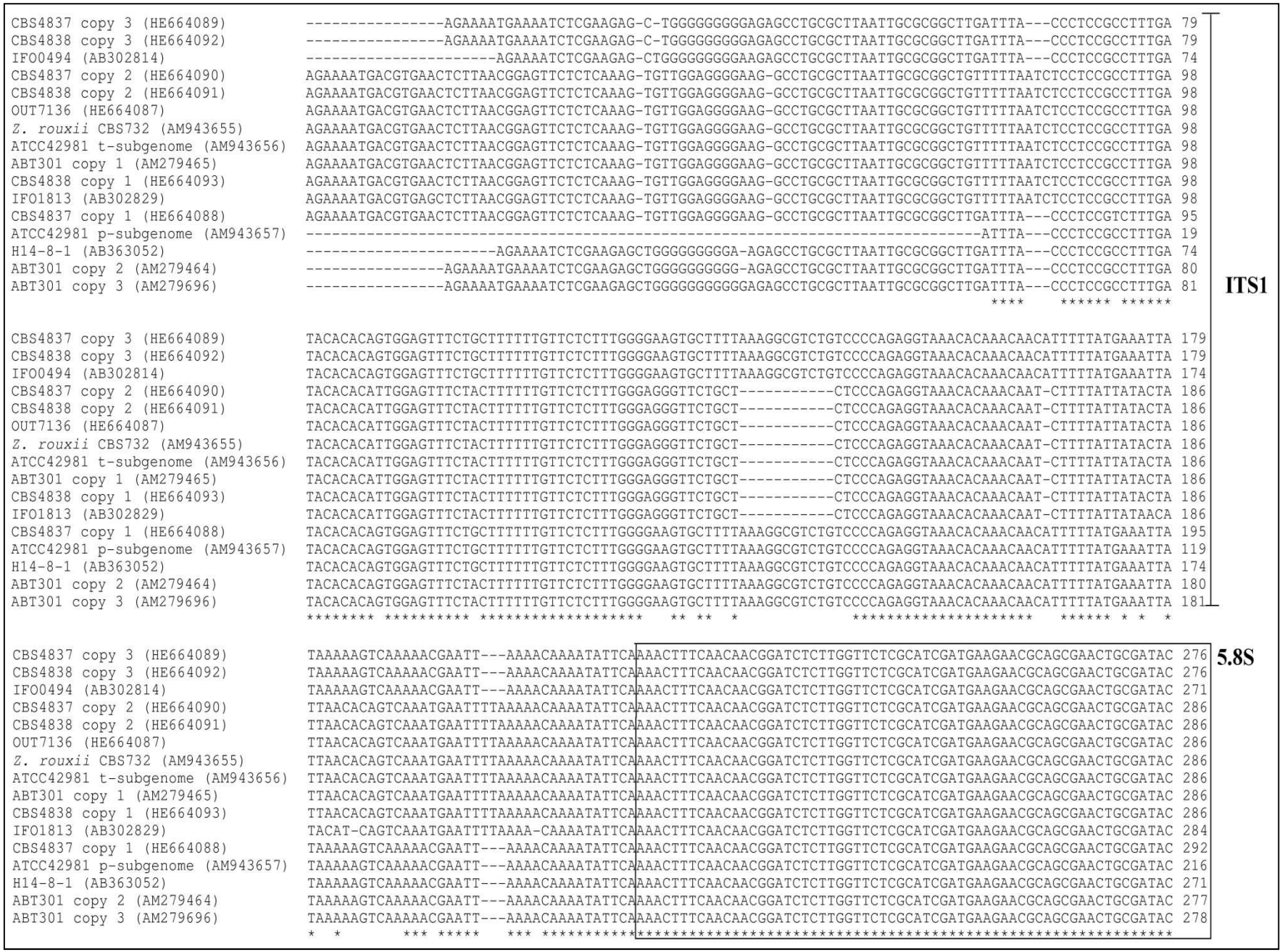


Figure 3.3. Phylogenetic analysis of 26S D1-D2 domain sequences of *Z rouxii* complex strains. *Zygosaccharomyces bailii* was used as outgroup. Copy 1 and copy 2 sequences indicate the presence of two different intragenomic sequences within a single individual. The sequences obtained in this study were reported in bold. The NJ method (Saitou and Nei, 1987) in Clustal-X (Thompson et al., 1997) was used, and gaps were not included. Numbers over the branches represent bootstrap coefficients from 1000 replicas (Felsenstein, 1985). Only bootstrap values over 50% are indicated.

26S D1-D2 domain, which was 100% identical to *Z. sapae*, and suggested that OUT7136 genome could harbor a patchy combination of *Z. rouxii* and *Z. sapae* sequences in different loci of rRNA gene arrays. The tree topology based on ITS sequences showed that clones from single individuals were scattered across the tree into three lineages, called S-group, R-group and recombinant group (**Figure. 3.5**). In particular, ITS copy 1 of CBS 4838 and copies 2 of CBS 4837 and CBS 4838 belonged to the R-group, whereas copy 1 of CBS 4837 and copy 3 from CBS 4838 and CBS 4837 belonged to the recombinant-group and displayed an nt identity higher than 98% with the strain IFO 0494, a uncharacterized *Zygosaccharomyces* yeast isolated from miso (Suezawa et al., 2008).


ITS1
5.8S

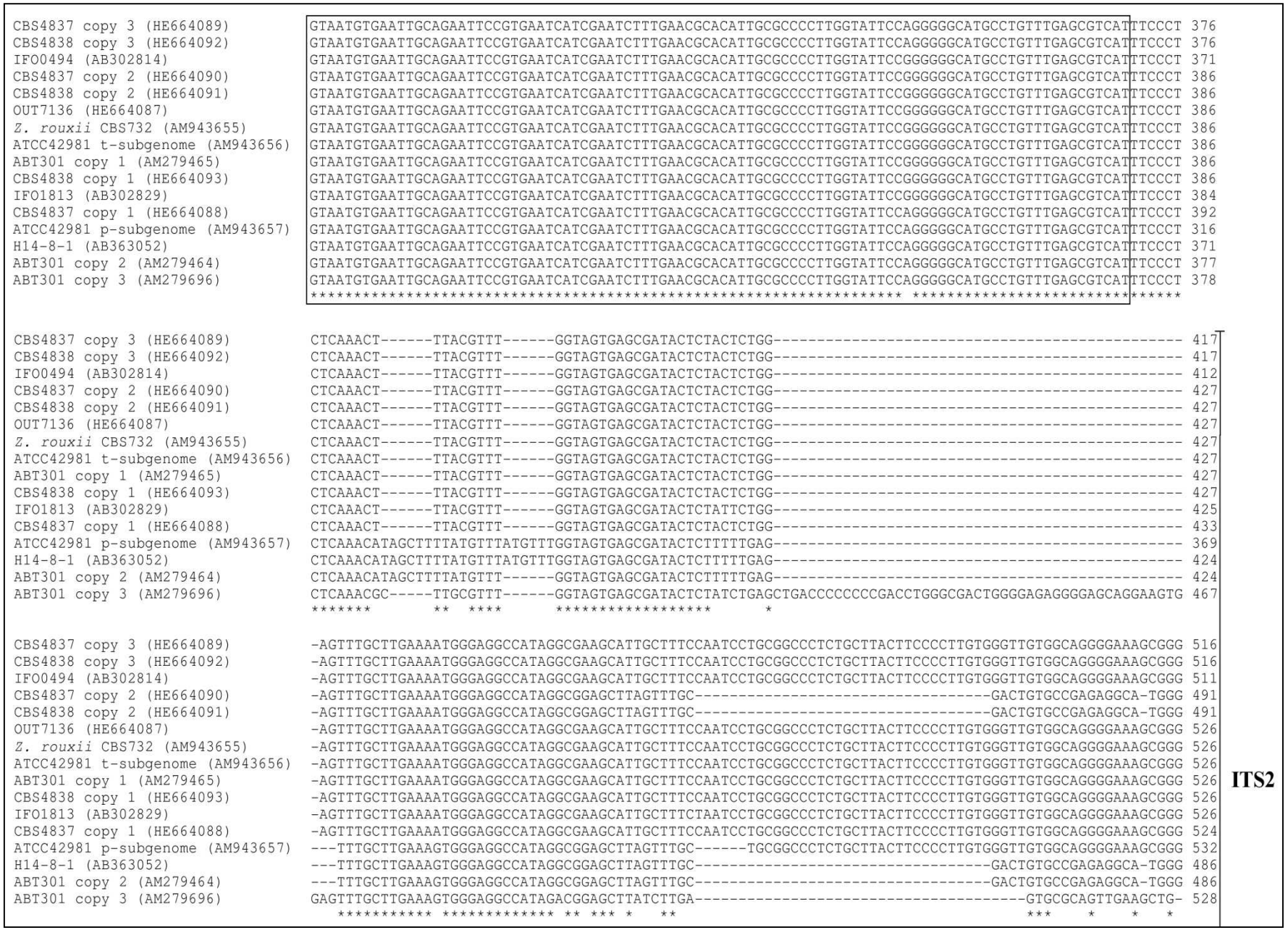

CBS4837 copy 3 (HE664089)
CBS4838 copy 3 (HE664092)
IFO0494 (AB302814)
CBS4837 copy 2 (HE664090)
CBS4838 copy 2 (HE664091)
OUT7136 (HE664087)
Z. rouxii CBS732 (AM943655)
ATCC42981 t-subgenome (AM943656)
ABT301 copy 1 (AM279465)
CBS4838 copy 1 (HE664093)
IFO1813 (AB302829)
CBS4837 copy 1 (HE664088)
ATCC42981 p-subgenome (AM943657)
H14-8-1 (AB363052)
ABT301 copy 2 (AM279464)
ABT301 copy 3 (AM279696)
ITS2

Figure 3.4. Alignment including intra- and intergenomic variable ITS sequences. Localization of SNPs and indels is mainly in low constrained ITS regions (indicated with arrow), whereas 5.8S rRNA genes (indicated with box) are highly conserved. Copies 1–3 after strain codes indicate sequences of different clones from a single individual.

Overall, the data suggested that the rRNA gene composition of CBS 4838 and CBS 4837 is similar to that of ATCC 42981, but there were some differences. Strain ATCC 42981 has a standard allodiploid rearrangement with two types of ITS regions and two types of 26S D1-D2 domains (Gordon and Wolfe, 2008), whereas strains CBS 4837 and CBS 4838 have two copies of 26S D1-D2 domain and three copies of ITS regions (**Table 3.2**). Strains CBS 4837 and CBS 4838 differed also from *Z. sapae* strains, which showed heterogeneity in ITS regions but not in 26S D1-D2 domains (Solieri et al., 2007). Three possible scenarios can explain the observed data: (1) these strains may be either aneuploid or diploid, with each chromosome of the pair of homeologous/homologous chromosomes bearing one rRNA gene variant; (2) there are different tandem repeat variants arranged in their rRNA gene arrays located on the same chromosome; (3) there are divergent rRNA gene arrays dispersed among different chromosomes.

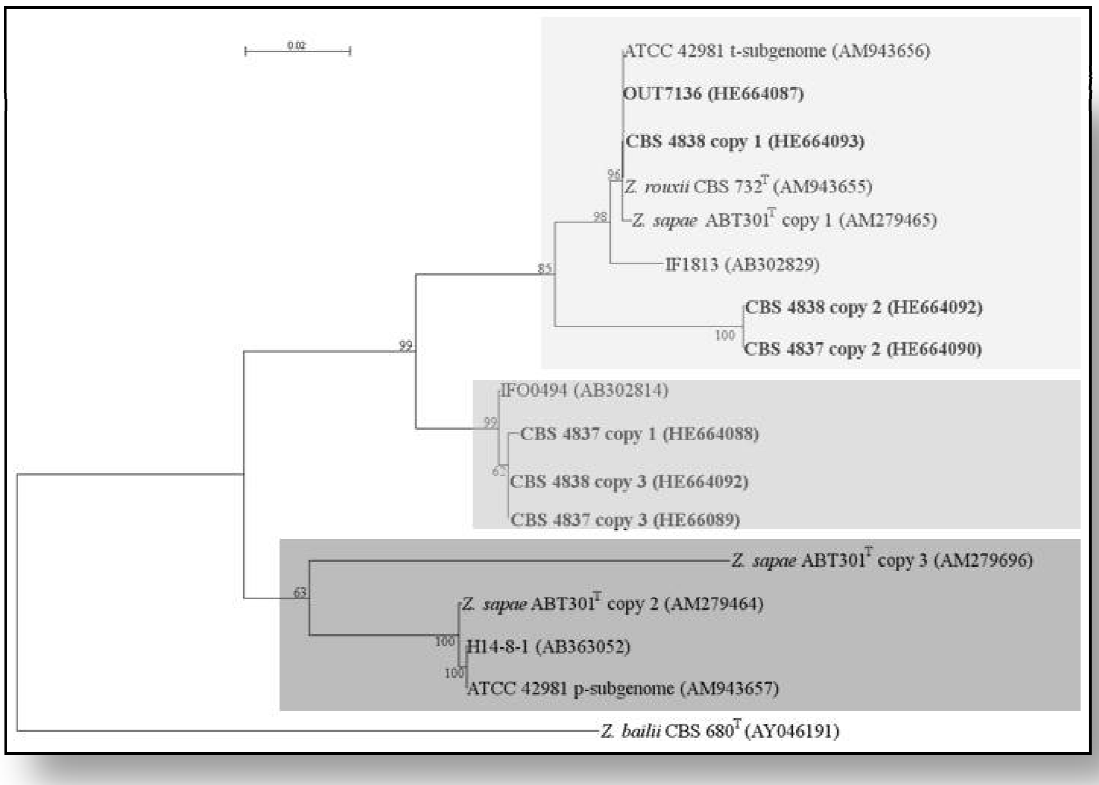


Figure 3.5. Phylogenetic relationships of *Z. rouxii* complex based on NJ analysis (Saitou and Nei, 1987) of ITS rDNA sequences. *Zygosaccharomyces bailii* was used as outgroup. The sequences obtained in this study are shown in

bold. The suffix (copy 1, copy 2 or copy 3) after strain code indicate the sequence belonging to different clones from a single individual. Shaded light grey indicates R-cluster; shaded medium grey indicates recombinant cluster; and shaded dark grey indicates S-cluster. All positions containing gaps and missing data were eliminated from the dataset. See other statistical details in the legend of **Figure 3.3**.

3.3.4. Sequencing of housekeeping markers

Previous studies have demonstrated that both *Z. sapae* and allodiploid strains ATCC 42981 and CBS 4837 possess divergent copies of the housekeeping markers *ZrSOD* and *HIS3* (Iwaki et al., 1998; Solieri et al., 2007). Conversely, a single copy of *ZrSOD* was found in CBS 732^{T} (Kinclova et al., 2001). We performed allele-specific PCR assays and sequencing of *ZrSOD* and *HIS3* genes for strains OUT7136 and CBS 4838 and results are shown in **Table 3.2**. Analysis of *ZrSOD* locus showed that strain OUT7136 harbours two *ZrSOD* variants, one 99.6% identical to ATCC 42981 *ZrSOD2* and the other one 100% identical to ATCC 42981 *ZrSOD22*. Similarly, strain CBS 4838 showed two copies, both 100% identical to ATCC 42981 *ZrSOD2* and *ZrSOD22* genes, respectively. The same combination of variants has been reported for strain CBS 4837 (James et al., 2005), suggesting that all members of the mosaic lineage share the identical pattern of *ZrSOD* variants. Differently from the mosaic lineage, *Z. sapae* possesses partially divergent *ZrSOD* variants: one copy was 98.6% identical to *ZrSOD2-22*, a variant previously found only in CBS 732^{T}, whereas the other one was similar to ATCC 42981 *ZrSOD22* (99.8 and 98.7% identity in ABT301^{T} and ABT601, respectively) (Solieri et al., 2007). Finally, when we constructed a phylogenetic tree, two divergent lineages were delineated (**Figure 3.6, panel I**). With the exception of CBS 732^{T}, each strain was represented in the two different lineages (designated as A and B, **Figure 3.6, panel I**), suggesting that copies from different strains, orthologues (sequences within a single lineage), are more similar to each other than are the two copies within the same strain.

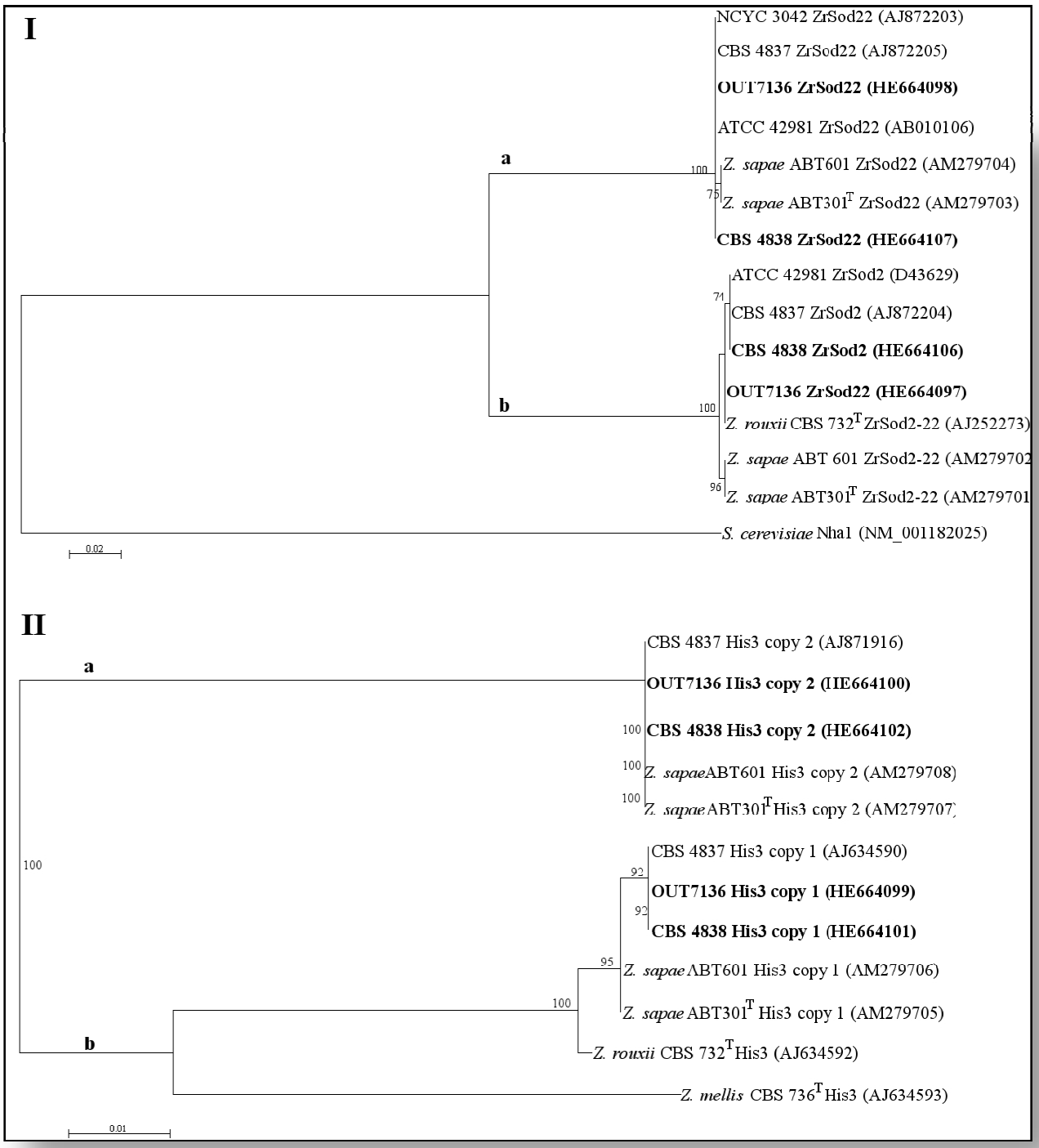


Figure 3.6. Phylogenetic relationships of haplotype sequences for *ZrSOD2* (521 bp) (a) and *HIS3* (362 bp) (b) genes. *Saccharomyces cerevisiae* (a) and *Zygosaccharomyces mellis* (b) were used as outgroups. The sequences obtained in this study are shown in bold. The suffix (copy 1 or copy 2) after strain code indicate haplotype amplified and sequenced with haplotype-specific primer pairs from each single strain. See other statistical details in the legend of **Figure 3.3**.

Similarly, allele-specific amplification and sequencing of *HIS3* gene were performed using two different primer pairs targeting CBS 732^T (copy 1) and ATCC 42981 (copy 2) *HIS3* genes, respectively. A successful amplification resulted from both primer pairs in strains OUT7136 and CBS 4838, confirming the presence of two different *HIS3* variants in their genomes. Similar results were found in all the other strains considered in this study, but not in *Z. rouxii* CBS 732^T (**Table 3.2**). The topology of *HIS3*-derived NJ tree was very similar to that of the *ZrSOD* marker and confirmed that two *HIS3* variants in each strain genome belong to two divergent phylogenetic lineages (**Figure 3.6, panel II**). The presence of orthologues more similar to each other than the two paralogues within the same strain suggested that the ancestor of all strains belonging to both *Z. sapae* and the mosaic lineage had at least two variants of each tested housekeeping marker.

Locus	Strains	Sequence type	Number of subcloned copy
5.8S-ITS‡	CBS 732^T	Homogeneous†	1
	OUT7136	Homogeneous	1
	ABT301^T	Heterogeneous†	3
	ABT601	Heterogeneous	3
	CBS 4837	Heterogeneous	3
	CBS 4838	Heterogeneous	3
	ATCC 42981	Heterogeneous	2
26S D1/D2 domain	CBS 732^T	Homogeneous	1
	OUT7136	Homogeneous	1
	ABT301^T	Homogeneous	1
	ABT601	Homogeneous	1
	CBS 4837	Heterogeneous	2
	CBS 4838	Heterogeneous	2
	ATCC 42981	Heterogeneous	2
ZrSOD2	CBS 732^T	Monomorphic	1
	OUT7136	Polymorphic	2
	ABT301^T	Polymorphic	2
	ABT601	Polymorphic	2
	CBS 4837	Polymorphic	2
	CBS 4838	Polymorphic	2
	ATCC 42981	Polymorphic	2
HIS3	CBS 732^T	Monomorphic	1
	OUT7136	Polymorphic	2

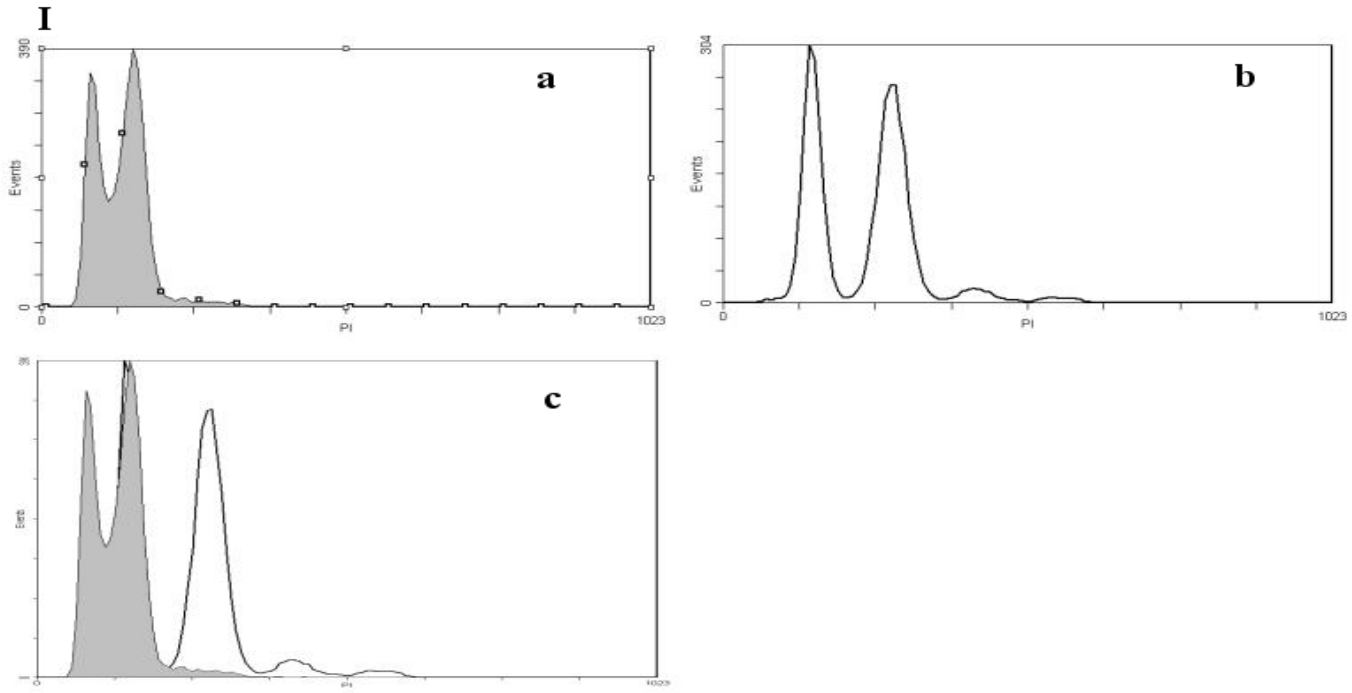
	ABT301^{T}	Polymorphic	2
	ABT601	Polymorphic	2
	CBS 4837	Polymorphic	2
	CBS 4838	Polymorphic	2
	ATCC 42981	Polymorphic	2
COX2	CBS 732^{T}	Monomorphic	1
	OUT7136	Monomorphic	1
	ABT301^{T}	Monomorphic	1
	ABT601	Monomorphic	1
	CBS 4837	Monomorphic	1
	CBS 4838	Monomorphic	1
	ATCC 42981	Monomorphic	1

The sequence analysis is based on the cloned PCR fragments for rRNA gene marker and on copy-specific PCR fragments for housekeeping markers. If only one sequence/copy was found in a strain, the locus was designated as monomorphic/homogeneous. If there was more than one copy, the locus was designated as polymorphic/heterogeneous. ‡ITS1, ITS2 and intervening 5.8S rRNA gene. †The terms 'heterogeneous/homogeneous' and monomorphic/polymorphic have been used for rRNA gene/regions and nuclear and mitochondrial genes, respectively.

Table 3.2. Sequence copy number (polymorphism) in *Z. rouxii* complex strains

3.3.5. Genome size and ploidy level

Analysis of housekeeping markers indicates that these yeasts could be not simple haploids (the classical state of *Z. rouxii* strains) and that at least some parts of the genome are duplicated Accordingly, strain ATCC 42981 and *Z. sapae* have been determined previously to be diploid (Solieri et al., 2008a). The occurrence of different variants in the genomes of the remaining strains CBS 4837, CBS 4838 and OUT7136 could suggest that they can be either diploid or aneuploid. We addressed the question by measuring total DNA content using an FCM approach. For CBS 732^{T} we obtained two peaks corresponding to two subpopulations, one with a 1n content (G0/G1 phases) and the other with a 2n content (G2/M phases) of the haploid genome (**Figure 3.7**). A clear shift of the two peaks towards a double amount of DNA was observed for strains OUT7136, CBS 4837 and CBS 4838 (**Figure 3.7**, panels I, II and III, respectively). In all cases, the first peak (G0/G1) was close to the second peak (G2/M) of the haploid CBS 732^{T}, consistent with a diploid or at least an aneuploid status of these strains.


I
a
b
c
Events
PI
0
1023
390
304

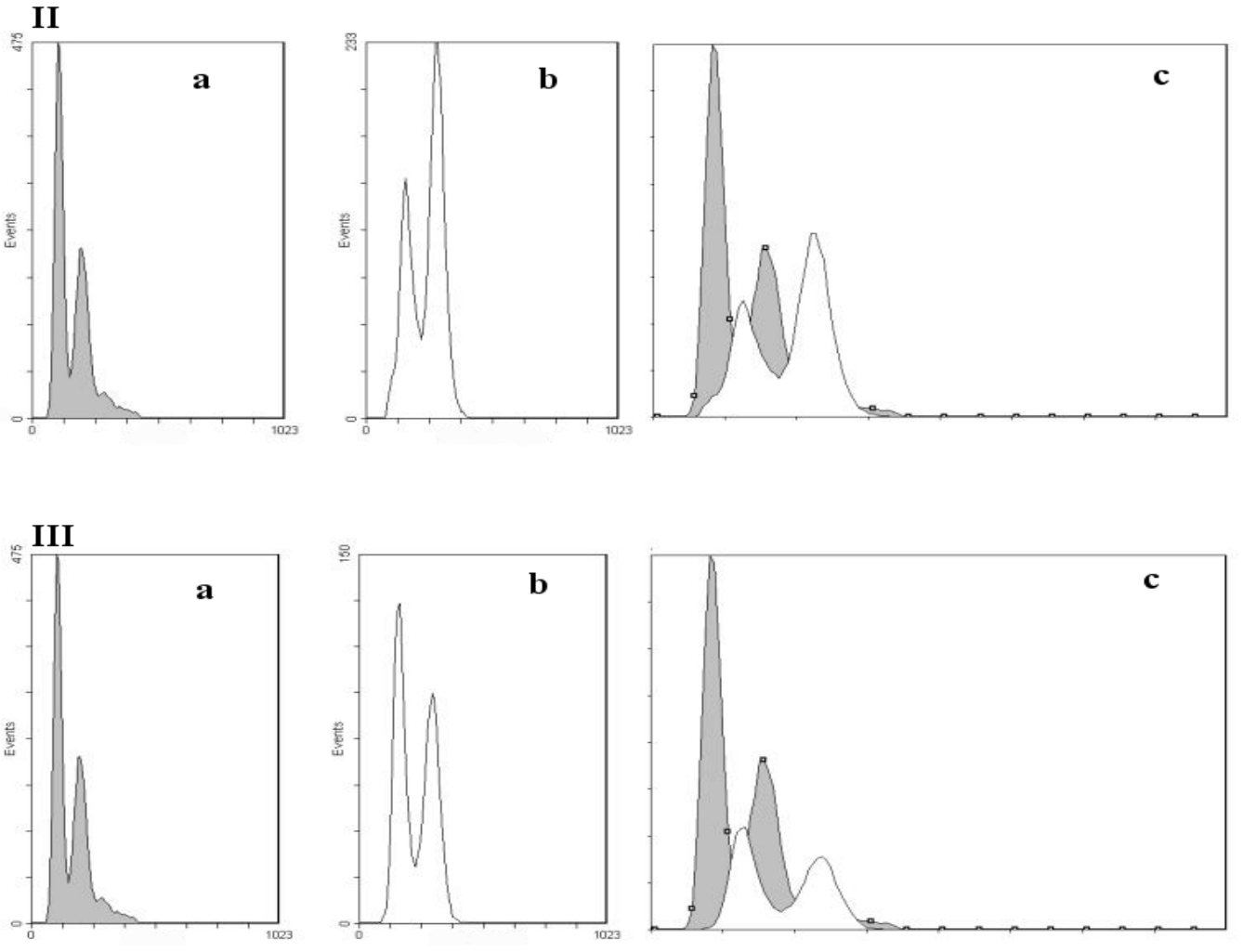

II
III
a
b
c
Events
0
1023
475
233
150

Figure 3.7. Fluorescence histograms of various *Z. rouxii* complex strains after propidium iodide staining. In each panel fluorescence histograms of haploid CBS 732^T (a) (shading light grey) and one mosaic strain (b) (white) were compared (c). (I) *Zygosaccharomyces* sp. OUT7136; (II) *Zygosaccharomyces* sp. CBS 4837; (III) *Zygosaccharomyces* sp. CBS 4838. The *y* axis of the graphs represents the total cell counts (×10) per sample, and the *x* axis indicates the relative fluorescence intensity of the samples.

The genome sizes were estimated from the G0/G1 peak median fluorescence intensity (MFI), using the genome size of *S. cerevisiae* as the calibration (**Table 3.3**). *Zygosaccharomyces sapae* strains ABT301^T and ABT601, and allodiploid ATCC 42981 have genome sizes ranging from 21.9 to 28.1 Mb (Solieri et al., 2008a). Strains OUT7136, CBS 4837 and CBS 4838 have estimated genome sizes of 19.57–21.94 Mb, significantly higher than the estimated 9.8-Mb genome of CBS 732^T (Souciet et al., 2009). We used PFGE to separate chromosomes and compare haploid and diploid strains in order to get an insight into the diversity of chromosome structure in *Z. rouxii* complex. Karyotyping showed that strains contained from six to 11 chromosomes. In some strains, a few bands showed a higher intensity (**Figure 3.8**), suggesting that some chromosomes overlap in size. The chromosome separation of strain CBS 732^T displays, as expected, a total of six bands, the 1.5-Mb band being a doublet made of chromosomes C and D. With the exception of a few chromosome length polymorphisms, the strains belonging to the mosaic lineage tend to harbor a similar number of chromosomal bands, amounting to eight chromosome bands per strain (**Table 3.3**). The chromosome sizes ranged from 1.2 to 2.2 Mb for strain OUT7136 and from 1.2 to 2.3 Mb for strains CBS 4837 and CBS 4838. Combination of FCM results and karyotypic data resulted in a ploidy level of 1.74 for OUT7136, 1.96 for CBS 4837, and 1.90 for CBS 4838, confirming the aneuploid/diploid status of these strains (**Table 3.3**). On the other hand, the chromosome separation of *Z. sapae* strains gave a higher number of chromosomes ranging from 10 to 11 bands, consistent with their higher genome sizes (Solieri et al., 2008a). The heterogeneity in karyotype within *Z. rouxii* complex suggests that the genome has rearranged very fast upon the separation of single lineages.

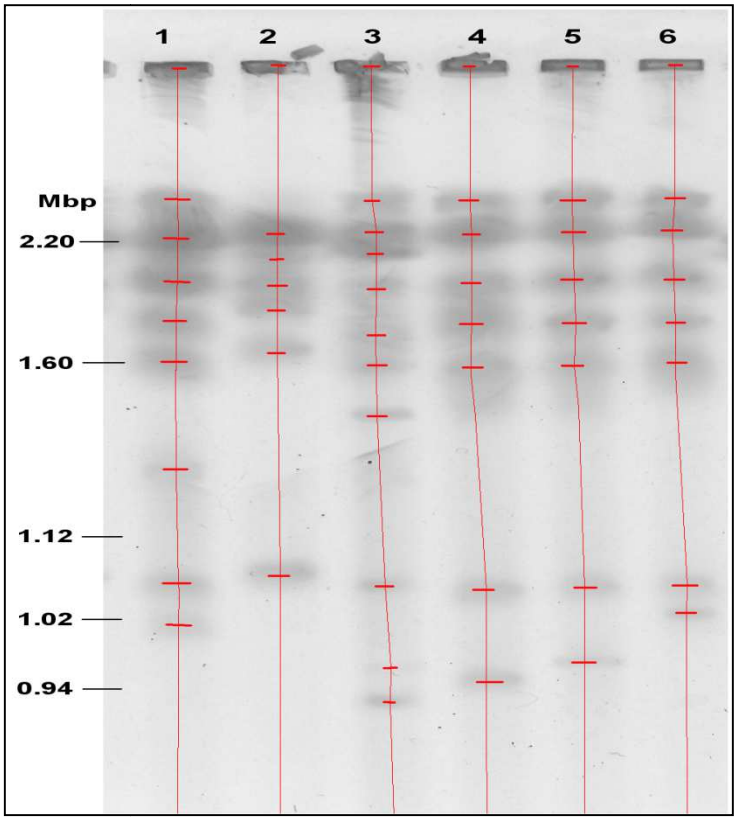


Figure 3.8. Chromosome separation in *Zygosaccharomyces rouxii* complex. PFGE of 6 strains shows a large variation in the numbers and sizes of chromosome bands. (1) *Zygosaccharomyces* sp. ATCC 42981; (2) *Z. rouxii* CBS 732^{T}; (3) *Z. sapae* ABT301^{T}; (4) *Zygosaccharomyces* sp. OUT7136; (5) *Zygosaccharomyces* sp. CBS 4837; (6) *Zygosaccharomyces* sp. CBS 4838. The vertical lines indicate one lane, and the larger horizontal lines indicate a chromosome band. The chromosomal size ladder (*Saccharomyces cerevisiae* S288C, Bio-Rad Laboratories) is in megabase pairs (Mb).

Strains	MFI of G_0/G_1 cells ±SD	DNA In_1	DNA In_2	Genome size±SD (Mb)	Chr N°	Σ PFGE bands (Mb)	Ploidy ratio
S. cerevisiae							
BY4742	103.3±3.0	1.00	1.03	13.1*	16	11.68[c]	1.12
BY4743	198.0±4.03	2.00	2.06	26.2‡	16	-	2.24
Z. rouxii relatives							
OUT7136	141.80±3.39	1.4	1.5	19.57±0.47	8	11.27	1.74
CBS 4837	164.50±2.40	1.59	1.77	22.70±0.33	8	11.56	1.96
CBS 4838	158.95±1.48	1.54	1.71	21.94±0.20	8	11.52	1.90

Genome sizes of *S. cerevisiae* *haploid and ‡diploid strains include repeated rRNA gene sequences (Goffeau et al., 1996). †Addition of PFGE bands reported for haploid strain YNN295 used as CHEF DNA size standard (Bio-Rad). Genomic size was expressed in megabase (Mb) as means of two replicas ±standard deviation (SD).

Table 3.3. Genome size, chromosome number and ploidy ratio in strains of the mosaic lineage.

3.4. Discussion

By evaluating genomic properties and nuclear and mitochondrial markers, we recognized two lineages well-separated from the *Z. rouxii* strain CBS 732^{T}, namely *Z. sapae* and the mosaic lineage (**Table 3.4**). With the exception of strain OUT7136, the strains considered in this study have an unusual level of heterogeneity in rRNA gene regions, but the pattern of this heterogeneity varies significantly between *Z. sapae* and the mosaic lineage (**Table 3.4**).

The co-occurrence of rRNA gene variants in the genome of a single individual suggests relaxation of concerted evolution, a recombination-driven process that is responsible for homogenizing rRNA gene repeats (Birky, 1996). Relaxation was demonstrated to occur in transition stages of concerted evolution, when location of rRNA gene loci on nonhomologous chromosomes potentially disrupts concerted evolution, when organisms are polyploid, as a result of interspecific postulated hybridization, or when the mutation rate exceeds the rate of concerted evolution, as in length variants in the intergenic spacer (Paun et al., 2007; Albertin and Marullo, 2012). In particular, the concerted evolution model assumed that the co-evolving genes have to be organized in tandem arrays. A previous study showed that the repeats of the 5S ribosomal genes of filamentous fungi species are dispersed among the genomes and are escaping the concerted evolution model (Rooney and Ward, 2005). The Birth-and-Death evolution model was more appropriate to describe the dynamics of rRNA genes not organized in tandem arrays (Nei and Rooney, 2005). In that model, the genetic evolution is regulated by a balance between gene duplication, turnover and maintenance, which allows the occurrence of several haplotypes within the same genome.

As of today, nothing is known about the organizational pattern of the ribosomal genes in *Z. sapae* and the mosaic lineage and therefore we cannot exclude that the Birth-and-Death evolution model could explain the heterogeneity detected in the present study. Intra-individual variations in

rRNA gene regions combined with the presence of divergent pairs of *ZrSOD* and *HIS3* paralogues may be the expression either of a duplicated gene/genome structure, which preceded the divergence of the lineages, or of a postulated hybridization event (Birky, 1996). Analysis of these nuclear housekeeping markers demonstrated that two homologues for each gene are present in diploid and aneuploid strains and that the orthologous sequences originating from one of the paralogous types are more similar in different strains to each other than to the sequences of the other paralogous type. Moreover, multiple lines of evidence demonstrated that *Z. sapae* and the mosaic lineage include diploid and aneuploid strains, in contrast to the haploid state reported for strains CBS 4837 and CBS 4838 in the early literature (Wickerham and Burton, 1960). These findings led us to suppose an evolutionary model in which the presumptive diploid-like status originated prior to sorting of *Z. sapae* and the mosaic lineage into different groups.

Genome polyploidization has been detected in many asexual taxa (Otto and Whitton, 2000) and it is an effective way to buffer the genome against the effects of accumulating mutations (Kondrashov, 1994). Interestingly, putative asexual fungal species seem to have a propensity for within-individual rRNA gene variation (Simon and Weiß, 2008). In predominantly asexual yeasts, such as *C. albicans* (Rustchenko, 2007), Candida *glabrata* (Muller et al., 2009), *Dekkera bruxellensis* (Hellborg and Piškur, 2009), and *D. hansenii* (Jacques et al., 2010), pronounced karyotype variability is common and a relaxed control over the chromosome structure has been supposed to increase the genome variability and competitiveness (Poláková et al., 2009). Significantly, chromosome number and genome size vary significantly also within *Z. rouxii* complex. This chromosome polymorphism makes it difficult to believe that these strains regularly undergo meiotic recombination. The variability in genome organization may ‘isolate’ the strains and prevent them from successful recombination, if meiosis occurs. Furthermore, in *Zygosaccharomyces*, mating between heterothallic haploid vegetative cells should precede meiosis and sporulation (Wickerham and Burton, 1960). The

reproductive mode, outlined in **Figure 3.9** panel a and b, precludes the process of genome renewal by self-diploidization occurring in *S. cerevisiae*, and allows for reticulation.

Under this hypothetical evolutionary scenario, mating between two divergent haploid cells would have resulted in a diploid ancestor, which lost meiosis ability and gave rise to stable diploid lineages (namely *Z. sapae* and the mosaic lineage), which clonally reproduce and independently evolve (**Figure. 3.9c, I**). The sequence divergence between variants of rRNA gene gene and housekeeping marker within an individual, as well as chromosome variability, supported such suggestion. However, evidence of rare conjugated asci in *Z. sapae* (Solieri et al., 2013) and successful mating in laboratory conditions between sibling heterothallic strains CBS 4837 and CBS 4838 (Mori and Onishi, 1967) has been reported. These observations gave rise two hypotheses. The first is that mating between diploid cells results either in a transient tetraploid zygote that undergoes meiosis or in a massive chromosome loss in order to restore the diploid status (**Figure 3.9, panel c, II**), a mechanism similar to what occurs during the parasexual cycle of *C. albicans* (Hull et al., 2000). The second is that the resulting transient tetraploid zygote is dikaryotic and the nuclei segregate into mitotic spores without karyogamy both in conjugated and unconjugated asci (**Figure. 3.9, panel c, III and IV**, respectively). In *Zygosaccharomyces bailii*, a close relative of *Z. rouxii*, absence of nuclear fusion and, as a result, the formation of mitotic spores has been reported (Rodrigues et al., 2003). The occurrence of binucleate cells in late stages of the *Z. bailii* life cycle was further confirmed by Dato et al. (2008). Similarly, *M. farinosa* allodiploid strains did not undergo meiosis, producing mainly diploid spores (Mallet et al., 2012). The abundance of *Z. sapae* cells with conjugation tubes and without zygote supported the second rather than the first hypothesis: mating between diploid cells might occur without nuclear fusion and meiosis.

Properties	CBS 732^T	*Z. sapae*		Mosaic lineage			
		ABT301^T	ABT601	OUT7136	CBS 4838	CBS 4837	ATCC 42981
DNA content	12.7	28.1	39.0	19.57	22.5	21.7	21.9
Ploidy	haploid	diploid	Diploid	aneupolid	diploid	Diploid	diploid
Chr. N.	6	10	11	8	8	8	7
ZSOD2	*ZrSOD2-22*	*ZrSOD2-22-ZrSOD22*	*ZrSOD2-22-ZrSOD22*	*ZrSOD22-ZrSOD2*	*ZrSOD22-ZrSOD2*	*ZrSOD22-ZrSOD2*	*ZrSOD2-ZrSOD2*
HIS3	Zr	2(Zr+Zp)	2(Zr+Zp)	2(Zr+Zp)	2(Zr+Zp)	2(Zr+Zp)	2(Zr+Zp)
5.8S-ITS	Zr	2(Zr+Zp)+1‡	2(Zr+Zp)+1	Zp	2(Zr+Zp)+1	2(Zr+Zp)+1	2(Zr+Zp)
26S D1-D2	Zr	Zp	Zp	Zp	2(Zr+Zp)	2(Zr+Zp)	2(Zr+Zp)
COX2							

Zr, *Zygosaccharomyces rouxii*-like copy; Zs, *Zygosaccharomyces sapae*-like copy.
*Genome size in Mb. †Genome size of CBS 732^T was estimated at 9.8 Mb as result of final assembly of genome project (Souciet et al., 2009) and of 12.7 Mb according to PFGE determination (Solieri et al., 2008a). ‡Additional recombinant copy.

Table 3.4. Overview of the main molecular and genetic properties of strains belonging to *Z. rouxii* complex.

Finally, the extent of rRNA gene variants within *Z. rouxii* complex, has evolutionary and diagnostic implications and points emphasizes the limit of classical phylogenetic analysis, which may lead eventually to erroneous trees. These phenomena create a network of paralogous sequence relationships potentially confounding accurate phylogenetic reconstruction. The question has been addressed by other authors (Mallet, 2007; Liti et al., 2006; Wu et al., 2008; Casaregola et al., 2011), who reported how ploidy and heterozygosity assessments, as well as phylogenetic networks and incorporation of population genetics into phylogenetic analysis, are pivotal tools to reveal mosaic species.

Overall, our results indicate that the group of strains included in this work is a complex of haploid and diploid heterogeneous species, including: (1) the haploid species *Z. rouxii*, e.g. CBS 732^T; (2) the diploid species *Z. sapae*, isolated from high sugar environments; (3) a diploid mosaic lineage that includes strains retrieved from salt environment. We hypothesize that the alternative mode of propagation to classical sexuality described in *Saccharomyces* species can account for this

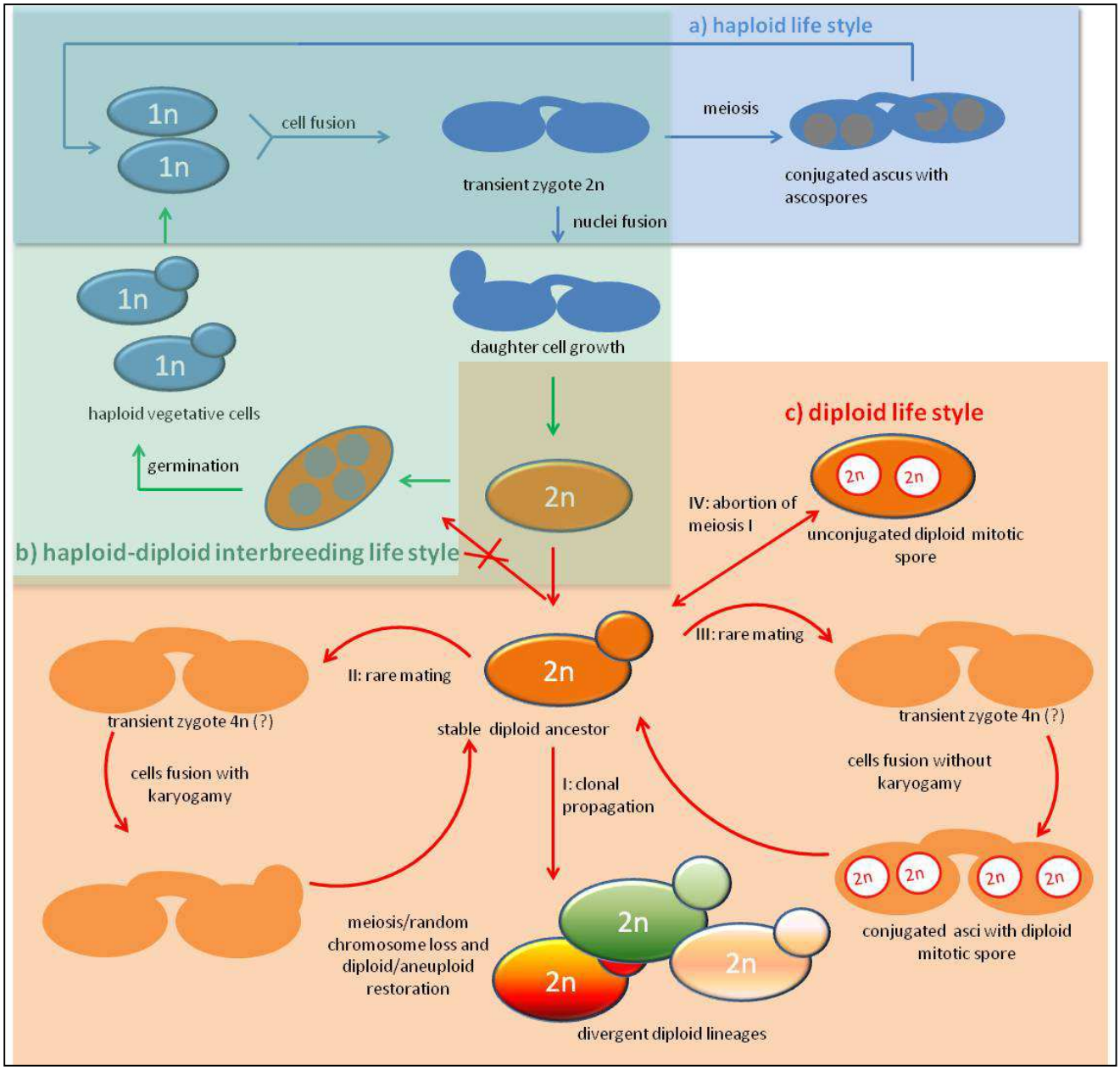


Figure 3.9. A population genetic perspective of the life cycle of *Z. rouxii* complex (modified from Mori, 1973). (a) In the haploid *Z. rouxii* life cycle, haploid vegetative cells reproduce by budding, giving rise to mother and daughter cells. Mating may occur between haploids and leads to the formation of a transient heterokaryotic zygote with the 'dumbbell' configuration typical of *Zygosaccharomyces* clade. The diploid zygote undergoes meiosis and sporulation in order to restore the haploid state. (b) Alternatively, cell mating is followed by nuclei fusion, resulting in a diploid zygote able both to reproduce by clonal propagation and to switch to haploid phase by meiosis. In the latter case, unconjugated asci should be produced (haploid–diploid interbreeding life style). (c) Lastly, coalescence of divergent diploid lineages results from fusion of nuclei from two haploid cells, giving rise to a stable diploid zygote, which reproduces permanently by clonal propagation (I). If mating occurs, transient heterokaryotic tetraploid zygote can perform nuclei fusion followed by either

meiosis or massive chromosome loss to restore the diploid state (II) or originate conjugated asci with mitotic spores (III). Abortion of the reductional division (meiosis I) can produce unconjugated asci harbouring mitotic diploid spores (IV).

genomic complexity and makes *Z. rouxii* complex prone to genome mosaicism and reticulate evolution. An increasing amount of evidence has shown that genome mosaicism is widespread in protoploid hemiascomycetes other than *Z. rouxii* complex, such as *D. hansenii* (Jacques et al., 2010), *P. sorbitophila* (Louis et al., 2012) and *M. farinosa* (Mallet et al., 2012). Notably, all these yeasts are able to survive to harsh conditions, such as a high extracellular concentration of osmolytes, suggesting that this genome complexity may occur in response to environmental cues (Gompert et al., 2006).

Chapter 4:

Sex determination system in diploid yeast *Zygosaccharomyces sapae*

4.1. Introduction

Sexual reproduction is ubiquitous in eukaryotic organisms, from yeasts to human (Hadany and Comeron, 2008). Sexual reproduction and breeding systems are driving forces for genetic diversity. The mating-type (*MAT*) locus represents a mutation and chromosome rearrangement hotspot in yeasts. *Zygosaccharomyces rouxii* complex yeasts are naturally faced with hostile low a_w environments and are characterized by gene copy number variation, genomic instability, and aneuploidy/allodiploidy. Hemiascomycetes in particular have evolved a homothallic and heterothallic repertoire of bipolar mating strategies orchestrated by a single *MAT* locus, which encodes key transcription factors that govern sexual identity and compatibility (Fraser and Heitman, 2003). In contrast, other yeasts, such as *Candida albicans*, have developed alternative cryptic sexual cycle governated by a same-sex mating. The variability in mating systems and sex chromosomes may drastically affect population genetic structure, pathogen evolution, and ecological processes of survival and adaptation (Fraser et al., 2005; Bubnick and Smulian, 2007; Hsueh and Heitman, 2008), offering in-deep understanding of the factors that shape sex evolution, one of the a major challenges in biology (Billiard et al., 2012).

In the haplo-diplontic yeast *Saccharomyces cerevisiae* the *MAT* locus is located in centromeric region of chromosome III (CEN-*MAT* linkage) in two versions (idiomorphs), either *MAT***a** or *MATα* genes, enabling yeast to specify three cell types: haploid a, haploid α, and diploid a/α. In heterothallic strains of *S. cerevisiae*, mating takes place between cells bearing complementary *MAT* idiomorphs. However, *S. cerevisiae* exists in nature mainly as homothallic diploid strains (Mortimer 2000; reviewed in Greig and Leu, 2009), and sexually reproduces in

clonal cell population by meiosis followed by mother–daughter mating (also referred as haplo-selfing) (Knop, 2006). A cassette model for mating-type switching has been proposed (Hicks et al., 1977; Herskowitz et al., 1988) and further experimentally verified to explain haplo-selfing in *S. cerevisiae*. Mating-type switching is a programmed DNA rearrangement process that occurs in haploid budded cells and converts a *MAT***a** allele into a *MATα* idiomorphs, or vice versa (Strathern et al., 1982; Haber, 1998). During switching, DNA at the *MAT* locus is removed and replaced with DNA copied from the heterochromatic silent cassettes near the telomeres of the chromosome III, either *HML* or *HMR*. The gene conversion is mediated by a HO endonuclease, which catalyzes a site-specific double stand break (DSB) at the boundary between the Y sequences unique to the *MAT*α or *MAT***a** alleles and the shared flanking Z sequences (reviewed in Haber, 2012).

Based on comparative genomic analyses, the HO-catalyzed homothallic switching in the family Saccharomycetaceae arose from an obligate heterothallic ancestor system via a two-step process: i) the origin of the silent cassettes (after the divergence of family Saccharomycetaceae from other families such as Debaryomycetaceae and the *Candida albicans* clade); ii) the recruitment of *HO* gene, after the occurrence of a whole genome duplication (WGD) event which split off the Saccharomycetaceae into the pre-WGD and post-WGD species, respectively (Wong et al., 2002; Butler et al., 2004). Despite the conservation of *HML* and *MAT* in *cis*, and of the α genotype at *HML*, the family Saccharomycetaceae displays consistent variability in ideomorph content and chromosomal organization of *MAT* locus (Tsong et al., 2003; Butler et al., 2004; Fabre et al., 2005; Gordon et al., 2011). Unlike *S. cerevisiae* and closest relatives, *HMR* is not constrained to be linked to *MAT* and *HML* loci on sex chromosome in other yeasts (Fabre et al., 2005). Moreover, the *S. cerevisiae MAT* loci code for only three proteins (the homeodomain proteins a1 and α2 and the ‘‘α-domain’’ protein α1), while an additional gene (*MAT***a**2) coding for an HMG domain DNA-binding protein is present in the *MAT***a** idiomorph of several species (Butler et al., 2004). Almost all the pre-WGD species retain stable chromosomal organization with a restricted set

of ancestrally conserved genes flanking *MAT* locus. On the contrary, in post-WGD species the *MAT* locus is subjected to a continual process of erosion leading different genes incorporated into the Z and X regions, making the sex chromosome an hotpsot for deletion and transposition (Martin et al., 2010; Gordon et al., 2011).

The protoploid *Zygosaccharomyces rouxii* is one of the few pre-WGD species that split off from post-WGD species after the gain of *HO* gene (Butler et al., 2004). *Z. rouxii* strains commonly inhabit low a_w environments and have been used for centuries as fermented food starters for the production of sugary and salty food, but they can also act as food spoilager accounting huge economical loss to food industry (Dakal et al., 2014). *Z. rouxii* has traditionally been considered as predominantly haploid yeast with a bipolar mating system (Wickerman and Burton, 1960). As sporulation requires a diploid DNA content, the species with a haploid lifestyle, such as *Z. rouxii*, must first undergo mating between heterothallic a and α cells in response to osmo-stress. The resulting transient a/α diploid zygote usually enters in meiosis, producing from two to four haploid gametes. Syngamy of homothallic strains is also possible between genetically identical haploid cells by mating type switching, followed by meiosis to restore the haploid status. Remarkably, alternative modes of reproduction have been observed but poorly investigated. Cell fusion could be not followed by nuclear fusion, resulting in a dikaryon which produces haploid buds (Mori, 1973). Zygote may lose the meiotic ability and begins clonal euploid/aneuploid lineages (Solieri et al., 2013a). Recently, *Z. rouxii* in yeast culture collections have been demonstrated considerable variation in ploidy and karyotype (James *et al*. 2005; Gordon and Wolfe 2008; Solieri et al., 2007, 2013a, b), that correspond with phenotypic variability to survive under stress cues (Solieri et al., 2014a). Based on these evidences, at least three groups have been delineated and globally referred to as *Z. rouxii* complex: the group of haploid *Z. rouxii* including the strain CBS 732^T, an allopolyploid group composed of strain ATCC 42981 and aneuploid relatives, and the novel diploid species *Zygosaccharomyces sapae*, which display mainly a clonal reproduction and rarely goes

through meiosis resulting in ascospores (Gordon and Wolfe, 2008; Solieri et al., 2013a, b). The co-existence in the same phylogenetic group of very closely related species of sexual and putative asexual taxa with similar ecological and physiological properties raises several questions: (i) is *Z. sapae* truly asexual having thus no traces of *MAT* genes in their genomes? (ii) alternatively, has asexual species formed recently and therefore still exhibit unfunctional sex related genes? (iii) is mating type imbalance possibly responsible for asexual lineages? Recently, the analysis of the *MAT* structure in haploid strains belonging to *Z. rouxii* lineages revealed a remarkable rearrangement of sex chromosome by ectopic recombination, leading to strains with unusual genetic make-up ααα and αααα (Watanabe et al., 2013). These evidences support that sex chromosome is prone to non-homologous recombination in *Z. rouxii* species complex. However, no evidences about the *MAT* loci organization have been reported in diploid lineages. In this study, we surveyed the presence and integrity of *MAT* and *HO* genes in *Z. sapae* diploid type strain ABT301^T.

4.2. Materials and Methods

4.2.1. Strains and mating tests

The *Z. sapae* ABT301^T strain was retrieved from high sugary Traditional balsamic vinegar (Solieri et al., 2006, 2013b) and deposited to the Yeast Collection of the Centraalbureau voor Schimmelcultures (CBS; Utrecht, The Netherlands) and to the Mycothéque de l'Université catholique de Louvain (MUCL; Louvain-la-Neuve, Belgium) under the codes CBS 12607^T and MUCL 54092^T, respectively. Strains CBS 732^T, CBS 4837 (mating type a) and 4838 (mating type α) were achieved from CBS collection. Strains were cultured and maintained in the yeast extract-peptone-glucose (YPG) medium (1.0% yeast extract, 1.0% peptone and 2.0% glucose, w/v). To study sexual compatibility, 2-4 day-old culture of ABT301^T mixed to CBS 4837 or CBS 4838 both on malt extract agar (MEA) (Difco) and MEA supplemented with 6% (w/v) NaCl, incubated at 27° for 1 week and examined microscopically using phase-contrast optics for production of conjugated asci.

4.2.2. DNA manipulation

Genomic DNA (gDNA) was extracted from early stationary cultures using the phenol/chloroform method (Hoffman and Winston, 1987). The restriction enzymes were purchased from Fermentas (Burlington, ON, Canada), rTAQ DNA polymerase and high-fidelity Phusion DNA polymerase from Takara (Takara Bio Inc., Shiga, Japan) and ThermoFisher Scientific (Waltham, MA), respectively, DNA ligation kit from Promega (Madison, WI). Plasmid preparations, PCR reactions, and other standard molecular biology techniques were carried out as described elsewhere (Sambrook et al., 1989) or as instructed by suppliers. In particular, standard PCR mixtures (25–50 µl) contained 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 200 µM of each dNTP, 0.4 µM of each primer, 0.02 U/µl of rTaq DNA polymerase and 100-200 ng of template DNA. The thermal program consisted of one cycle of 1 min at 94° followed by 35 to 40 cycles of 94° for 45 sec, 58° for 1 min, and 72° for 2 min. For amplification of DNA fragments > 2 kb, PCR mixtures (20 µl) contained 1X Phusion HF Buffer, 200 µM of each dNTP, 0.5 µM of each primer, 0.02 U/µl of Phusion DNA polymerase, and 100-200 ng of template DNA. The thermal program consisted of 1 cycle of 98° for 1 min, 25-35 cycles of 98° for 10 sec, 60-68° for 30 sec, 72° for 30sec/kb, followed by 1 cycle of 72 ° for 10 min. All the PCR reactions were performed with BioRad T100 Thermalcycler (Bio-Rad Laboratories, Hercules, CA). Primer design was carried out using the Primer3 software (Untergasser et al., 2012). Screening of cloning libraries containing PCR products from degenerate primers were performed by sequencing at least three plasmids. All the sequencing reactions were carried out by a custom sequencing service provider (BMR Genomics, Padova, Italy)

4.2.3. Cloning of *MAT* loci

Schematic strategy of *MAT* idiomorphs cloning is outlined in **Figure 4.1**. Briefly, degenerate primers were designed based on a set of amino acid sequences that represents highly conserved regions of homologous proteins MAT**a**1, MATα1, and MATα2 from the species *S.*

cerevisiae, and *Z. rouxii* (**Table 4.1**). These degenerate primers pairs were used to PCR amplify similar conserved regions of *Z. sapae* gDNA. Individual gel bands from amplified PCR products showing predicted size were gel-extracted by using the QIAquick column method (Qiagen) and cloned into pGEM-T Easy (Promega Corporation). At least three plasmids with inserts of expected size were sequenced in both directions with primers T7 and SP6. Protein-coding regions were determined using BLAST search (Altschul et al., 1997).

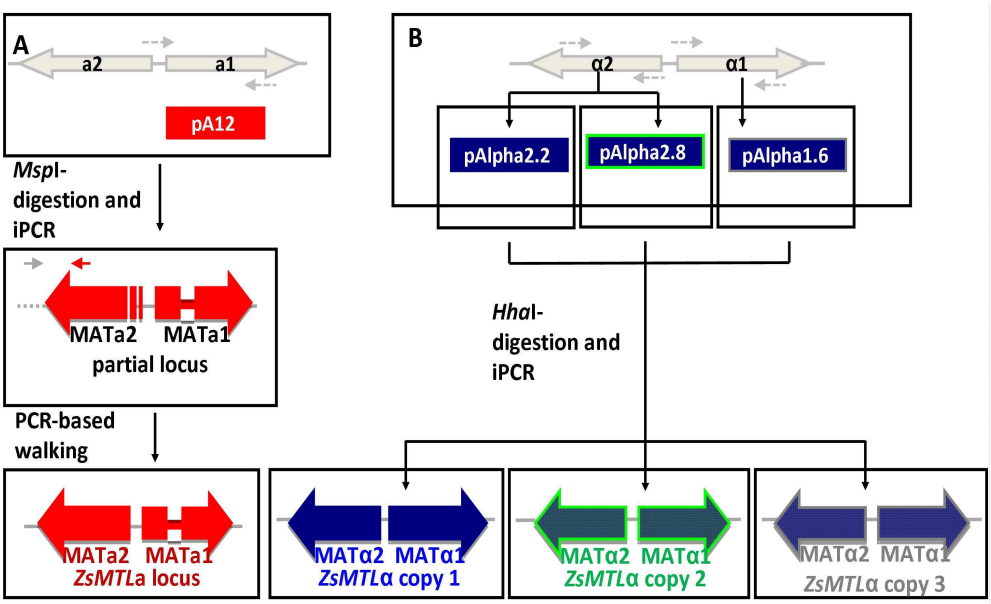


Figure 4.1. Schematic outline of *Z. sapae* mating type loci cloning strategy. Small arrows indicate primers; small dotted arrows indicate degenerate primers. Dotted lines represent unknown sequences. *ZsMTL*a and *ZsMTL*α loci are colored in red and blue, respectively. Divergent copies of *ZsMTL*α are surrounded by blue (1), green (2), and gray (3), respectively.

The *MAT* sequences were further extended by inverse PCR (iPCR). Briefly, gDNA (200 ng) was digested with *Hha*I and *Msp*I for sequencing *MAT*α and *MAT***a** loci, respectively. DNA digests were ligated with T4 DNA ligase (Promega Corporation) and the digestion/ligation products were 10 fold diluted and 1 µl was used for 25 µl PCR reaction using rTaq polymerase (Takara Bio. Inc.) and the primers listed in **Table 4.2**. To complete *MAT***a**2 sequence, the primer 301_MATA2F1, spanning the 5'UTR region of *MAT***a**2 open reading frames (ORFs) in *Z. rouxii* ZYRO0C18326g locus, was used together with a internal *MAT***a**2-specific primer (301_MATA2R1) in PCR reaction (**Table 4.2**).

Protein	Domain (N-C-terminal ends)	Primer	Sequence (5'-3')	PCR strategy	Plasmid	Insert length (bp)
MATa1	MRNCKG	Zr-MATA1F1	ATG MGi AAY TGY AAR GGN AA	Zr-MATA1F1/Zr-MATA1R1	pA1.12	153
	PIQVRIW	Zr-MATA1R1	CCADATNCKiACYTGDATNGG			
MATα1	NSFMAF	Zr-MATalpha1-F2	AAY WSi TTY ATG GCN TTY	Zr-MATalpha1-F2/Zr_MATalpha1_R1 followed by Zr-MATalpha1-F2/Zr-MATalpha1-R2	pAlpha1.6	495
	WDTFAQQFN	Zr_MATalpha1_R1	RAAYTGYTGNGCRAANGTRTCCCA			
	CGFVEWV	Zr-MATalpha1-R2	CCCAYTCNACRAANCCRCA			
MATα2	MNKIPIE	Zr_MATALPHA2_F1	ATGAAYAARATHCCNATHGAR	Zr_MATALPHA2_F1/Zr_MATALPHA2_R1	pAlpha2.2	578
	PQIKNWV	Zr_MATALPHA2_R1	ACCCARTTYTTDATYTGNGG		pAlpha2.8	578
HO	55KHRAFEGE62	ZrHO_F2	AARCAYMGNGCNTTYGARGGNGA	ZrHO_F2/ZrHO_R2	pHO2.3	258
	133DFPMTPEG140	ZrHO_R2	ACCYTCNGGNGTCATNGGRAARTC		pHO2.8	258
	284LRKNNPFW292	ZrHO_F3	YTNMGNAARAAYAAYCCiTTYTGG	ZrHO_F3/ZrHO_R3	pHO3.5	147
	324FLAGLIDSD332	ZrHO_R3	TCNSWRTCDATNARNCCNGCiARRAA			

Table 4.1. **Degenerate primers used in this study.**

Target	Primer code	Sequence (5'-3')	Description
ZsMTLα	301_MATa1F1	CCAAGAACTCTCGAAGAAGCTG	Primer specific for plasmid pAlpha1.6 and used to extend MATα1p-like coding sequenceby *Hha*I-iPCR
	301_MATa1R1	GGCGGTGATGGAATCTTAGT	Primer specific for plasmid pAlpha1.6 and used to extend MATα1p-like coding sequenceby *Hha*I-iPCR
	301_MATa1F2	GTTCGGAGAAGCCACTCAATTC	Primer specific for plasmid pAlpha1.6 and used to extend MATα1p-like coding sequenceby *Hha*I-iPCR
	301_MATa1R2	TCATCCGCTATACACTTCCC	Primer specific for plasmid pAlpha1.6 and used to extend MATα1p-like coding sequenceby *Hha*I-iPCR
	301_MATa2F4	ATGGAGACTAAGTTATCGGGACC	Primer specific for plasmid pAlpha2.2 and used to extend MATα2p-like coding sequence by *Hha*I-iPCR
	301_MATa2R4	CTCTTGATGTACTGGGTTGAGC	Primer specific for plasmid pAlpha2.2 and used to extend MATα2p-like coding sequence by *Hha*I-iPCR
	301_MATa2F2	GAGCATCCTTACCTGCAAAC	Primer specific for plasmid pAlpha2.8 and used to extend MATα2p-like coding sequence by *Hha*I-iPCR
	301_MATa2R2	GGAAGTATCTTCTCTAGCTCTGC	Primer specific for plasmid pAlpha2.8 and used to extend MATα2p-like coding sequence by *Hha*I-iPCR
*ZsMTL***a**	301_MATA2F1	GCAACATGGTCATGGTCAAC	Primer specific for plasmid pA12 and used to extend MATa1p-like coding sequence by *Msp*I-iPCR
	301_MATA2R1	TGAAGAGCACTGGCATCTAAA	Primer specific for plasmid pA12 and used to extend MATa1p-like coding sequence by *Msp*I-iPCR
	301_MATA1F2	GCTGTAGCAGCTAATTGTGG	Primer specific for plasmid pA12 and used to extend MATa1p-like coding sequence by *Msp*I-iPCR
	301_MATA1R2	CTCTTTCTCTCAAATACACGTTC	Primer specific for plasmid pA12 and used to extend MATa1p-like coding sequence by *Msp*I-iPCR
	301_MATA2F2	ACA GGT CTT CGA CGT TTA GC	Primer used for PCR-based 5' walking of *ZsMTLa* locus
	301_MATA2R2	CAT GTG TCT GCA ATC ACT TCA C	Primer used for PCR-based 5' walking of *ZsMTL*a locus
HO	301_5'HOF1	CTACGTCGAGAGATCCATCATAG	primer specific for plasmid pHO2.3 used in combination with 301_5'HOR1
	301_5'HOF3	TCAGTGGCACATCAGCTT	primer specific for plasmid pHO2.8 used in combination with 301_5'HOR1
	301_5'HOR1	GCTTCACGCACCTGTAAATC	primer specific for plasmid pHO3.5
	UpHOCBS732F2	ACGAGTGGTGGTGGGATAGACTTA	primer targeting 5' UTR of *Z. rouxii* CBS 732 *HO* gene; used for 5' PCR walking
	301_verylikeHOR3	CGCGAATCTACCGGTACTATT	*HO* copy 1-specific primer used in combination to UpHOCBS732F2; used for 5' PCR walking
	301_likeHOR3	CTACAAACCTACCGGTGTTAGA	*HO* copy 2-specific primer used in combination to UpHOCBS732F2; used for 5' PCR walking
	ZrHO_R5	CCNSWCCARTCNCKRTARAARTA	primer targeting the aa domain FYRDWSG at the C-terminal of *Z. rouxii* Ho; used for 3' PCR walking
	DownHOCBS732R1	TCACCAAGGCTATGTCTTCTCGCT	primer targeting 3' UTR of *Z. rouxii* CBS 732 *HO* gene; used for 3' PCR walking
	301_very_likeHOF5	TGTGATGGACATCGCAGAAATCGC	*HO* copy 1-specific primer used in combination to ZrHO_R5; used for 3' PCR walking
	301_very_likeHOF7	TGCATGCGGTGATCATTGTAAGGC	*HO* copy 1-specific primer used in combination to DownHOCBS732R1; used for 3' PCR walking
	301_likeHOF5	GGACATCGTAGAAACCGCCATTTG	*HO* copy 2-specific primer used in combination to ZrHO_R5; used for 3' PCR walking
	301_likeHOF7	ATGTTGTGGGCGTAACAGTTG	*HO* copy 2-specific primer used in combination to DownHOCBS732R1; used for 3' PCR walking

Table 4.2. List of primer used for iPCR and PCR walking of *ZsMTL* loci and *HO* genes.

4.2.4. Cassette system determination

To verify if gene organizations around the *MAT* loci in *Z. sapae* resembles those described in other yeast species (Butler et al., 2004; Watanabe et al., 2013), PCR amplification of gDNA was carried out by using primer sets spanning putative *MAT*-flanking genes (**Table 4.3**). Briefly, the first round of long-range PCR reaction was done with high fidelity DNA polymerase (Phusion, ThermoFisher Scientific) and the external primers 1, 2, 3, A, B, B', C (Watanabe et al., 2013), and DownMATa1R1 (this study) in 20 µl reaction volume, following the manufacturer's instructions. Subsequently, a semi-nested PCR amplification was done using a 1:20 dilution of the previous PCR and nested *MAT*-specific primers. In case of negative results in first round of PCR reactions, we tested alternative combinations of MAT flanking genes, by direct PCR amplifications from gDNA with the following primers sets: 1, 2, 3/reverse nested *MAT*-specific primer (for 5' end flanking genes) and forward nested *MAT*-specific primers/A, B, B', C, and DownMATa1R1 (for 3' end flanking genes). Sequencing of purified PCR products were carried out as reported above.

4.2.5. Cloning of *HO* genes

Homology comparison among the HO proteins from the species *S. cerevisiae* (AAA34683; NP_010054) and *Z. rouxii* (ZYRO0C10428p), as well as the *S. cerevisiae* VMA intein (AAL18609) was performed by Clustal-W2 alignment (Larkin et al., 2007) to identify four highly conserved amino acid sequences. Relied on these conserved motifs, two degenerate primer pairs, ZrHOF2/ZrHO_R2 and ZrHOF3/ZrHO_R3 were designed and used to amplify the N- and C-terminal coding regions of the putative *Z. sapae HO* gene, respectively. PCR fragments of expected length were gel extracted and cloned as previously reported. The plasmids pHO2.3 and pHO2.8 bearing two inserts coding for putative HO N-terminal portions, and pHO3.5 containing an insert covering the HO C-terminal portion were identified by sequencing in both directions. Genomic portions cloned in pHO2.3 and pHO3.5 and in pHO2.8 and pHO3.5 were joined by PCR

amplification and sequenced with primers pairs 301_5'HOF1/301_5'HOR1 and 301_5'HOF3/301_5'HOR1, respectively (**Table 4.2**). The resulting two partial *HO* contigs were referred as copy 1 and copy 2. Subsequently, the full-length ORF sequences of *HO* copies 1 and 2 were achieved by PCR-based walking. For upstream walking, a forward primer targeting 5' UTR of *Z. rouxii* CBS 732^{T} *HO* gene (ZYRO0C10428g) was combined with two *HO* copy specific reverse primers (**Table 4.2**). The sequences flanking the 3 ends of both copies were covered through a two-steps PCR walking strategy. In the first step, *HO* copy specific forward primers were combined with degenerate reverse primer targeting the HO conserved amino acid domain FYRDWSG. In the second one, forward *HO* copy specific primers were exploited together with a downstream reverse primer, targeting 3' UTR of *Z. rouxii HO* gene (**Table 4.2**). Schematic outline of cloning strategy was reported in **Figure 4.2**.

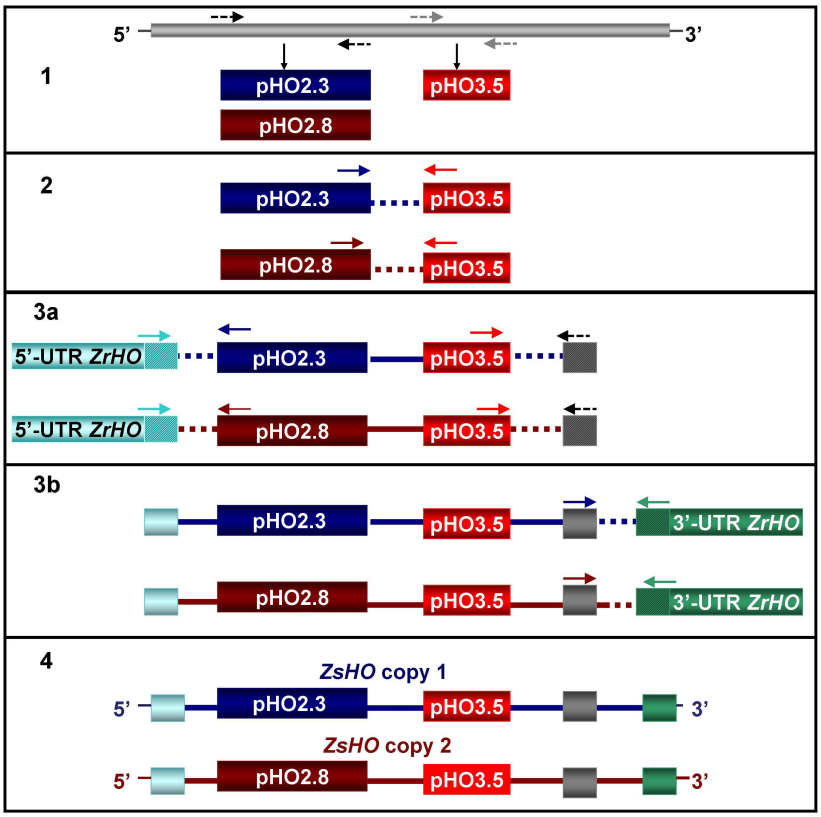


Figure 4.2. Outlined strategy used for cloning *Zygosaccharomyces sapae HO* genes. Numbers from 1 to 4 indicate the cloning and PCR walking steps. Dotted arrows represent degenerate primers and dotted lines undetermined sequences. Abbreviation: *ZrHO*, *Zygorsaccharomyces rouxii HO* gene; *ZsHO*, *Zygorsaccharomyces sapae HO* gene. Plasmid names according to **Table 4.1**.

Locus	PCR walking	Primer	Sequence (5'->3')	Description	Reference
MATα1	3' PCR walking	DownMATa1R1	TTYGARTTYTAYGCNGAYTG	degenerate primer targeting FEFYADC conserved aa sequence of *Z. rouxii* CBS 732^{T} *SLA2* gene	this study
		301_MATa1F1	CCAAGAACTCTCGAAGAAGCTG		this study
		301_MATa1F2	GTTCGGAGAAGCCACTCAATTC		this study
		301_MATa1rouF1	CCGCCGAAGAATTTACTTAGAG	MATα1 *Z. rouxii*-specific forward primer	this study
		301MATa1a2-cp1F1	TTCCTTCACCGCCAGAGGTTC	MATα1 copy 1-specific forward primer	this study
		301MATa1a2-cp2F3	TTCCTTCACCTCCGGAGAACC	MATα1 copy 2-specific forward primer	this study
		A	CCAGTTAGTGTGTTATCGATAAGTC		Watanabe et al., 2013
		B	TCTATTTCGTCCGTTTATCGTTGGT		Watanabe et al., 2013
		B'	CAGAGACTAATAATGAGAGAAAAGC		Watanabe et al., 2013
		C	TCAGTACCAGAAGTGGTCTTTGAAA		Watanabe et al., 2013
MATα2	5' PCR-walking	301_MATa2RouR1	TTAGGAGATAAAGGTAAGAATAGG	MATα2 *Z. rouxii*-specific reverse primer	this study
		301-MATa2CP1R1	CTT GGT AAT ACA GGT AAA GAG GGT	MATα2 copy 1-specific reverse primer	this study
		301MATa1a2-cp2R1	GACACATTGCATTCTGTTAAACGT	MATα2 copy 2-specific reverse primer	this study
		1	GCTACTCCCTCATTAGAACATGAAA		Watanabe et al., 2013
		2	CGCATGATATGAAACGAAGATGCAA		Watanabe et al., 2013
		3	TACTTACTGGATGAATCTTCTGTGA		Watanabe et al., 2013

Table 4.3. List of primers used for cassette system determination.

4.2.6. PFGE and Southern blotting assays

Chromosomal DNA preparation in plug, gel preparation, and pulsed field gel electrophoresis (PFGE) were performed as previously reported (Solieri et al., 2008a). Chromosomal DNAs separated by PFGE were transferred onto a Hybond-N+ membrane (GE Healthcare, Buckinghamshire, UK) by upward capillary transfer. Experiments were performed using a PCR digoxigenin probe synthesis kit (Roche Diagnostics, Basel, Switzerland) and detection was carried out by chemiluminescence, using an antidigoxigenin antibody and CDP-star (Roche) according to the manufacturer's instructions (**refer section 7.2.6, Chapter 7**). Primers engaged in probe synthesis were listed in **Table 4.4**.

Target	Primer name	Sequence (5'→3')	Probe lenght (bp)
ZsMTLα1	301_MATa1F2	GTTCGGAGAAGCCACTCAATTC	329
	301_MATa1R3	GCTGGCACAAGCTTCTCAACTCTA	
ZsMTLa1	301_MATA1F3	GTAGCTTCCACAAGGTCTTCAAGG	585
	301_MATA1R3	GTGTCCAATCTACTTGTCAGACCCA	
HO	301_5'HOF4	CGCTGAGGACATCGATGAAA	631
	301_3'HOR2	TTCAAATTCACCACGCAGTTCC	

Table 4.4. Primers for probe synthesis used in PFGE-Southern blotting.

4.2.7. Sequence analysis, phylogenetic construction, and protein domain analysis

Database searches were run with the BLAST server at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST). Multiple sequence alignments were performed with the Clustal-W program at the European Molecular Biology Laboratory (http://www.ebi.ac.uk/clustalw). Subsequent adjustments of these alignments were done manually. Phylogenetic trees were constructed by the neighbor-joining (NJ) method using MEGA version 5.0 from Clustal-W alignment (Tamura et al., 2011). Bootstrap support was estimated using 1000 replicates for distance analysis. Statistics relating to the identification of Pfam domains of predicted proteins were obtained from version 27.0 of PFAM protein families database (Punta et al., 2012).

Sequences have been submitted to the EMBL/GenBank databases (Submission ID numbers: Hx2000037341, Hx2000037270, Hx2000037266).

4.3. Results

4.3.1. Mating tests

We first assessed the mating behavior of ABT301^T strain in pure and mixed cultures with the mating partners CBS 4837 (mating type a) and CBS 4838 (mating type α), respectively. Our previous observations show that ABT301^T formed rarely asci in pure culture on MEA medium, that involved mother and daughter cells which remained attached to each other (Solieri et al., 2013b). No conjugated asci were observed on 6%NaCl-MEA after 12 days of incubation. Strain ABT301^T showed no mating reaction with *MAT***a** CBS 4837 or or *MAT*α CBS 4838 tester strains, even after three weeks of incubation both on MEA and 6%NaCl-MEA media (data not showed), suggesting the homothallic state for ABT301^T or that ABT301^T doesn't respond to *Z. rouxii* pheromone signaling or that its pheromone expression may be repressed or defective.

4.3.2. Isolation of conserved domain of *Z. sapae MAT*α1 and *MAT*α2

To determine how mating type information is retained in *Z. sapae* genome, we cloned *MAT*α loci from ABT301^T. Based on amino acid sequence homology comparisons between *S. cerevisiae* and *Z. rouxii*, two highly conserved regions were identified for MATα1 and MATα2, respectively. Assuming that these sequences would also be conserved in *Z. sapae*, due to the high level of genetic relatedness with *Z. rouxii* (Solieri et al., 2013b), two degenerate primer pairs were designed for *MAT*α1 and *MAT*α2 ORFs, respectively. Using these primers and *Z. sapae* DNA, two distinct PCR products, of 578 bp and 495 bp, were detected. Sequence analysis of the cloning library revealed that inserts from recombinant plasmids pAlpha2.2, pAlpha2.8, and pAlpha1.6 contained partial ORFs sequences. A BLAST-based sequence similarity search revealed that the two 578-bp partial

ORFs were two divergent *MAT*α2 genes, while the 495-bp partial ORF was derived from a locus homologous to *Z. rouxii MAT*α1. In particular, pAlpha2.2 and pAlpha2.8 harbored two 578 bp inserts 81 and 99% identical to *Z. rouxii MAT*α2 gene (ZYRO0F15818g), respectively, whereas the 495 bp-insert of pAlpha1.6 shared 84% of identity with *Z. rouxii MAT*α1 gene (ZYRO0F15840g).

4.3.3. Characterization of *Z. sapae MTL*α loci

Chromosome walking by iPCR was used to further extend the *MAT*α idiomorph-like sequences, resulting in three *Z. sapae* mating type-like loci α, referred to as Zs*MTL*α locus copy 1, copy 2 and copy 3. Based on a BLAST-type search, two ORFs were predicted in each Zs*MTL*α locus, encoding proteins of 201 and 226 amino acid homologous to *Z. rouxii* MATα1 and MATα2, respectively, and separated by a intervening 343-bp intergenic sequence (**Figure 4.1**). All three Zs*MTL*α loci displayed an identical structural organization with the *MAT*α1 and *MAT*α2 genes orientated in opposite direction on complementary DNA strands, suggesting a configuration similar to those in *S. cerevisiae* and other hemiascomycetes (Butler et al., 2004). To establish the genomic location of *ZsMTL*α loci, we combined Southern blot analysis and PFGE-karyotyping. As previously reported (Solieri et al., 2008a), PFGE-Southern blotting failed to clearly resolve the highest molecular weight chromosomes spanning from 1.6 to 2.2 Mb, and labeled I, L and L', respectively (**Figure 4.2**). Hybridization of PFGE-Southern blot with a α-idiomorph specific probe resulted in a double band spanning from chromosome I and L, suggesting that *ZsMTL*α loci reside on at least two similar high molecular weight chromosomes difficult to discriminate.

*ZsMTL*α1 copies 2 and 3 diverged from the *ZsMTL*α1 copy 1 for 68 and 82 nt substitutions, respectively. The MATα1 like ORF from Zs*MTL*α locus copy 1 was 99.8% and 35.2% identical to *Z. rouxii* and *S. cerevisiae MAT*α1 genes, respectively, whereas the deduced 201-amino-acid sequence shared 99.5% overall identity with *Z. rouxii* MATα1. In Zs*MTL*α locus copy 2, the MATα1 like-coding sequence and its deduced 201-amino acid sequence (ZsMATα1 copy 2)

exhibited a identity of 86.4 and 87.5% with *Z. rouxii* counterparts, and of 35.8 and 32.0% with *S. cerevisiae* counterparts, respectively. The MATα1 like-coding sequence in Zs*MTL*α locus copy 3 and the deduced 201-amino acid sequence were the most divergent from those of *Z. rouxii* (identity of 15.9 and 81.5%, respectively) and *S. cerevisiae* (36.2 and 30.9%, respectively). The NJ-tree was constructed using a selection of MATα1 sequences from representative taxa of post and pre-WGD species. As expected, *ZsMAT*α1 copies 2 and 3 did not grouped to *Z. rouxii* counterpart, but instead clusterized separately (bootstrapping values of 93 and 100%, respectively), with copy 2 more closer to *Z. rouxii* MATα1 than copy 3 (**Figure 4.3A**).

Alignment of the ZsMATα1 copies with *Z. rouxii* and *S. cerevisiae* MATα1 proteins revealed a region of highest similarity inside the α1-box region and the acidic carboxyl terminal end (**Figure 4.3B**), whose integrity is required for DNA binding and vegetative incompatibility respectively (Philley et al., 1994). Searching in PFam-A database revealed that α1-box from ZsMATα1 copy 2 adhered a little better to the consensus profile (PF04769; *E*-values 6.6e-09) than the homologous regions in copy 1 and 3 (*E*-value 1.8e-08 and 1.7e-08, respectively). For example, α1-box in ZsMATα1 copy 2 had a E159 residue as in *S. cerevisiae* while in the same position this amino acid was replaced by alanine in ZsMTLα1p copy 1 and 3, as well as in *Z. rouxii* MATα1. Although the consensus profile does not consider this substitution as conservative, it is still detectable in MATα1 box of related species such as *Torulaspora delbrueckii*, *Vanderwaltozyma polyspora*, and *Candida glabrata*. In addition, the α1-box of ZsMATα1 copy 3 displayed a H163N substitution compared to ZsMATα1 copies 1 and 2 and *Z. rouxii* MATα1. However, this position is poor conserved in the consensus profile for α1-box even inside Saccharomycetes. The amino acid substitutions among ZrMATα1 copies occurred mainly at their amino terminal ends, with the most divergent copy 3 displaying 17 unique residues, while sharing 12 and 7 substitutions with copy 2 and copy 1, respectively. Noteworthy, in the last case they all involved lysine or arginine residues.

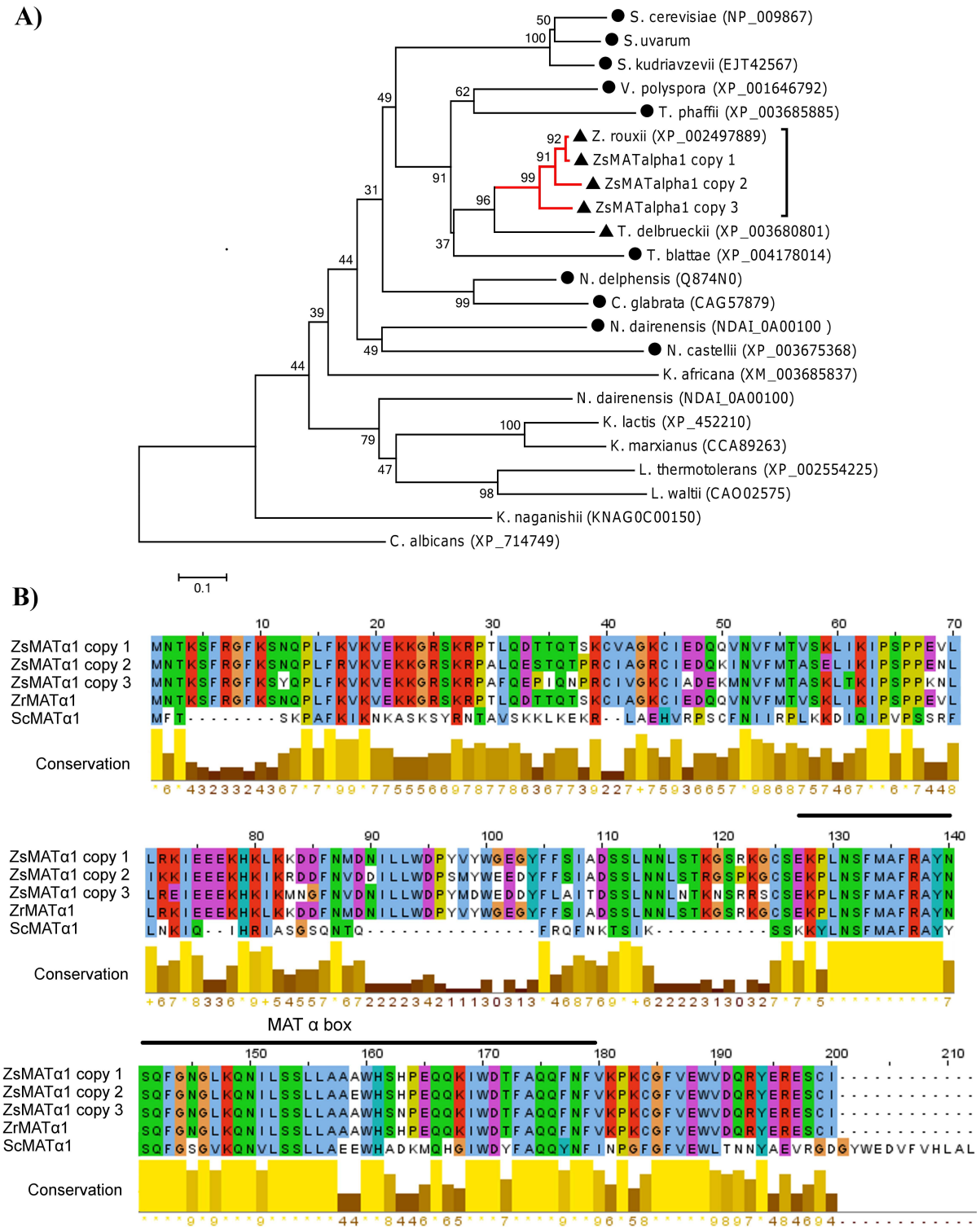


Figure 4.3. Phylogenetic analysis and sequences comparison of MATα1 proteins. A) A neighbor joining tree showing the phylogenetic relationships between *Z. sapae* and other hemiascomycetes inferred from MATα1 proteins. Number on branches indicated bootstrap support (1000 pseudoreplicates) from NJ. Red branch indicates ZsMATα1 sequences, dark dot indicates post-WGD species; dark triangle indicates pre-WGD species with *HO* gene. **B)** Amino acid alignment of putative MATα1 copies isolated from *Z. sapae* (ZsMATα1 copy 1, 2 and 3; GenBank under submission) and the orthologs from *Z. rouxii* (GenBank: XP_002497889) and *S. cerevisiae* (GenBank: NP_009867).

The conserved domain for mating-type proteins MATα1 is showed (MATα-box; Pfam PF04769). The amino acid identities were colored according the Clustal-X color scheme, and the conservation index at each alignment position were provided by Jalview (Waterhouse et al., 2009).

The MATα2 coding sequences from *ZsMTL*α loci copies 1, 2, and 3 showed 99.9, 77.6, and 66.7% of identities with *Z. rouxii MAT*α2 orthologs, respectively. Identity percentages of *ZsMAT*α2 copy 1, 2, and 3 with *S. cerevisiae MAT*α2 gene ranged from 36.3 to 42.4%. Three deduced 226-amino acid sequences, namely ZsMATα2 copies 1, 2, and 3, shared 99.5, 80.5, and 65.6% identity with *Z. rouxii* MATα2 and 38.6, 39.1, and 35.2% identity with *S. cerevisiae* MATα2, respectively.

Phylogeny inferred based on the MATα2 amino acid sequences of post and pre-WGD species showed a tree topology congruent with the species relationships established based on MATα1 sequences. ABT301^{T} genome harbors three MATα2 variants, one (copy 1) clustered with *Z. rouxii* MATα2, whereas the other (copies 2 and 3) related but phylogenetically different based on high level of amino acid sequence divergence (**Figure 4.4A**). All three copies contained a conserved HD1 class homeodomain (Pfam PF00046; *E*-values 4.1e-7, 2.0e-7 and 9.5e-8, for ZsMATα2 copy 1, copy 2 and copy 3, respectively), consisting in a three-helix globular domain which contacts both major groove bases and the DNA backbone (Wolbergret et al., 1991) (**Figure 4.4B**). Six residues in helix 3 that contact the backbone with their side chains in *S. cerevisiae* MATα2 homeodomains were also conserved in *Z. rouxii*, *Z. sapae* along with the tyrosine residue (Y145 in ZrMATα2) just upstream at N-terminal of helix 1 (**Figure 4.4**). A further phenylalanine residue with the same structural role in *S. cererevisiae* MATα2 was indeed replaced by lysine in *Z. rouxii* (Y150L). The three residues of *S. cerevisiae* MATα2 which form additional interactions with the DNA minor groove were conserved both in *Z. sapae* and *Z. rouxii* (R146, G147 and R149). However, portions of the protein outside of the homeodomain which mediate interactions with accessory proteins had a different degree of conservation. The unstructured carboxy-terminal tail of α2 is required for the formation of a stable al/α2-operator complex in *S. cerevisiae* and, thus, for the heterodimer-mediated repression of transcription. This domain is fully conserved in *Z. sapae* and

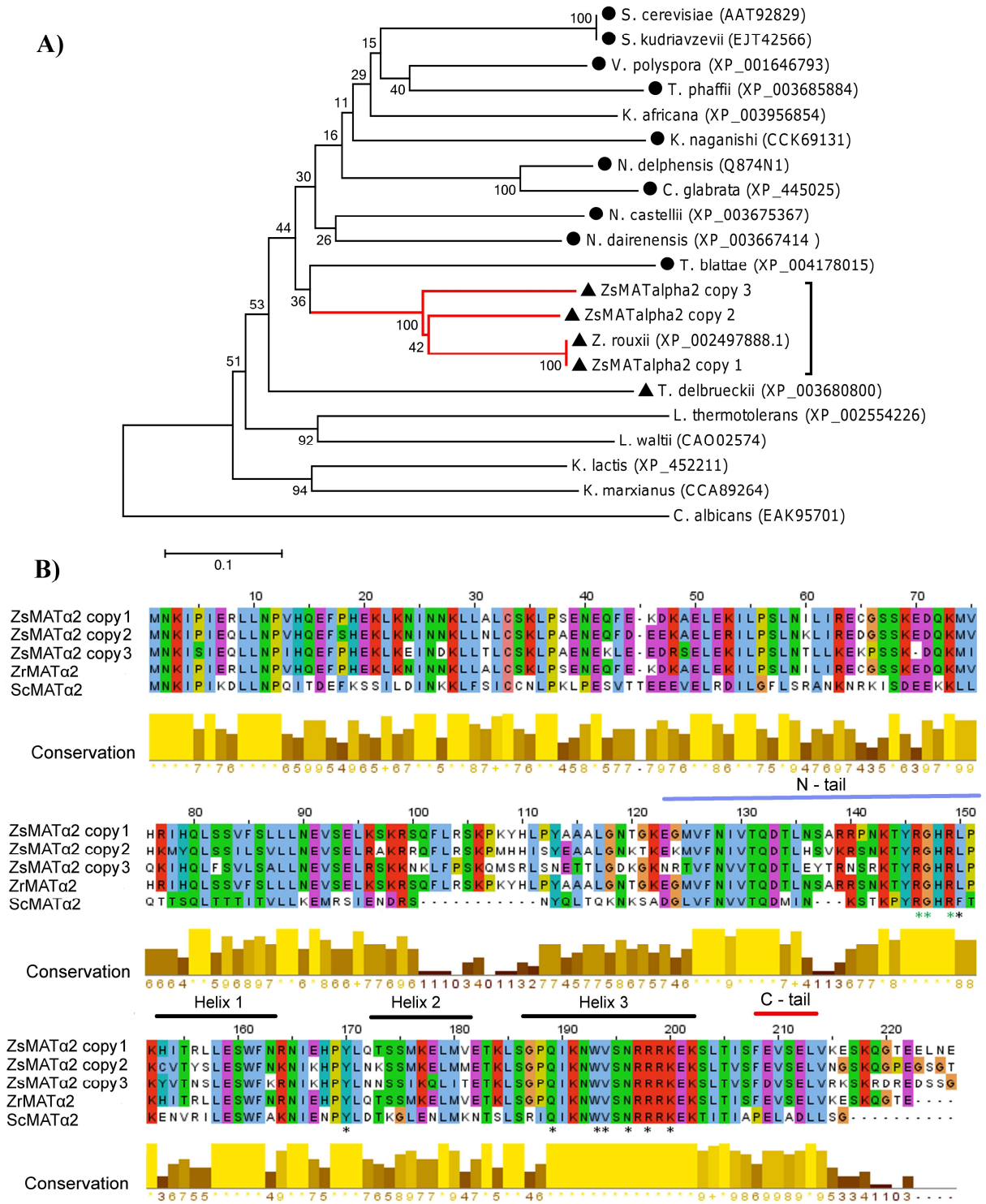


Figure 4.4. Phylogenetic analysis and sequences comparison of MATα2 proteins. A) Neighbour-joining phylogeny as inferred from MATα2 sequences depicting evolutionary relationships between *Z. sapae* and other hemiascomycetes. Number on branches indicate bootstrap support (1000 pseudoreplicates) from NJ. Red branch indicates ZsMATα2 sequences, dark dot indicates post-WGD species; dark triangle indicates pre-WGD species with *HO* gene. B) Alignment of deduced amino-acid sequences from putative MATα2 genes cloned in *Z. sapae* (ZsMATα1 copy 1, 2 and 3; GenBank under submission) and orthologos MATα2 annotated in *Z. rouxii* (GenBank: XP_0024978881) and *S. cerevisiae* genomes (GenBank: NP_009866). The DNA binding homeodomain of MATα2

(Pfam PF00046) consisting in three three-helix globular domain which contact major groove bases and the DNA backbone are indicated by horizontal bars. Evolutionary conserved key residues involved in DNA binding are highlighted with black asterisks. Green asterisks denote amino-acids which take additional interactions with the DNA minor groove in *S. cerevisiae* MATα2, present in the unstructured tail at the N-terminal side of homeodomain (violet bar). The unstructured carboxy-terminal tail of *S. cerevisiae* MATα2 required for formation of a stable a1/α2-operator complex is also showed (red bar)

Z. rouxii MATα2 and largely resembled that found in *S. cerevisiae* (Mark et al., 1993). The intervening flexible hinge which connect the amino-terminal domain and the homeodomain of *S. cerevisiae* MATα2 mediates the interaction of MATα2α2 homodimer with two subunits of MCM1 and hence its operator binding capacity (Vershom et al., 1993). This sequence is more divergent in ZsMATα2 copy 3 compared to ZsMATα2 copies 1 and 2, and between MATα2 proteins in *Z. rouxii* and *S. cerevisiae*. The ability of MATα2 to form both homodimers (α2/α2) and heterodimers (α2/a) mainly relies on the integrity of the N-terminal portion (Ho et al., 1994, 2002). N-terminal homology between MATα2 in *Zygosaccharomyces* species and *S. cerevisiae* is less than that found for the homeodomains, probably revealing a specie-specific co-evolution of the dimerization binding motifs. ZsMATα2 copy 3 was the most divergent from copies 1 and 2, owing to unique amino acid replacements even if, in a few positions, the residue was different in all three copies, suggesting that these amino acid substitutions were less affected by functional constrains.

4.3.4. Isolation of a conserved domain of *Z. sapae MAT*a2 and characterization of *MTL*a locus

A cloning strategy similar to that used for Zs*MTL*α loci was carried out for determining the *MAT***a**-like locus in *Z. sapae*. Homology comparison of MAT**a**1 amino acid sequences from *S. cerevisiae* and *Z. rouxii* allowed establishing two conserved motifs that presumably could be also conserved in *Z. sapae*. Two degenerated primers were designed that encompassing 153 bp of the *Z. rouxii MAT***a**1 gene. One PCR product of 153 bp was identified and cloned, and the resulting plasmid pA1.12 was sequenced. Sequence analysis of pA1.12 showed that PCR insert was 100% identical to *Z. rouxii MAT***a**1 gene (ZYRO0C18348g).

4.3.5. Characterization of *MTL***a** locus

The complete *MAT***a** idiomorph locus was obtained by a combination of iPCR and PCR walking using the 153-bp MAT**a**1 partial sequence as a starting point. The resulted 2088-bp consensus sequence (*Zs MTL***a** idiomorph) included two ORFs encoding putative MAT**a**1 and MAT**a**2-like proteins, respectively, separated by a 279-bp intergenic sequence. The ORF encoding a putative MAT**a**1 was 474 bp in length and displayed a putative 51-bp intron. The deduced ZsMAT**a**1 140-aa sequence was 100 % identical to *Z. rouxii* MAT**a**1 (**Figure 4.5A**). With respect to genomic location, PFGE-Southern blotting showed that *ZsMTL***a** locus resides on the single high molecular weight chromosome L poorly resolved from chromosome L' (**Figure 4.6**). The MAT**a**1 harbored a conserved HD2 class homeodomain (HD) (Pfam *E*-value, 8.1-e10, PF00046), consisting of an unstructured N-terminal arm and three helices linked by two loops (**Figure 4.5A**) (Anderson et al., 2002).

The 888-bp ORF encoding a putative MAT**a**2 was shorter in length that the *Z. rouxii* ortholog (ZYRO0C18326g) due to a 26-bp deletion. Thus, the deduced MAT**a**2p amino acid sequence is 9 amino acids shorter than *Z. rouxii* MAT**a**2 due to the lack of the domain $_{219}$(QAQAANAQA)$_{227}$ (**Figure 4.5B, C**). MAT**a**2 was provided with single MATA_HMG-box, class I member of the HMG-box superfamily of DNA-binding proteins (NCBI's Conserved Domain Database code: cd01389; residues 72-145; *E*-value 4.31e-06; **Figure 4.5B**), coding by a sequence spanning across Y**a** and X regions. Beyond this putative functional domain, there were a very few spotted similarities with MAT**a**2 annotated in close related species. The inferred joint point responsible for peptide removal from MAT**a**2 in *Z. sapae* laid on X region and went through an imperfect tandem sequence $(CAAGCA/G)_3$ at the nucleotide position 653 (**Figure 4.5C**).

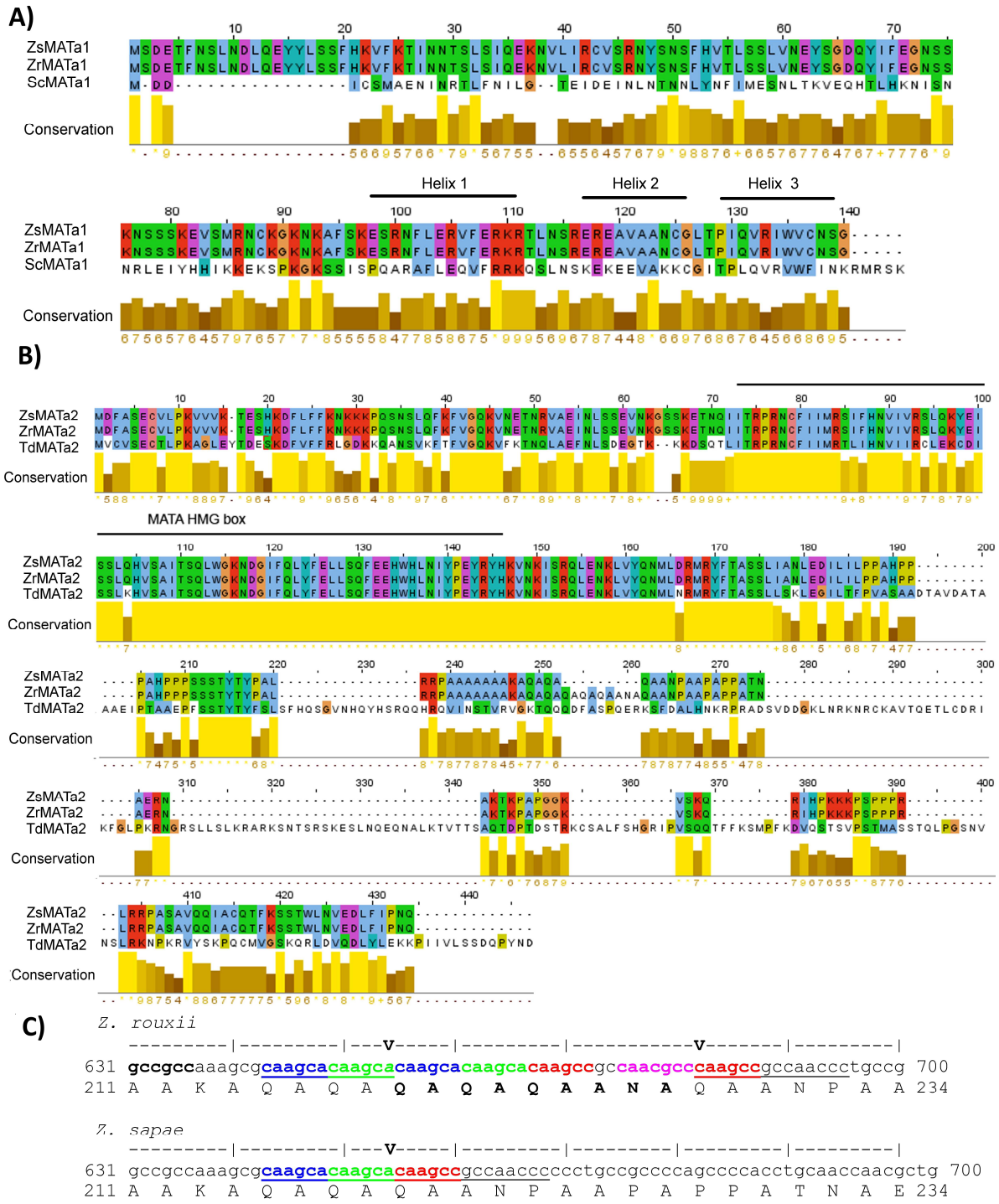


Figure 4.5. Amino acid sequence alignments of MATa1 and MATa2. A) Alignment of MAT**a**1 from *Z. sapae* (ZsMAT**a**1, GenBank under submission), *Z. rouxii* (ZrMAT**a**1, GenBank: XP_002496431) and *S. cerevisiae* (ScMAT**a**1, GenBank: NP_010021). The three alpha helices that characterized the homeodomain were highlighted. **B)** Alignment of MAT**a**2 from *Z. sapae* (ZsMAT**a**2, GenBank under submission), *Z. rouxii* (ZrMAT**a**2, GenBank: XP_002496430) and *Torulaspora delbrueckii* (TdMAT**a**2, GenBank: XP_003682598). The MATA_HMG-box which binds the minor groove of DNA is pointed out. In both alignments, the amino acid identities were colored according the

Clustal-X color scheme and the conservation index at each alignment position were provided by Jalview (Waterhouse *et al*. 2009). **C)** Partial nucleotide sequence alignment shows indel junction boundaries (V) in *Z. rouxii* and *Z. sapae* MAT**a**2. Imperfect tandem repeat units are highlighted in different colors.

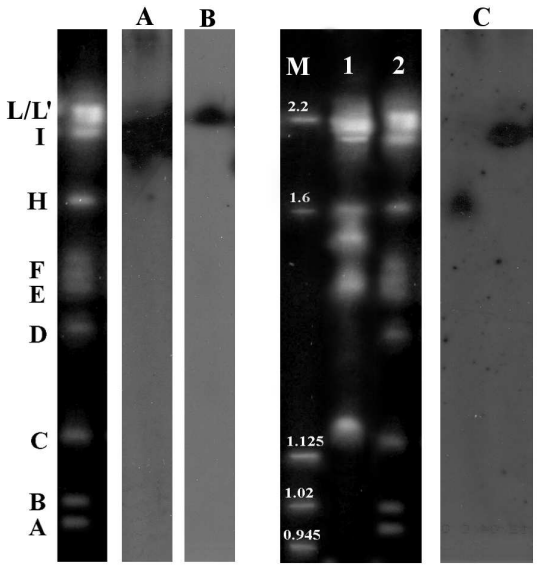


Figure 4.6. Mapping of *ZsMTLα*, *ZsMTL*a and *ZsHO* loci. Chromosome were separated by PFGE for *Zygosaccharomyces rouxii* CBS 732^T (1) and *Zygosaccharomyces sapae* ABT301^T (2) and Southern blotting analyses were carried out with probes labelling to α-idiomorph loci (A), **a**-idiomorph locus (B), and *HO* gene (C), respectively. M indicates the chromosomal size ladder (*Saccharomyces cerevisiae* S288C, Bio-Rad Laboratories) is in megabase pairs (Mbp). *Z. sapae* chromosomes are indicated in uppercase letters.

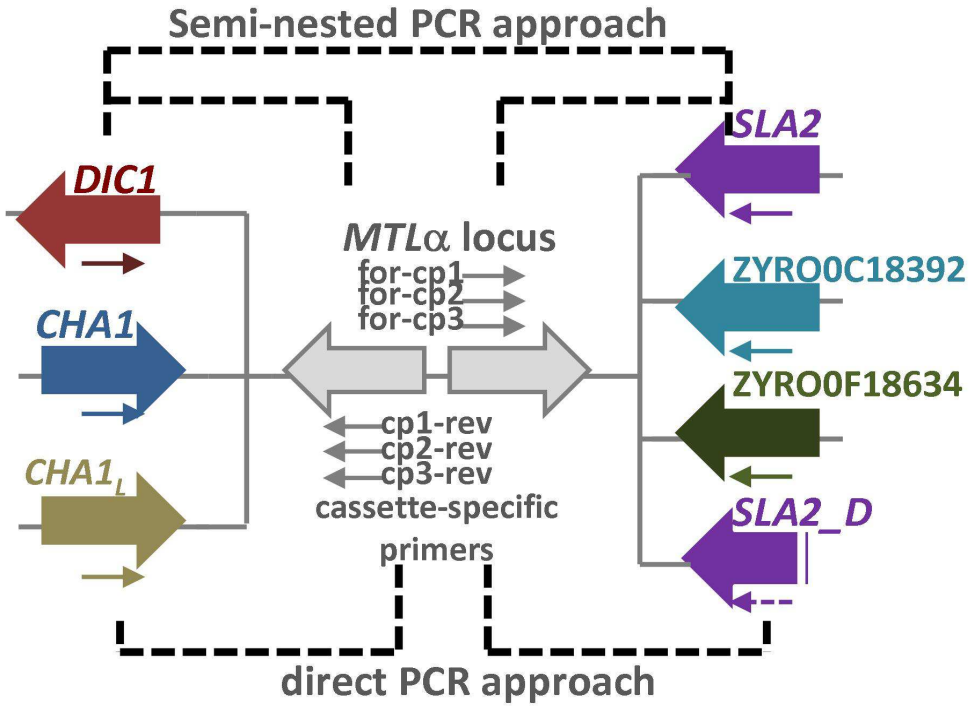
4.3.6. Cassette system analysis

In *S. cerevisiae*, the functional *MAT*α locus is flanked by *BUD5* at the 3′ end of *MAT***a**1 and by *TAF2* at the 3′ end of *MAT***a**2, while the silent *HMR* and *HML* loci are flanked by YCRWDDta12/YCR097W-a and YCL068C/HCl065W, respectively. In *Z. rouxii*, several chromosomal arrangements have been revealed in different strains or in different collection cultures of the same strain (Watanabe et al., 2013), suggesting that the *MAT* loci are ectopic recombination hotspots. Analysis of the 3' end flanking genes showed that *SLA2* gene is frequently linked both to *MAT* and *HML* cassettes in all the chromosomal rearrangements described in *Z. rouxii* (Watanabe et al., 2013) and in other hemiascomycetes (Gordon et al., 2011). To assign chromosomal positions and establish neighboring genes of Zs*MTL* idiomorphs, PCR amplifications across the whole cloned

cassettes were performed, employing primer sets designed on genes flanking all *MAT*, *HML*, and *HMR* cassettes observed in other *Z. rouxii* strains (**Figure 4.7A**). To capture possible divergent sequences of 3' ends flanking regions of cloned *ZsMTL* idiomorphs, we designed a further degenerate primer DownMATa1R, spanning the motif FEFYADC of *Z. rouxii SLA2* gene (ZYRO0F15862g) (*SLA2_D*). Positive PCR products were obtained with the primer pairs 2/A, 3/C, 2/DownMATa1R and further screened via semi-nested approach using primers specific for *ZsMTL***a**, *ZsMTL*α copies 1, 2, and 3, respectively (**Figure 4.7B**). No PCR products were gained with the primer 1 in *CHA1* (ZYRO0F15774g) gene sequence at *MAT*. To exclude alternative combinations of flanking genes other than those described by Watanabe et al. (2013), direct PCR was done combining *ZsMTL*-copy specific primers and primers laying on other potential flanking genes and the results of both approaches are reported in **Figure 4.7A and Figure 4.7B**. Four *ZsMTL*α cassettes were arranged in the following gene order: *CHA1* (ZYRO0F18524g-like)-*ZsMTL*α copy 1-*SLA2* (ZYRO0F18364g-like); *DIC1-ZsMTL*α copy 1-*SLA2* (ZYRO0F15862g-like); *DIC1-ZsMAT*α copy 2-*SLA2* (ZYRO0F15862g-like); *DIC1-ZsMAT*α copy 3-*SLA2* (ZYRO0F15862g-like). The arrangement *CHA1* (ZYRO0F18524g-like)-*ZsMTL*α copy 1-*SLA2* (ZYRO0F18364g-like) is consistent with the designation of this locus as silent *HML* cassette (*ZsHML* copy 1). The arrangement *DIC1-MAT-SLA2* indicates that all three α-idiomorph *ZsMTL* cassettes are orthologous to *MAT* expression loci in other pre-WGD species and are labeled as Zs*MAT*α copies from 1 to 3. The 2/DownMATa1R PCR amplicons were positive to all three *ZsMTL*α-copy specific internal primers, resulting in three α-idiomorphs cassettes with a *Z. rouxii DIC1*-like upstream region and a downstream region divergent from that found for *Z. rouxii MAT* or *HML*-flanking genes, where *SLA2* (ZYRO0F15862g) and ZYRO0F18524g locus have been detected, respectively. Based on the syntenic pattern *DIC1-ZsMTL*α*-SLA2_D*, the three *ZsMTL*α cassettes were referred as Zs*HML*_D copies 1, 2 and 3. Finally, the downstream region of *ZsMTL***a** cassette resulted to be orthologous to the *Z. rouxii SLA2* gene (ZYRO0F15862g), while the gene at

its 5' end was still unknown, since all the PCR amplifications failed, suggesting the *ZsMTL***a** upstream region is not conserved between *Z. rouxii* and *Z. sapae*.

A



B

Synteny (5'->3')	Primer pair (forward/reverse)	*ZsMTL* loci			
		α copy 1	α copy 2	α copy 3	a
CHA1-MAT-SLA2	1/A	-/5'-/3'+	-/5'-/3'+	-/5'-/3'+	-/5'-/3'+
CHA1-MAT-ZYRO0F18392	1/B	-/5'-/3'-	-/5'-/3'-	-/5'-/3'-	-/5'-/3'-
CHA1-MAT-ZYRO0F18634	1/C	-/5'-/3'+	-/5'-/3'-	-/5'-/3'-	-/5'-/3'-
CHA1-MAT-SLA2_D	1/DownMATa1R1	-/5'-/3'+	-/5'-/3'+	-/5'-/3'+	-/5'-/3'+
DIC1-MAT-SLA2	2/A	+/5'+/3'+	+/5'+/3'+	+/5'+/3'+	-/5'-/3'+
DIC1-MAT-ZYRO0C18392	2/B	-/5'+/3'-	-/5'+/3'-	-/5'+/3'-	-/5'-/3'-
DIC1-MAT-ZYRO0F18634	2/C	-/5'+/3'+	-/5'+/3'-	-/5'+/3'-	-/5'-/3'-
DIC1-MAT-SLA2_D	2/DownMATa1R1	+/5'+/3'+	+/5'+/3'+	+/5'+/3'+	-/5'-/3'-
CHA1$_L$*-MAT-SLA2*	3-A	-/5'-/3'+	-/5'-/3'+	-/5'-/3'+	-/5'-/3'+
CHA1$_L$*-MAT*-ZYRO0F18392	3-B	-/5'-/3'-	-/5'-/3'-	-/5'-/3'-	-/5'-/3'-
CHA1$_L$*-MAT*- ZYRO0F18634	3-C	+/5'+/3'+	-/5'-/3'-	-/5'-/3'-	-/5'-/3'-
CHA1$_L$*-MAT-SLA2_D*	3/DownMATa1R1	-/5'+/3'+	-/5'-/3'+	-/5'-/3'+	-/5'-/3'-

Figure 4.7. Variability of *MTL*-flanking regions in *Z. sapae* ABT301^T. A) PCR-based strategies used for determining chromosomal three cassette system in *Z. sapae*. Cassette specific-primers were used to screen PCR

products obtained by using all possible combinations of primers spanning putative *MAT* flanking genes (semi-nested PCR approach); in case of negative results, 5' and 3' PCR walking was done by using all possible combinations of cassette specific-primers and *MAT* flanking genes primers (direct PCR approach). Small arrows indicate primers; small dotted arrows indicate degenerate primers. B) PCR amplification results of the Zs*MTL* loci from ABT301^{T}. Left symbols indicate presence or absence of PCR products after semi-nested PCR; middle symbols indicate presence or absence of direct PCR products at 5' end; right symbols indicate presence or absence of direct PCR products at 3' end. Abbreviation: *SLA_D*, divergent *SLA2* gene partial sequence determined with the degenerate primer 1-DownMATa1R; *CHA1*$_{L}$, *CHA1* (ZYRO0F18524) located near to the silent *HML* cassette in CBS 732^{T} genome.

4.3.7. Analysis of Z and X regions

MAT, *HMR* and *HML* share two homologous regions flanking the Y sequences, termed X and Z. The Z and X regions are virtually identical among the three copies within each genome and were found to be among the most slowly evolving sequences in the genome of *Saccharomyces* species (Kellis et al., 2003). Since HO creates a DSB within *MAT* locus at the junction between Y and Z sequences (Haber, 1998) and single base substitutions at the region near the Y/Z border are sufficient to inhibit *HO*-cut *MAT* switching (Weiffenbach et al., 1983; Nickoloff et al., 1990), we determined the extent of the sequence homology in the 3' flanking Z sequences among the Zs*MAT* cassettes to take information about their functional state. As expected in species with HO endonucelase, Y–Z junction has been stabilized to a site in ZsMAT**a**1 and all ZsMATα1 copies. The eight *Z. sapae* mating type cassettes (3 *ZsMATα*-D copies, 3 *ZsMATα*-like copies, the *ZsHML*-like copy 1, and the *ZsMAT***a** locus, respectively) were always found with the HO site-specific sequence CGCAGC at the first site of the Z regions. This sequence was also found in *C. glabrata* (Butler et al., 2004), and represents a variant of the canonical *S. cerevisiae* recognition sequence (5'- CGCAAC -3') for the HO site-specific enzymatic cleavage of *MAT* during switching (**Figure 4.8**). Both HO site-specific sequences have been shown to be cleaved efficiently by the *S. cerevisiae* HO *in vivo* (Nickoloff et al., 1990). The high level of conservation at the Y/Z borders suggest that all the mating type loci cassettes could functional either as putative *MAT* or *HML*/*HMR* donor sequences. Otherwise, base substitutions were observed at the 3' end of Z region.

```
                                HO-specific site
CBS 732                 CGCAGCAGTTTAATTTTGTCAAGCCGAAGTGTGGGTTTGTGGAGTGGGTG 50
ZsMATa                  CGCAGCAGTTTAATTTTGTCAAGCCGAAGTGTGGGTTTGTGGAGTGGGTG 50
ZsMATalpha copy 1       CGCAGCAGTTTAATTTTGTCAAGCCGAAGTGTGGGTTTGTGGAGTGGGTG 50
ZsHML copy 1            CGCAGCAGTTTAATTTTGTCAAGCCGAAGTGTGGGTTTGTGGAGTGGGTG 50
ZsMATalpha copy 2       CGCAGCAGTTTAATTTTGTCAAGCCGAAGTGTGGGTTTGTGGAGTGGGTG 50
ZsMATalpha copy 3       CGCAGCAGTTTAATTTTGTCAAGCCGAAGTGTGGGTTTGTGGAGTGGGTG 50
ZsHML_D copy 1          CGCAGCAGTTTAATTTTGTCAAGCCGAAGTGTGGGTTTGTGGAGTGGGTG 50
ZsHML_D copy 2          CGCAGCAGTTTAATTTTGTCAAGCCGAAGTGTGGGTTTGTGGAGTGGGTG 50
ZsHML_D copy 3          CGCAGCAGTTTAATTTTGTCAAGCCGAAGTGTGGGTTTGTGGAGTGGGTG 50
                        **************************************************

CBS 732                 GATCAAAGATATGAGCGGGAGAGTTGTATTTAGTTTGTTAAGAGTTGTTG 100
ZsMATa                  GATCAAAGATATGAGCGGGAGAGTTGTATTTAGTTTGTTAAGAGTTGTTG 100
ZsMATalpha copy 1       GATCAAAGATATGAGCGGGAGAGTTGTATTTAGTTTGTTAAGAGTTGTTG 100
ZsHML copy 1            GATCAAAGATATGAGCGGGAGAGTTGTATTTAGTTTGTTAAGAGTTGTTG 100
ZsMATalpha copy 2       GATCAAAGATATGAGCGGGAGAGTTGTATTTAGTTTGTTAAGAGTTGTTG 100
ZsMATalpha copy 3       GATCAAAGATATGAGCGGGAGAGTTGTATTTAGTTTGTTAAGAGTTGTTG 100
ZsHML_D copy 1          GATCAAAGATATGAGCGGGAGAGTTGTATTTAGTTTGTTAAGAGTTGTTG 100
ZsHML_D copy 2          GATCAAAGATATGAGCGGGAGAGTTGTATTTAGTTTGTTAAGAGTTGTTG 100
ZsHML_D copy 3          GATCAAAGATATGAGCGGGAGAGTTGTATTTAGTTTGTTAAGAGTTGTTG 100
                        **************************************************

CBS 732                 TAGATTTGTATTTGCATTGATGATGTCTTGTTGGGAGGGGGAAAGTAGTG 150
ZsMATa                  TAGATTTGTATTTGCATTGATGATGTCTTGTTGGGAGGGGGAAAGTAGTG 150
ZsMATalpha copy 1       TAGATTTGTATTTGCATTGATGATGTCTTGTTGGGAGGGGGAAAGTAGTG 150
ZsHML copy 1            TAGATTTGTATTTGCATTGATGATGTCTTGTTGGGAGGGGGAAAGTAGTG 150
ZsMATalpha copy 2       TAGATTTGTATTTGCATTGATGATGTCTTGTTGGGAGGGGGAAAGTAGTG 150
ZsMATalpha copy 3       TAGATTTGTATTTGCATTGATGATGTCTTGTTGGGAGGGGGAAAGTAGTG 150
ZsHML_D copy 1          TAGATTTGTATTTGCATTGATGATGTCTTGTTGGGAGGGGGAAAGTAGTG 150
ZsHML_D copy 2          TAGATTTGTATTTGCATTGATGATGTCTTGTTGGGAGGGGGAAAGTAGTG 150
ZsHML_D copy 3          TAGATTTGTATTTGCATTGATGATGTCTTGTTGGGAGGGGGAAAGTAGTG 150
                        **************************************************

CBS 732                 CTTGGAAGTAGAGTTGAGAAGCTTGGGCCAGCAGTGGCGATGGATTGTGT 200
ZsMATa                  CTTGGAAGTAGAGTTGAGAAGCTTGGGCCAGCAGTGGCGATGGATTGTGT 200
ZsMATalpha copy 1       CTTGGAAGTAGAGTTGAGAAGCTTGGGCCAGCAGTGGCGATGGATTGTGT 200
ZsHML copy 1            CTTGGAAGTAGAGTTGAGAAGCTTGGGCCAGCAGTGGCGATGGATTGTGT 200
ZsMATalpha copy 2       CTTGGAAGTAGAGTTGAGAAGCTTGGGCCAGCAGTGGCGATGGATTGTGT 200
ZsMATalpha copy 3       CTTGGAAGTAGAGTTGAGAAGCTTGGGCCAGCAGTGGCGATGGATTGTGT 200
ZsHML_D copy 1          CTTGGAAGTAGAGTTGAGAAGCTTGTGCCAGCACTGGCGATGGATTGTGT 200
ZsHML_D copy 2          CTTGGAAGTAGAGTTGAGAAGCTTGTGCCAGCACTGGCGATGGATTGTGT 200
ZsHML_D copy 3          CTTGGAAGTAGAGTTGAGAAGCTTGTGCCAGCACTGGCGATGGATTGTGT 200
                        ************************* ******* ****************

CBS 732                 TGTAGTAGTGTAGTTATCTGATGCGTTAATTGTATTATAGT 241
ZsMATa                  TGTAGTAGTGTAGTTATCTGATGCGTTAATTGTATTATAGT 241
ZsMATalpha copy 1       TGTAGTAGTGTAGTTATCTGATGCGTTAATTGTATTATAGT 241
ZsHML copy 1            TGTAGTAGTGTAGTTATCTGATGCGTTAATTGTATTATAGT 241
ZsMATalpha copy 2       TGTAGTAGTGTAGTTATCTGATGCGTTAATTGTATTATAGT 241
ZsMATalpha copy 3       TGTAGTAGTGTAGTTATCTGATGCGTTAATTGTATTATAGT 241
ZsHML_D copy 1          TGTAGTAGTGTACTGATCTGATGGGTTAATTGTATTAGGCT 241
ZsHML_D copy 2          TGTAGTAGTGTACTGATCTGATGGGTTAATTGTATTAGGCT 241
ZsHML_D copy 3          TGTAGTAGTGTACTGATCTGATGGGTTAATTGTATTAGGCT 241
                        ************ * ******** ************* . *
```

Figure 4.8. Z regions sequence comparisons from *Z. sapae* strain ABT301^{T} and *Z. rouxii* CBS 732^{T}. Aligned Z sequences of eight *Z. sapae* (Zs) mating type cassettes: *ZsMAT*α copies 1, 2, 3, *ZsHML*_D copies 1, 2, and 3, *ZsHML* copy 1, and *ZsMAT***a** cassette

```
CBS 732                 AGCGATTTGCTGGACGGCGGAGGCGGGGCGGCGGAGGCGGGGCGGGGGCG 50
ZsHML copy 1            AGCGATTTGCTGGACGGCGGAGGCGGGGCGGCGGAGGCGGGGCGGGGGCG 50
ZsMATalpha copy 1       AGCGATTTGCTGGACGGCGGAGGCGGGGCGGCGTAGGCGGGGCGGGGGCG 50
ZsMAT**a**              AGCGATTTGCTGGACGGCGGAGGCGGGGCGGCGTAGGCGGGGCGGGGGCG 50
ZsMATalpha copy 2       AGCGATTTGCTGGACGGCGGAGGCGGGGCGGCGTAGGCGGGGCGGGGGCG 50
ZsHML_D copy 1          AGCGATTTGCTGGACGGCGGAGGCGGGGCGGCGTAGGCGGGGCGGGGGCG 50
ZsHML_D copy 2          AGCGATTTGCTGGACGGCGGAGGCGGGGCGGCGTAGGCGGGGCGGGGGCG 50
ZsMATalpha copy 3       AGCGATTTGCTGGACGGCGGAGGCGGGGCGGCGTAGGCGGGGCGGGGGCG 50
ZsHML_D copy 3          AGCGATTTGCTGGACGGCGGAGGCGGGGCGGCGTAGGCGGGGCGGGGGCG 50
                        ********************************* ***************

CBS 732                 ATGGTTTTTTCTTGGGGTGGATTCGCTGCTTGGAGACTTTGCCGCCGGGG 100
ZsHML copy 1            ATGGTTTTTTCTTGGGGTGGATTCGCTGCTTGGAGACTTTGCCGCCGGGG 100
ZsMATalpha copy 1       ATGGTTTTTTCTTGGGGTGGATTCGCTGCTTGGAGACTTTGCCGCCGGGG 100
ZsMAT**a**              ATGGTTTTTTCTTGGGGTGGATTCGCTGCTTGGAGACTTTGCCGCCGGGG 100
ZsMATalpha copy 2       ATGGTTTTTTCTTGGGGTGGATTCGCTGCTTGGAGACTTTGCCGCCGGGG 100
ZsHML_D copy 1          ATGGTTTTTTCTTGGGGTGGATTCGCTGCTTGGAGACTTTGCCGCCGGGG 100
ZsHML_D copy 2          ATGGTTTTTTCTTGGGGTGGATTCGCTGCTTGGAGACTTTGCCGCCGGGG 100
ZsMATalpha copy 3       ATGGTTTTTTCTTGGGGTGGATTCGCTGCTTGGAGACTTTGCCGCCGGGG 100
ZsHML_D copy 3          ATGGTTTTTTCTTGGGGTGGATTCGCTGCTTGGAGACTTTGCCGCCGGGG 100
                        ************************************************

CBS 732                 GCGGGTTTTGTTTTTGCATTTCTCTCAGCGTTGGTTGCAGGTGGGGCTGG 150
ZsHML copy 1            GCGGGTTTTGTTTTTGCATTTCTCTCAGCGTTGGTTGCAGGTGGGGCTGG 150
ZsMATalpha copy 1       GCGGGTTTTGTTTTTGCATTTCTCTCAGCGTTGGTTGCAGGTGGGGCTGG 150
ZsMAT**a**              GCGGGTTTTGTTTTTGCATTTCTCTCAGCGTTGGTTGCAGGTGGGGCTGG 150
ZsMATalpha copy 2       GCGGGTTTTGTTTTTGCATTTCTCTCAGCGTTGGTTGCAGGTGGGGCTGG 150
ZsHML_D copy 1          GCGGGTTTTGTTTTTGCATTTCTCTCAGCGTTGGTTGCAGGTGGGGCTGG 150
ZsHML_D copy 2          GCGGGTTTTGTTTTTGCATTTCTCTCAGCGTTGGTTGCAGGTGGGGCTGG 150
ZsMATalpha copy 3       GCGGGTTTTGTTTTTGCATTTCTCTCAGCGTTGGTTGCAGGTGGGGCTGG 150
ZsHML_D copy 3          GCGGGTTTTGTTTTTGCATTTCTCTCAGCGTTGGTTGCAGGTGGGGCTGG 150
                        ************************************************

CBS 732                 GGCGGCAGGGTTGGCGGCTTGGGCGTTGGCGGCTTGTGCTTGTGCTTGTG 200
ZsHML copy 1            GGCGGCAGGGTTGGCGGCTTGGGCGTTGGCGGCTTGTGCTTGTGCTTGTG 200
ZsMATalpha copy 1       GGCGGCAGGGTTGGCGGCTTG--------------------------TG 173
ZsMAT**a**              GGCGGCAGGGTTGGCGGCTTG--------------------------TG 173
ZsMATalpha copy 2       GGCGGCAGGGTTGGCGGCTTG--------------------------TG 173
ZsHML_D copy 1          GGCGGCAGGGTTGGCGGCTTG--------------------------TG 173
ZsHML_D copy 2          GGCGGCAGGGTTGGCGGCTTG--------------------------TG 173
ZsMATalpha copy 3       GGCGGCAGGGTTGGCGGCTTG--------------------------TG 173
ZsHML_D copy 3          GGCGGCAGGGTTGGCGGCTTG--------------------------TG 173
                        ********************                          **

CBS 732                 CTTGTGCTTGCGCTTTGGCGGCGGCGGCGGCGGCGGCAGGACGCCGCAGG 250
ZsHML copy 1            CTTGTGCTTGCGCTTTGGCGGCGGCGGCGGCGGCGGCAGGACGCCGCAGG 250
ZsMATalpha copy 1       CTTGTGCTTGCGCTTTGGCGGCGGCGGCGGCGGCGGCAGGACGCCGCAGG 223
ZsMAT**a**              CTTGTGCTTGCGCTTTGGCGGCGGCGGCGGCGGCGGCAGGACGCCGCAGG 223
ZsMATalpha copy 2       CTTGTGCTTGCGCTTTGGCGGCGGCGGCGGCGGCGGCAGGACGCCGCAGG 223
ZsHML_D copy 1          CTTGTGCTTGCGCTTTGGCGGCGGCGGCGGCGGCGGCAGGACGCCGCAGG 223
ZsHML_D copy 2          CTTGTGCTTGCGCTTTGGCGGCGGCGGCGGCGGCGGCAGGACGCCGCAGG 223
ZsMATalpha copy 3       CTTGTGCTTGCGCTTTGGCGGCGGCGGCGGCGGCGGCAGGACGCCGCAGG 223
ZsHML_D copy 3          CTTGTGCTTGCGCTTTGGCGGCGGCGGCGGCGGCGGCAGGACGCCGCAGG 223
                        ************************************************
```

Figure 4.9. X regions sequence comparisons from *Z. sapae* strain ABT301^{T} and *Z. rouxii* CBS 732^{T}. Partial aligned X sequences of eight *Z. sapae* (Zs) mating type cassettes: *ZsMAT*α copies 1, 2, 3, *ZsHML*_D copies 1, 2, and 3, *ZsHML*, and ZsMAT**a** cassette.

In particular, the four mating type cassettes flanked by *Z. rouxii*-like 5' and 3' regions, namely the loci *ZsHML*-like copy 1, *ZsMATα*-like copies from 1 to 3, and *ZsMAT***a**, displayed the Z regions 100% identical to those found in haploid *Z. rouxii* CBS 732^T, whereas the Zs*HML*_D copies 1, 2, and 3 differed for 8 SNPs from the canonical *Z. rouxii* Z sequences. Finally, X region analysis showed that *ZsMAT***a**2 extends into the X region, whereas the X/Yα junction is upstream the codon stop of *ZsMAT*α2 genes. Consistently to this organization, alignment of X regions showed that seven *Z. sapae* α and a-idiomorphs loci differed from those of *Z. rouxii* and *ZsHML* copy 1 for the aforementioned 26-bp deletion found in *ZsMAT***a**2 gene (**Figure 4.9**).

4.3.8. Cloning of *HO* gene

The occurrence of *HO*-cleavable site in Z region of *ZsMAT* cassettes suggests that ABT301^T genome could harbor *HO* endonuclease gene. To test this hypothesis, a degenerate primer pair was exploited to determine *Z. sapae* homologs of *Z. rouxii HO* gene (ZYRO0C10428g) (**Figure 4.2**). Two putative full-length ORFs, namely *ZsHO* copy 1 and copy 2, were identified with 100 and 86.2% identities to *Z. rouxii HO* gene, respectively. The predicted *Z. sapae* HO proteins have 100 and 92.6% sequence identities to *Z. rouxii* HO protein. NJ-based phylogeny inferred from amino acid HO sequences showed that ZsHO copy 2 is clearly distinct from *ZsHO* copy 1 and *Z. rouxii HO* (**Figure 4.10A**). To determine where HO copies are located on ABT301^T genome we performed a PFGE-Southern blotting with probe labeling to *HO* gene. The results showed that the chromosomal localization of *ZsHO*s differed from that of *Z. rouxii* ortholog. Consistently to *Z. rouxii* genome project (Souciet et al., 2009), CBS 732^T genome harbors *HO* gene on the low molecular weight chromosome C, whereas, in ABT301^T genome, *ZsHO* genes appears to be localized on the same high molecular weight chromosome I, which harbors the *ZsMTL*α loci (**Figure 4.6**).

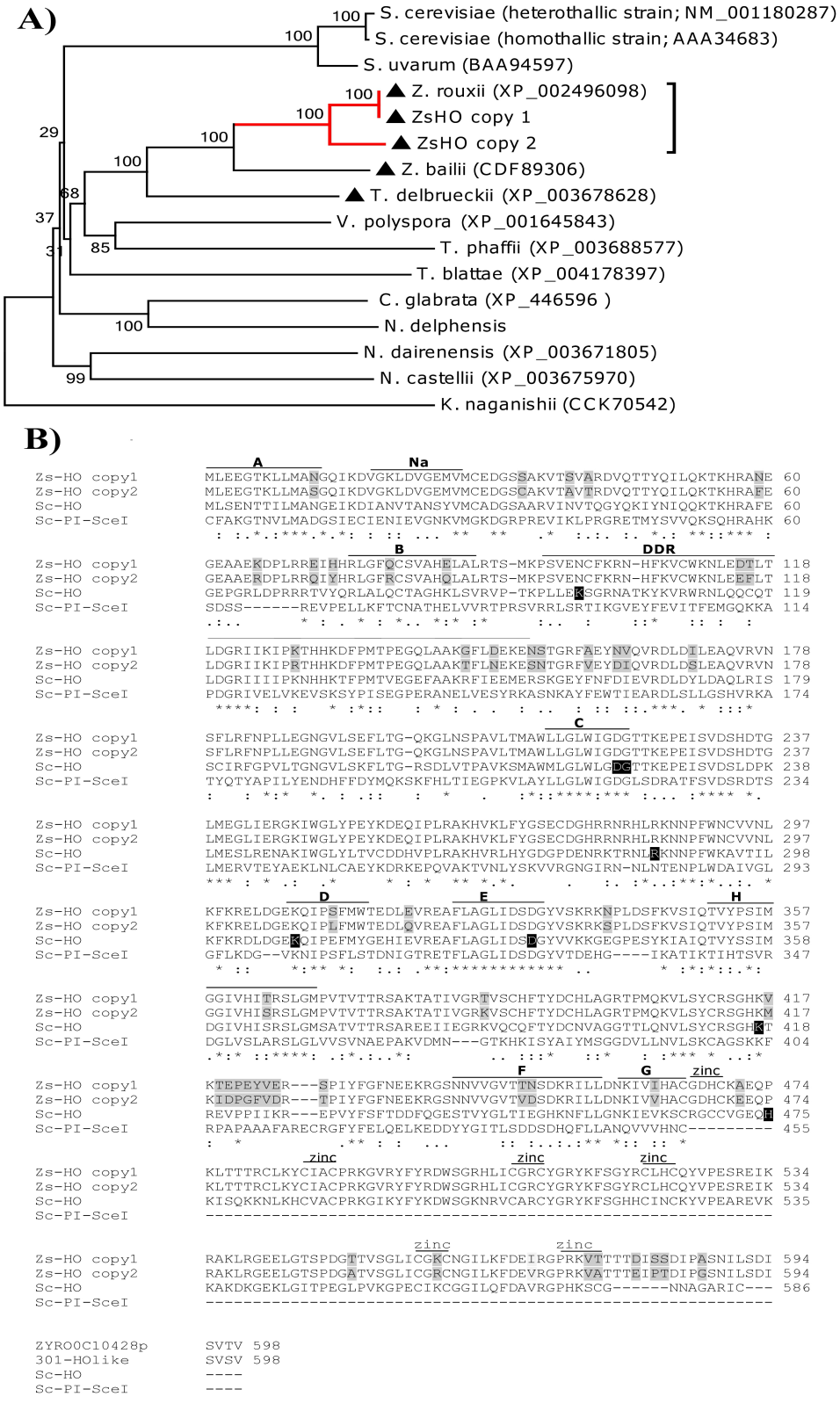


Figure 4.10. Phylogenetic analysis and amino acid sequences comparison of HO endonucleases. A) Neighbour-joining phylogeny as inferred from HO sequences depicting evolutionary relationships between *Z. sapae* and other

hemiascomycetes. Numbers on branches indicate bootstrap support (1000 pseudoreplicates) from NJ. Red branch indicates clusters including ZrHO and ZsHO sequences, dark triangle indicates pre-WGD species. B) Functional domains in PI-SceI and HO endonucleases. Primary amino acid alignment of *S. cerevisiae* PI-SceI (Sc-PI-SceI; GenBank: AA98762) and HO cloned in *S. cerevisiae* (Sc-HO; GenBank: CAA98806) and *Z. sapae* (ZsHO copy 1, GenBank: under submission; ZsHO copy 2, GenBank: under submission). Protein splicing domain with Hint motifs: A, N2, G, F and G. Endonuclease domains C, D, E, H. The DNA recognition region (DDR) and C-X2-C motifs of three putative zinc finger motifs (zinc) at PI-SceI carboxyl-terminal are also showed. In dark black shade are amino acid position functionally characterized in *S. cerevisiae* HO mutants. In light grey divergent positions between HO copies 1 and 2 in *Z. sapae*. Amino acid identities are reported below the alignment following Clustal-W rules: *, identity; :, conservative substitution; . semi-conservative. substitution.

In *S. cerevisiae*, HO endonuclease is necessary for the completion of the sexual cycle in clonal populations by inducing the formation of cells of opposite mating types within a clone. Since *Z. sapae* is unable to mate heterothallic sexual partner, its sexual reproduction may depend on the ability of some cells to switch mating types and fuse with related cells. As shown in Figure 5.10B, the highest homology between both *HO* genes cloned in *Z. sapae* and single copy *HO* found in *Z. rouxii* and *S. cerevisiae* genomes corresponded to conserved motifs characteristic of intein-encoded LAGLIDADG endonucleases (Hafez and Hausner, 2012; Stoddard, 2005; Belfort and Roberts, 1997). The two *Z. sapae* HO copies mostly differed in positions outside these functional domains (**Figure 4.10B**). With a few exceptions, *Z. sapae* HOs shared high identity in eight intein motifs lying at their C- and N- terminals which form the relic of the protein-splicing domain in HO proteins. The intervening sequence organized around the LAGLIDADG motif and organized in four amino acid domains responsible for HO endonuclease activity are conserved as well. The C-terminal end of *S. cerevisiae* HO harbors three zinc finger domains thought enhancing the specificity of HO binding (Bakharat et al., 2004). In ZsHOs these finger domains had the same organization in the primary sequence, even the last HX2C residues was absent. However, this motif can also be deleted from S. cerevisiae HO, without affecting the switching type activity (Bakharat et al., 2004). Structural and mutagenesis studies of LAGLIDADG endonuclease such as HO and PI-SceI in *S. cerevisiae*, revealed that the region downstream the B motif, the DDR region, although not well conserved in its primary sequence, probably contacts the phosphate DNA backbones of

target site through charged lateral chains of key amino acid residues (Moure et al., 2002; He et al., 1998). This hypothesis was supported by the effect of K99A substitution in *S. cerevisiae* HO that abrogated the switching type activity (Bakharat et al., 2004). Indeed high identity was also found in primary sequence of putative DDR region in HOs cloned in *Z. sapae*, while there was poor similarity with *S. cerevisiae* HO. Noteworthy the K99 residue in *S. cerevisiae* HO, was conservatively replaced by another positive charged amino acid (N97) in *Z. sapae* HOs. Similarly, a few numbers of amino acid residues that once replaced in *S. cerevisiae* HO hampered the binding and/or endonuclease activities *in vivo* or *in vitro*, were still conserved in both *Z. sapae* HOs (*i.e.* D222, G223, R286, K308, D333, K417) (Meiron et al., 1995; Ekino et al., 1999; Bakharat et al., 2004; Ezov et al., 2010). On the other hand, few exceptions to this conservation were found. For example, residue H475 in *S. cerevisiae* HO, involved in DNA binding of endonuclease target sequence, was substituted by proline in both *Z. sapae* HOs (Meiron et al., 1995; Ekino et al., 1999).

4.4. Discussion

Non conventional yeasts isolated from highly stress environments have been recently attracted increasing attention both for biotechnological exploitation and genome evolution studies. Chronic osmotic stress triggers aneuploidy (Pfau and Amon, 2012), increases the genome DNA content (Gerstein et al., 2006; Dhar et al., 2011), and favors chromosome instability (Aguilera and García-Muse, 2013). The frequency of sex and the nature of breeding systems have profound effects on genome variation and adaptation to stress environments (Lee et al., 2010; Balloux et al., 2003). Although *Z. rouxii* and relatives are the most relevant osmo and halotolerant food yeasts, research into its mating system is restricted to the haploid *Z. rouxii* strains (Butler at al., 2004; Gordon et al., 2011; Watanabe et al., 2013). Previous analysis demonstrated that *Z. sapae* diploid strains are genetically and phylogenetically different from *Z. rouxii*. Here, we examined mating type system in *Z. sapae* strain ABT301^{T} and found that the pattern of *ZsMTL* loci is completely different from those described for haploid *Z. rouxii* strains. Based on genome project (Souciet et al., 2009), haploid

strain CBS 732^T displayed the *MAT*α and *HML*α cassettes on chromosome F and the *HMR*a cassette located on chromosome C. The *MAT*α and *HML*α loci contain copies of *MAT*α1 and *MAT*α2 genes identical to their homologs from *HML* locus. While this work was in progress, Watanabe et al. (2013) used a PCR-based method for tagging 5' and 3' *MAT* flanking conserved regions. This study revealed the *MAT* loci arrangements in haploid *Z. rouxii* cells and demonstrated a variable mating type loci organization in haploid strains and in different cultures of CBS 732^T. Here, we exploited three experimental approaches, *i.e. MAT* gene cloning, PCR *MAT* cassette placement and PFGE-Southern blotting, to enroll the *MAT* loci co-occurring in ABT301^T genome and to inspect their genome configuration. Firstly, we provide evidences that *Z. sapae* ABT301^T possesses four independent mating type-like loci, resulting in an unusual a,α,α,α genotype. In addition to one *ZsMTL***a** locus harboring *MAT***a**2 and *MAT***a**1 genes, we found three *MTL*α loci, each containing pairs of *MAT*α1 and *MAT*α2 genes. Remarkably, in two out of three *ZsMTL*α loci *MAT*α1 and *MAT*α2 genes were slightly divergent from those harbored in the canonical *Z. rouxi MAT*α locus (ZYRO0F15840g and ZYRO0F15818g, respectively). A similar pattern of mating type genes expansion has been recently found in *Hortaea werneckii*, a highly halotolerant heterothallic black yeast, which possesses two divergent *MAT1-1-1* genes (Lenassi et al., 2013). We hypothesize that the presence of three divergent *ZsMTL*α loci variants could arisen form two alternative events. One route may consist in an amplification of a chromosomal segment containing the ancestral linked *MAT*α1 and *MAT*α1 genes, resulting in paralogs which undergoing to progressive accumulation of mutations in the post-duplication period. Potentially, this duplication could also involve the entire sex chromosome due to chromosome missegregation during mitosis, leading to a diploid progeny with three chromosomes harboring progressively divergent *ZsMTL*α loci. In the second route, the acquisition of extra *ZsMTL* loci on homeologous sex chromosomes may take place after horizontal gene transfer (HGT) or interspecific introgression events. In fungi, interspecies *MAT* HGTs have been documented in clonal populations with increased adaptive phenotypes to new environments, but the mechanisms underlying these gene transfers are yet poorly understood (reviewed in

Richards et al., 2011). In the absence of significant evidence to preferentially support one of the proposed alternatives, we can not exclude any hypothesis about the generation of divergent *ZsMTLα* variants. Noteworthy, the amino acid sequence analysis reveals that substitutions among ZsMAT**a**1 or ZsMAT**a**2 copies are not randomly distributed since they mainly affect no functionally relevant regions. Accordingly, also most of key residues crucial for transcription activities of MATα1 and MATα2 in *S. cerevisiae* are still conserved in putative *Z. sapae* orthologs. These findings convey that no silent mutations in diverging *ZsMTL* genes have been under such kind of selective driving force aimed to maintain the functional integrity of encoded transcription factors. The retention of three divergent and putatively functional *ZsMTLα* loci variants could be favored by the divergent transcription of *MATα*1 and *MATα*2 from the intervening promoter located on the intergenic region within each variant. Alternatively, the acquisition of *ZsMTLα* extra loci could be very recently occurred via HGT without accumulation of deleterious mutations.

The second goal was to establish whether *Z. sapae* has a cassette system like that in *S. cerevisiae* and *Z. rouxii*. Consistent with the conservation of eight intein motifs and the amino acid residues involved in DNA binding, strain ABT301^{T} seems to possess two divergent *HO* genes, coding two putatively functional edonucleases. Again these data hint that both the *ZsHO* genes are under the same selective pressure and that SNPs in *ZsHO* copies 1 and 2 are selectively neutral mutations, with negligible effects on gene function. However in-vitro switching tests are advisable to prove this hypothesis. Moreover, the high degree of divergence observed between *ZsHO* copies 1 and 2 suggests that these genes did not arisen either from recent duplication or postulated hybridization events. Since all species that have *HO* genes have silent cassettes (Butler et al., 2004), we studied how the *ZsMTL* copies could be arranged as *MAT*, *HML* and *HMR* cassettes, by exploring the *ZsMTL* gene surroundings. Although most yeast species contain highly variable organization of mating type *MAT* locus and *HMR* and *HML* silent loci, pre-WGD species retain ancestral arrangement *DIC-MAT-SLA2* which distinguishes mating type *MAT* locus from silent

cassettes *HML* or *HMR* (Gordon et al., 2011). One *MTL***a** and three *MTL*α variants has been anchored to flanking regions by PCR amplification using one primer specific to Yα (copy 1, 2 and 3, respectively) or Y**a** together with a primer annealing on common neighboring sequences. Three *ZsMTL*α loci resulted duplicated in two syntenic patterns. One set, namely *ZsMAT*α copies 1, 2 and 3, exhibits the canonical synteny *DIC1-MAT-SLA2* conserved in the *MAT* expression loci of the majority of pre-WGD species (Gordon et al., 2011). The other set includes three *ZsMTL*α loci with an arrangement *DIC1-ZsMAT*α*-SLA2_D,* regardered as *ZsHML*_D. The *ZsMTL*α copy 1 locus also fits to *CHA1-MAT-SLA2* gene organization (*ZsHML* copy 1). The *SLA2* gene lays at the 3' end of *ZsMTL***a** locus while the gene at its 5' end remained unknown. Since the arrangement of *SLA2* gene on the right side of *MAT* is conserved in a number of pre-WGD (Butler et al., 2004; Gordon et al., 2011), we considered the *ZsMTL***a** locus as *MAT***a** expression locus. This hypothesis is supported by observing that in *S. cerevisiae* diploid cells express the MAT**a**1-MATα2 repressor, necessary to turn off the transcription of a set haploid-specific genes. As being ABT301^{T} a diploid strain, it should express MAT**a**1 with the same functional role. Our preliminary RT-PCR results indicate that *MAT***a**1 and *MAT*α2 are transcribed in ABT301^{T} strain in both standard and salt stressed conditions, excluding that the *ZsMTL***a** locus is a cryptic mating-type locus *HMR* (data not shown).

To explain the peculiar genetic make-up of mating system in *Z. sapae*, we inferred two nonexclusive scenarios of chromosomal arrangement, considering two constrains: i) that *MAT* and *HML* loci are linked in hemiascomycetes (Gordon *et al*. 2011); ii) *HMR* and/or *MAT* loci are located on different chromosomes in *Zygosaccharomyces* species (Fabre et al., 2005; Souciet et al., 2009; Watanabe et al., 2013). Based on model 1, diploid ABT301^{T} genome bears two genetically distinct sets of sex chromosome pairs, both lacking *HMR* cassettes. One set contains *MAT***a** and *MAT*α *Z. rouxii*-like sequences linked to *ZsHML_D* copy 1 and *ZsHML* copy 1, respectively. The other chromosome pair includes two slight divergent *MAT*α loci, namely Zs*MAT*α copy 2 and 3, linked with homologous *ZsHML*_D copy 2 and 3 loci, respectively (**Figure 4.11A**). In model 2, we

supposed that the diploid ABT301^T strain has an aα genotype, homozygous for the *MAT***a**-*HML* loci (*ZsMAT***a**-*HML_D* copy 1) and heterozygous for the *MAT*α-*HML* loci. In particular, the *MAT*α and *HML* loci could consist in similar alleles Zs*MAT*α copy 1-*ZsHML* copy 1 and Zs*MAT*α copy 2-*Zs*HML copy 2. Furthermore, consistently with this model, ABT301^T strain displays an homeologous extra-copy of sex chromosome (trisomy) which hosts the divergent cassettes Zs*MAT*α copy 3-*ZsHML_D* copy 3 (**Figure 4.11B**). This model implies that ABT301^T is not a euploid strain with a karyotype that is a multiple of the haploid complement, a status which partially disagrees with our previous data (Solieri et al., 2008a). By combining FACS and PFGE, strain ABT301^T and its conspecific ABT601 resulted to be diploid yeasts bearing additional number of chromosomes compared to the karyotype of *Z. rouxii* strains. Unfortunately, loss or gain of individual chromosomes would be not easily detectable even by combining FACS and PFGE, if the chromosomes are similar in size. Therefore, the hypothesis of additional copies of sex chromosomes bearing syntenic arrangement of Zs*MAT*α copy 3- *HML_D* copy 3 could be not excluded. However, in both models the lack of *HMR* cassette could suggests that ABT301^T may be unable to reproduce by haplo-selfing. The loss of *HMR* cassette has been previously documented in *S. cerevisiae* haploid cells, where mutation or deletion of the *MAT*α locus on chromosome III results in a reversion to the default *MAT***a** mating type, termed a-like fakers, allowing these *MAT*α cells to mate illegitimately with strains of the *MAT*α mating type (Strathern et al., 1982). This event involves mitotic crossover at a frequency of 3.1 x 10^{-6} (Hiraoka et al., 2000), leading to a deletion between *MAT* and *HMR* or a circular chromosome fusing *MAT* and *HML* (Hawthorne 1963; Strathern et al., 1979; Haber et al., 1980). α,α homozygous diploid strains have been found via same-mating sex in *Cryptococcus neoformans* (Lin et al., 2005) and via parasexual cycle in *C. albicans* (Magee and Magee, 2000; Wu et al., 2005; Forche et al., 2008). Among species having the silent cassette system, α,α,α strains have been found in *C. glabrata* (Srikantha et al., 2003), while α,α,α and α,α,α,α strains in *Z. rouxii* (Watanabe et al., 2013). In *Z. sapae*, a interchromosomal recombination may lead to the loss of *HMR* and the subsequent translocation of *ZsHOs* to the same

chromosomes harboring *ZsMTLα* loci, giving rise to a chromosomal configuration different from that of *Z. rouxii* CBS 732^{T}. The resulting aααα genotype is likely to produce a mating type imbalance, which determines the clonality as the main mode of reproduction observed in *Z. sapae* (Solieri et al., 2013a).

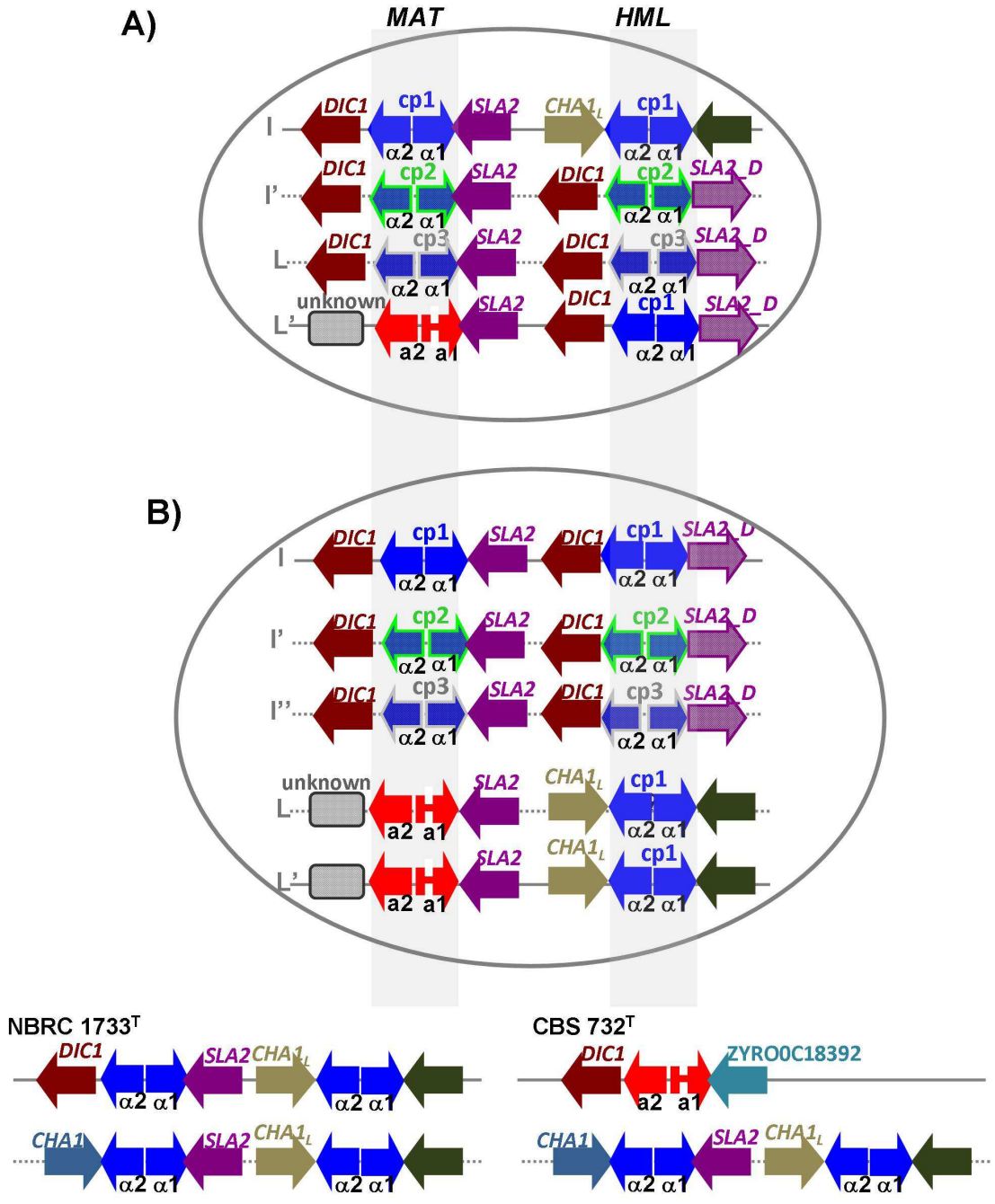


Figure 4.11. Inferred genomic organization of *Z. sapae* ABT301^{T}. A) First hypothesis considers a diploid genome with two sex chromosome pairs, namely I and I' and L and L', bearing *ZsMATα* cassettes 1, 2 and 3, and *ZsMAT***a**

cassette, which are linked to the putative silent *ZsHMLα*_D cassettes 1, 2 and 3 and to *ZsHML* copy 1, respectively. B) Second hypothesis considers an aneuploid number of chromosomes I. A set of three homeologous chromosomes, namely I, I', and I'', harbor *ZsMATα* cassettes 1, 2 and 3, arranged with *ZsHMLα*_D cassettes 1, 2 and 3, respectively. Chromosomes L and L' are homozygous for the *ZsMAT***a** locus, which is linked to a silent *ZsHML* copy 1 cassette. In both chromosomal arrangements, *ZsHO* copies 1 and 2 are located on chromosome set I. Chromosomal letters are according to Figure S2. Dotted arrows indicate divergent *ZsMTL* sequences from *Z. rouxii*. Chromosomal organization of three cassette system in *Z. rouxii* haploid strains NBRC 1733^{T} and CBS 732^{T} were reported for comparative purposes, according to Watanabe *et al*. (2013) and Souciet *et al.* (2009), respectively.

4.5. Conclusion

Our work provides a first insight to understand how the mating type system is arranged in *Z. sapae* diploid genome. A question much harder to be addressed, is to understand why the *Z. sapae* genome harbors a redundant number of divergent *MTLα* loci. Our hypothesis is that, although the *MAT* loci are typically non recombining regions (Idnurm, 2011), sex chromosome is a hotspot for DSBs, translocation and mutation in *Z. sapae*. As in the relative *Z. rouxii*, *Z. sapae* mating type information is arranged in two unlinked chromosomes which should favor outbreeding instead of inbreeding (Fraser and Heitman, 2003). Illegitimate recombination at these "hot spots" can be induced by the exposure of *Zygosaccharomyces* yeasts to environmental stresses such as high osmotic conditions. This hypothesis is consistent with the results recently reported for haploid *Z. rouxii* strains (Watanabe et al., 2013) and with many reports associating the increased DSBs frequencies with the gain of mutations and genome instability, including errors in DNA synthesis and microhomology-mediated jumps to ectopic templates (Hick et al., 2010). In *S. cerevisiae*, the *MAT*-bearing chromosome III was found to be the most unstable chromosome among all cases examined (Kumaran et al., 2013) and associated to haplo-insuffuciency (De Clare et al., 2010). When two specific DSBs are introduced simultaneously on separate chromosomes, DSBs-repair occurs via homologous recombination (with or without crossingover) (reviewed by Haber, 2006) and in the absence of homology via nonhomologous end joining (Yu and Gabriel, 2004), with reciprocal translocations and interchromosomal rearrangements. We speculate that under stress conditions imprecise mating switching and homeologous recombination between sex chromosomes

further enrich the repertoire of dodges that *Zygosaccharomyces* species can use to generate the wide range of genetic diversity, giving rise to aneuploid and/or euploid descendants with variable karyotype and diverse phenotype. The results may be a divergent adaptation associated with reproductive isolation in speciation (Dettman et al., 2007). Consistently with this thesis, increased recombination in response to stress (fitness-associated recombination; Hadany and Baker, 2003) is thought to promote the evolution of complex traits by accelerating the rate of adaptation in *S. cerevisiae* (Magowene et al., 2011) and *C. albicans* (Forche et al., 2011). The present study provides a methodological approach and sequence information to carry out a large-scale screening of *Z. sapae* and diploid *Z. rouxii*. This screening will be instrumental in verifying the role of genome instability and sex chromosome plasticity in stress adaptation.

Chapter 5:

Quantitative phenotypic analysis of multi-stress response in *Zygosaccharomyces rouxii* complex using *grofit* package

5.1. Introduction

Evolution and natural selection optimize an organism's genotype within the context of its environment. *Zygosaccharomyces rouxii* complex comprises three yeasts clusters sourced from sugar and salt-rich environments: haploid *Zygosaccharomyces rouxii*, diploid *Zygosaccharomyces sapae*, and allodiploid/aneuploid strains of uncertain taxonomic affiliations. Molecular mechanisms underlying cellular response to stress have been intensively characterized in model organism *Saccharomyces cerevisiae*, which exhibits moderate halo- and osmo-tolerance. In contrast, little efforts have been put in determining the molecular physiology in *Z. rouxii* (Watanabe et al., 1995, 2004; Iwaki et al., 1998, 1999; Pribylova et al., 2008). Recently, the release of *Z. rouxii* strain CBS 732^T genome sequence (Souciet et al., 2009) has boosted studies aimed at elucidating the molecular and genetic determinants involved in tolerance to harsh environments (Leandro et al., 2011; Stříbný et al., 2012; Hou et al., 2013; Watanabe et al., 2013; Wei et al., 2013). However, these studies are generally limited to few representative strains and, therefore, couldn't address the extent of physiological strain variation. Several lines of evidence suggest that *Z. rouxii* in yeast culture collections comprises at least three groups (globally referred to as *Z. rouxii* complex): the group of *Z. rouxii* type strains (as CBS 732^T), *Zygosaccharomyces sapae*, and an allopolyploid/aneuploid group (composed of ATCC 42981, CBS 4837 and CBS 4838) (Gordon and Wolfe, 2008; Solieri et al., 2013a, b). Whereas *Z. sapae* has been implicated in alcoholic fermentation of sugar-rich traditional balsamic vinegar (TBV) (Solieri et al., 2006), allodiploid/aneuploid strains ATCC 42981, CBS 4838, OUT7136, and CBS 4837 have been isolated from salty soy moromi and miso

(Wickerham and Burton, 1960; Kiuchi et al., 1980). Differently from *Z. rouxii* CBS 732^T, *Z. sapae* and allodiploid/aneuploid lineage display unusual rDNA heterogeneity with respect to the internal transcribed spacers (ITS) rRNA regions and/or the D1-D2 domains of large subunits (LSU) (Solieri et al., 2013b). Other differences can encompass whole chromosomal aneuploidy and polyploidy (Gordon and Wolfe, 2008; Solieri et al., 2013b). *Z. sapae* is diploid yeast with a genome size larger than those of the haploid CBS 732^T, allodiploid ATCC 42981, and aneuploid CBS 4837 and CBS 4838 strains. It has been demonstrated that changes in ploidy introduce potentially significant physiological effects, providing high or low fitness benefits mainly in highly stressful environments (Mable and Otto, 2001; Zeyl et al., 2003; Anderson et al., 2004). Although the *Z. rouxii* complex has been well characterized from a genetic point of view, the level of phenotypic diversity within this group has not yet studied, in relation to stress response, making it difficult to delimitate these lineages on the basis of physiological features. Some researchers have shown that high-dimensional molecular phenomics data sets can be leveraged accurately to predict phenotypic variation between strains, often with greater precision than afforded by DNA sequence analysis alone (Skelly et al., 2013).

One of the central principals of biology is the concept that a set of genetic instructions (or genotype) interacts with the environment to produce the characteristics (or phenotypes) of an organism. However, prediction of phenotype from genotype is generally a difficult task as large number of genes and gene products contribute to most phenotypes in concert with complex and changeable environmental influences. Phenomics, the study of the phenome, is a rapidly emerging area of science, which seeks to characterize phenotypes in a rigorous and formal way, and link these traits to the associated genes and gene variants (alleles). Because most phenotypes are determined by the interactions of genes and environment, the ideal situation is to collect large numbers of measures across multiple environments, at different growth stages, and for diverse strains or mutants. In order to better understand the quantitative characteristics and structure of

phenotypic diversity, measurement of gene expression, protein and metabolite abundance, and morphological traits in genetically diverse and mutant strains is generally taken into account (Skelly et al., 2013). Microorganisms are phenotypically assayed under a variety of conditions, depending upon the relevant question being asked. Warringer et al. (2003) presented a methodology for gene functional prediction based on extraction of physiologically relevant growth variables from all viable haploid yeast knockout mutants. This quantitative phenomics approach, applied to saline cultivation, identified marginal but functionally important phenotypes and allowed the precise determination of time to adapt to an environmental challenge (λ), rate of growth (μ), and efficiency of growth (A). They identified 500 salt sensitive gene deletions, the majority of which were previously uncharacterized and displayed salt sensitivity for only one of the three physiological features. They also reported a high correlation to protein–protein interaction data; in particular, several salt-sensitive subcellular networks indicating functional modules were revealed. This genome-wide analysis of hyperosmotic stress yielded 488 genes whose deletion increased sensitivity, including those responsible for glycerol production, ion homeostasis, cytoskeleton organization, signaling pathways and vacuolar protein transport.

Modern yeast biotechnology places a large emphasis on exploring potential biotechnological applications of so-called non-conventional yeasts, such as *Pichia*, *Zygosaccharomyces*, and *Kluyveromyces* (Porro and Branduardi, 2009). Within the genus *Zygosaccharomyces*, *Zygosaccharomyces rouxii* has a leading role in food industry both as fermentation-driving biocatalyzer and food spoiling agent. *Z. rouxii* strains have been also exploited as cell factories for producing enzymes (Kashyap et al., 2002; Iyer and Singhal, 2008, 2010), D-arbitol (Saha et al., 2007), chiral compounds (Vincenzi et al., 1997; Erdélyi et al., 2004, 2006), and heterologous proteins (Ogawa et al., 1990). Since industrial fermentations typically use complex substrates, tolerance to high osmo and electrolyte concentrations seems to be one of the essential traits a strain should possess. Many applications take advantage of *Z. rouxii* (and its

relatives) specific trait to grow in low a_w environments. These yeasts are constantly faced with environmental stresses and produce ethanol when growing in the presence of high salt (up to 3-4 M NaCl) or sugar (up to 5.0 M dextrose) concentrations (Martorell et al., 2007). Moreover, being petite-negative, *Z. rouxii* has been described as having a preferentially respiratory metabolism (Merico et al., 2007). By understanding how particular genotypes result in specific phenotypic properties is a core goal of modern biology that enables development of organisms with commercially useful characteristics.

In this study, we quantitatively analysed the stress response variation in *Z. rouxii* complex by modelling growth variables using *grofit* package. Besides this, the fitness of *Z. rouxii* complex strains has been quantitatively investigated under a pattern of nine single and six combined stress environments that comprise (i) alkali metal cations; (ii) glycerol consumption; and (iii) growth at 37°C. A crucial aim of this study was also to identify a reliable strategy to evaluate growth data collected in conditions highly different from the optimal ones. Therefore, both parametric models and model-free fits using cubic spline interpolation were implemented, tested, and compared.

5.2. Material and methods

5.2.1. Yeast strains

Eight strains were used in this study as listed in **Table 5.1**. Strains were obtained as lyophilized stocks from CBS (Centralbureau voor Schimmelcultures, Delft, the Netherlands) and ATCC (American Type Culture Collection). Strain OUT7136 was kindly provided by Prof. Y. Kaneko (Osaka University, Japan). *Zygosaccharomyces bailii* OUT7140 was used as reference strain or outgroup strain. Strains were routinely cultured on YPD agar slants [1% (w/v) yeast extract, 1% (w/v) peptone, 2% (w/v) dextrose, 2% (w/v) agar] at 27°C and maintained at 4°C for the duration of the study.

Strain	Taxonomic Affiliation	Source	rDNA heterogeneity[a]	Variant number		Genome size (Mb)	Ploidy
				SOD2	*HIS3*		
CBS 732^T	*Z. rouxii*	Must	-	1	1	9.8/12.7[b]	Haploid
ABT301^T	*Z. sapae*	TBV	+	2	2	28.1±1.3	Diploid
ABT601	*Z. sapae*	TBV	+	2	2	39.0±0.3	Diploid
ATCC 42981	Allodiploid	Miso	+	2	2	21.9±0.20	Diploid
CBS 4837	Allodiploid	Miso	+	2	2	21.7±0.33	diploid/aneuploid
CBS 4838	Allodiploid	Miso	+	2	2	22.5±0.20	diploid/aneuploid
OUT7136	Allodiploid	Soy moromi	-	2	2	19.57±0.47	Aneuploid
OUT7140	*Z. bailii*	Soy moromi	-	nd	Nd	Nd	Nd

[a] rDNA heterogeneity refers as to intra-individual polymorphisms including either internal transcribed spacers ITS1 and ITS2 or the large subunit (LSU) of rRNA gene (26S D1-D2) or both; [b] 9.8 Mbp is the result of final genome project assembly (Souciet et al. 2009), 12.7 Mbp is according to PFGE determination (Solieri et al. 2008a); Abbreviation: TBV, traditional balsamic vinegar; nd, not determined

Table 5.1. Strains details and genetic properties of *Z. rouxii* complex used in this study. Genetic features and taxonomical affiliation were reported according to Solieri et al. (2013b).

5.2.2. Preliminary phenotypic assays

Assimilation and fermentation tests were performed in triplicate according to standard methods (Kurtzman et al., 2011). For serial drop tests, early stationary phase pre-cultures (grown on YPD liquid medium at 27°C) were harvested by centrifugation and washed in 1 ml sterile MilliQ water. Five microliter of serially 10-fold dilution were spotted aseptically on YNB-NH_4-2%G [0.67% (w/v) yeast nitrogen base without amino acids (Difco), 2% (w/v) dextrose] solid medium supplemented either with NaCl or KCl at 2.5 and 3.0 M, respectively. For glycerol and dextrose tests, YP [1% (w/v) yeast extract, 1% (w/v) peptone] solid medium was supplemented with: 2.5% and 3.0% (v/v) glycerol; 60% and 70% (w/v) dextrose. Spots of 10-flod dilutions of the early stationary phase pre-cultures, on YPD agar plates were used as control condition. Plates were incubated at 27°C for 48-72 h and imaged using an Epson Expression 10,000 XL scanner operating in transmitted light mode.

5.2.3. Testing of the environmental stresses in batch experiments

Fifteen stress conditions were tested by growing the selected yeast strains in batch mode. Such experiments were performed in duplicate in 100 ml baffled Erlenmeyer (E-flask) flasks containing 50 mL of medium. Stress conditions and media are listed in Table 6.2. Water activity at 27°C of each medium (5 mL) was measured after autoclaving with a Novasin Lab MASTER-a_w electric hygrometer (Novasina, Pfaffikon, Switzerland). Basal medium YPD at 27°C±1°C was used as control condition. E-flask cultivations were preferred over microplates to assure sufficient aeration rate. Briefly, early-stationary cells were pre-cultured overnight as described above, re-suspended in peptone water, and transferred in the new medium with a starting OD (at 600 nm) equal to 0.07. E-flasks were kept at 27°C±1°C (or 37°C±1°C) and cultures subjected to orbital shaking at high intensity (200 rpm). OD measurements were manually performed at 600 nm on a Novospec II spectrophotometer (Pharmacia Biotech) and samples diluted at OD values from 0.4 to 0.6 (one OD unit corresponding to 2.7 x 10^7 CFU mL^{-1}).

5.2.4. Modeling of growth variables, parameter estimation and statistical analysis

To extract kinetic parameter values of λ (length of the lag phase), growth rate μ (corresponding to the slope), A (corresponding to the maximum cell growth), and the area under the curve (AUC), the growth curves resulting from stress tests were fitted by model-based and spline based approaches using the *grofit* package, (Kahm et al., 2010). The package *grofit* was originally built to derive dose-response curves and calculate descriptive pharmacological or toxicological values. Here, we focused on *grofit* intermediate output, which contains curve parameter estimates calculated both using model-based and model-free fitting approaches. Briefly, *grofit* package implemented logistic, Gompertz, modified Gompertz, and Richards mathematical models as model-based approaches, and compared non-linear least squares regression of each model according to the Akaike information criterion to select a best one. The model-free fit relied on cubic spline interpolation with the smoothing parameter smooth.gc. values set to default (i.e.,

NULL), 0.25, and 0.50, respectively. Finally, each growth curve was described by four sets of kinetic parameter values λ, μ, A, and AUC, one set resulting from the selected mathematical model, and the remaining three resulting from spline fitting with three different smoothing parameters. Full functionalities of *grofit* package are detailed at the function repository of R environment (http://cran.r-project.org/web/packages/grofit/).

Statistical analysis was performed using the software Statistica v.10 (Statsoft, Inc. Tulsa, USA). The correlation among *grofit* parameters was estimated using Pearson correlation coefficient (*r*), while the significance of their differences in different strains and/or conditions was quantified as confidence interval (CI) of the means, calculated at 95% level with normal distribution (1.96σ).

For each strain, the μ values obtained in control and stress conditions by setting the smoothing parameter to 0.25, were used to calculate the logarithmic strain coefficients LSC_{ctr} and LSC_{stress}, respectively, using *Z. rouxii* CBS 732^{T} as reference strain. LSC_{ctrl} and LSC_{stress} were compared forming the logarithmic phenotypic index (LPI_{rate}), as described in Warringer *et al.* (2003).

5.3. Results

5.3.1. Phenotypic assays: yeast serial drop tests

Z. rouxii complex includes seven strains split into three lineages, *Z. rouxii*, *Z. sapae* and, the allodiploid/anueploid group (Solieri et al., 2013b). These strains were originally isolated from two different environments such as sugar-rich Traditional Balsamic Vinegar and salty soy moromi (**Table 5.1**) and, here, they are submitted to physiological screening. Standard assimilation and fermentation tests showed that, other than discriminative phenotypic keys between *Z. sapae* and *Z. rouxii*, there was a slight inter-strain variation mainly with respect to rate in glycerol assimilation and growth response to high saline concentrations over time (**Supplementary Table S5.1**). These

results are in agreement with the previous observation that strain CBS 732^{T} and ATCC 42981 differ in tolerance to NaCl and glycerol catabolism (Pribylova et al., 2007a). End-point drop analyses confirmed the strain variability (**Figure 5.1**).

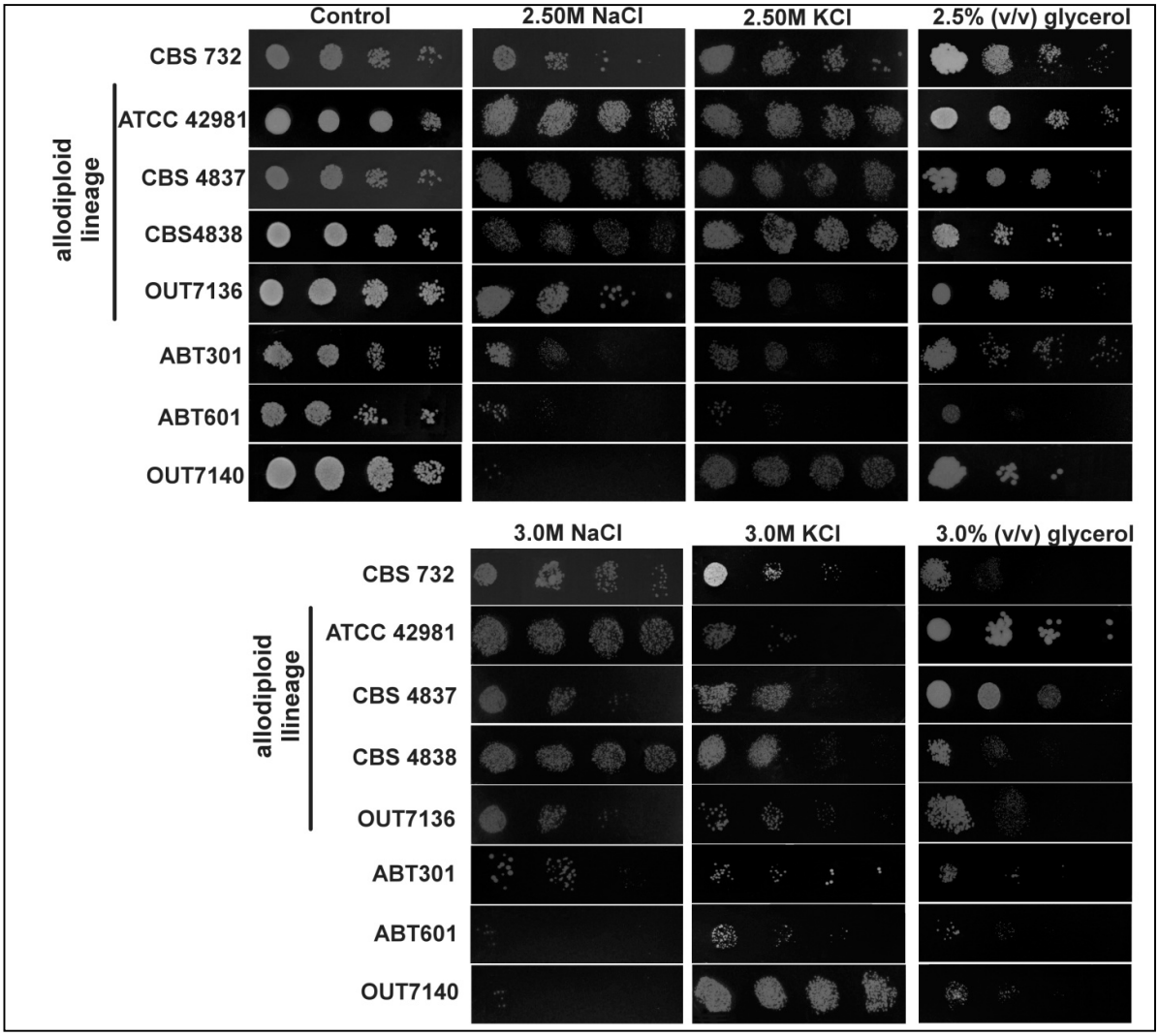


Figure 5.1. Yeast Serial Drop assays. Growth of *Z. rouxii* complex strains on YNB-NH_4-2%G plates supplemented with 2.5 and 3.0 M NaCl or KCl, and on YP agar plates supplemented with 3.5 and 3.0% (v/v) of glycerol, respectively. Growth on YPD agar plates was reported as control condition. A representative result from two independent experiments is shown.

When the strains were assessed for salt stress resistance, all strains grew up to the highest NaCl concentration (3.0 M) with different sensibilities. Allodiploid strains ATCC 42981, CBS 4838, and CBS 4837 were quite salt-tolerant, whereas *Z. sapae* ABT301^{T} and ABT601 resulted very sensitive to both salt concentrations. *Zygosaccharomyces rouxii* strains CBS732^{T} and

OUT7136 (allodiploid lineage) showed an intermediate phenotype, whereas, as expected, *Z. bailii* OUT7140 showed the most salt-sensitive phenotype. Slight inhibition of growth was observed on strains were exposed to 2.5 and 3.0 M KCl, with the exception of *Z. bailii*, which was the most fast-growing strain in both conditions. The third treatment was the exposure of the cells to glycerol as sole carbon source. *Zygosaccharomyces sapae* strains grew very slowly, while allodiploid ATCC 42981, CBS 4838, and CBS 4837 showed efficient glycerol assimilation with the resulting effective growth. No strains differences were detected when growing on 60 and 70% (w/v) dextrose (data not shown).

5.3.2. Parameter estimation from growth curves

In addition to end-point measurements, the tolerance of the *Z. rouxii* complex strains to environmental stress was analysed as quantitative changes in physiologically relevant growth variables, like time to adapt to environmental challenge (λ), kinetics of growth (μ), and efficiency of proliferation represented by stationary-phase cell density (A) (Warringer and Blomberg, 2003; Warringer et al., 2003). In particular, a panel of 15 environmental conditions was assessed (**Table 5.2**). These conditions were specifically selected to investigate whether there is a distinction among strains in adaptation to wide variations in salinity, a hallmark characteristic of all *Z. rouxii* complex strains, and those associated with handling other types of stresses. Growth parameters λ, μ, A, and AUC were estimated using both the model and the model-free spline fitting approaches of *grofit* R package (**Table 5.3 and Supplementary Table S5.1**).

Single stress	**Concentrations**			**Medium**
	1	**2**	**3**	
NaCl (M); [aw]	1.75; [0.925]	2.50; [0.901]	3.0; [0.885]	YNB-NH_4-2%G
KCl (M); [aw]	1.75; [0.927]	2.50; [0.905]	3.0; [0.892]	YNB-NH_4-2%G
Glycerol (v/v%); [aw]	3.00; [0.993]	4.0; [0.989]	-	YP
Combined stresses	**Concentrations**			**Medium**
	1	**2**	**3**	
Glycerol (v/v%); dextrose (w/v%); [aw]	3.50; 2; [0.985]	10.0; 2; [0.974]	20.0; 2;[0.941]	YP
Glycerol (%); NaCl (M); [aw]	3.0; 1.0; [0.947]	3.0; 2.0; [0.934]	-	YP
Glycerol (v/v%); $CuSO_4$ (μM); [aw]	3.0; 62.0; [0.939]	-	-	YP

Table 5.2. Environmental perturbations used for the quantitative phenotypic profiling.

Since the tested stress conditions affect significantly the strain survival, they provided growth curves without the typical sigmoid shape. Another potential problem relies on measurement artefacts induced by the presence of agglutinated cells which alters the growth curve shapes (Eilers and Marx, 1996). In *Z. sapae* cultures, mother and daughter cells have been reported to remain attached each other to form agglutinated flower-like structures (Solieri et al., 2013a). Moreover, stuck of fermentation induced by low levels of aeration imposed the use of E-flask over microplate. Therefore, the first aim was to identify a reliable estimation method that can efficiently deal with growth curves deviating from the standard sigmoid shape. The application of the parametric models to even slightly non-typical growth behaviours can lead to systematic errors and potentially to biologically unreasonable results (Gottschalk and Dunn, 2005; Vaas et al., 2012). The *grofit* R package implements both the best parametric model selected on the basis of Akaike criterion and the model-free fitting based on smoothed cubic spline, providing the opportunity to use both strategies to fit the growth curves. Depending on specific experimental stress conditions, 3.63% of the curves could not be fitted by the model-based approach, whereas ten experiments returned a negative estimate for λ. Nevertheless, negative λ values did not necessarily correspond to biologically unreasonable values for others parameters μ, A, and AUC. Indeed, although, in principle, biologically reasonable values for λ should not be negative or exceed the last time point of measurement, slightly negative λ values (mean between 3.3 and 3.0 h), can be judged as just negligibly inaccurate estimations of 0. The vast majority of negative λ occurred for A≤1, approximately representing negative reactions. In contrast, the spline fit yielded parameter estimates for every condition, but resulted in negative estimates for λ in 10.08% of the experiments (**Table 5.3**). The majority of them showed slightly negative values for λ, approximately representing negative reactions. In spline estimation, a major issue is the selection of an appropriate degree of smoothness. When using the default value for the smoothing parameter (i.e., smooth.gc.=NULL), *grofit* can estimate the optimal value of smoothness using a cross validation technique. However, this function has been described as a preliminary attempt and the use of a

fixed value is still recommended for datasets with few points, as those from batch cultivations in baffled E-flasks (Kahm et al., 2010). We run the spline fitting, using the default and two arbitrarily chosen values of the smoothness parameter (0.25 and 0.50) in order to reduce the number of un-interpretable λ values. Spline-fitting with smoothness parameters of 0.25 and 0.50 resulted in negative estimates for λ in 8.87% and 25.81% of the experiments, respectively. Particularly, setting the smoothing value to 0.25 returned slightly negative λ values only in highly stressful conditions, corresponding to either very irregular growth curves or intrinsically negative reactions.

Value	Dataset 1	Dataset 2	Tot
Total curves	124	124	248
Curves that cannot be fit by models	4.84%	2.42%	3.63%
Curves that cannot be fit by splines	0	0	0
Model parameter $\lambda < 0$	5.65%	2.42%	4.03%
Spline parameter $\lambda < 0$ (smooth.gc.=NULL)	12.10%	8.06%	10.08%
Spline parameter $\lambda < 0$ (smooth.gc.=0.25)	8.87%	8.06%	8.47%
Spline parameter $\lambda < 0$ (smooth.gc.=0.50)	25.81%	24.19%	25.00%

Table 5.3. Summary statistics from parameters estimation. Two datasets were analyzed using both the model and spline fitting approaches from the basic part of R's add-on package *grofit* (Kahm et al., 2010).

Pearson correlation coefficients between the curve parameters are listed in **Table 5.4**. In the model-fitting approach, the correlation between λ and the other parameters was quite low. Also, moderate correlations were observed between growth rate μ and efficiency of proliferation A (0.852) and between A and AUC (0.796). Similarly, when estimating the curve parameters using the spline fitting, λ resulted even less correlated to the other parameters, while the other correlations (μ *vs* A; A *vs* AUC) slightly increased to the range 0.904-0.933 and 0.810–0.816, respectively. Thus, on average λ was less strongly correlated with the other parameters in the case of the spline, whereas all other correlations were stronger. Accordingly, estimates for AUC correlated more strongly between model and spline, followed by A, μ, and finally λ in decreasing order. This evidence suggests that λ, as expected by its definition, is a weak parameter for discriminating between growing and not-growing conditions. Indeed, scatter plots of λ against any other parameter show the presence of multiple groups, depending on the final growth level.

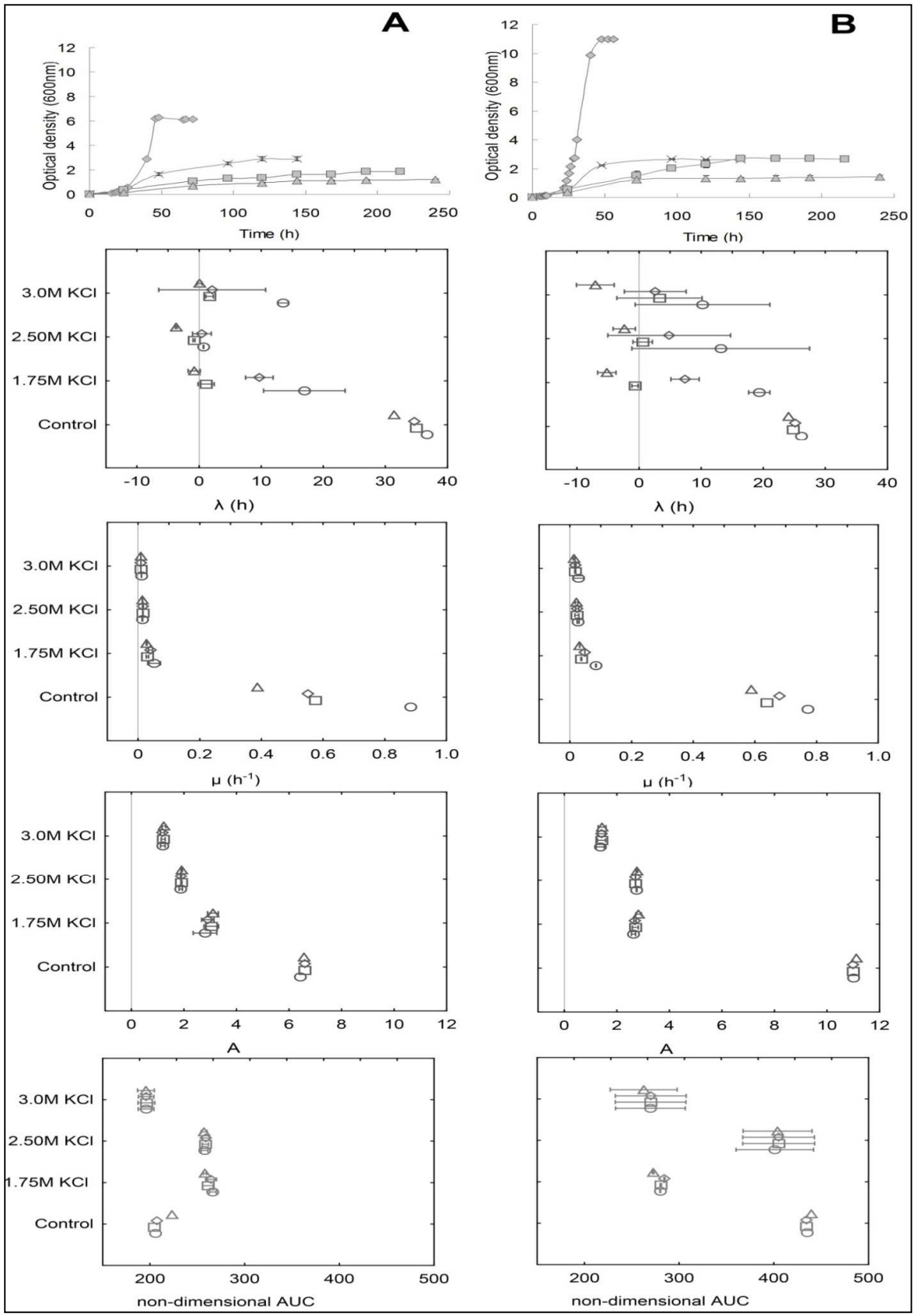


Figure 5.2. **Comparison of parameters and CI estimates from the same curve using four distinct fitting**

approaches. Left, plot of representative growth curves for *Z. sapae* ABT301^T (A) and allodiploid CBS 4837 (B) in basal medium (grey diamonds) and at 1.75 (grey plus), 2.50 (grey squares), and 3.0 M (grey triangles) KCl, respectively. Right, values and 95% CIs for each of the four parameters lag phase (λ), slope (μ), maximum (A) and area under the curve (AUC) estimated from the eight curves depicted on the left using model fitting (circles); spline approach smoot.gc=NULL (squares); spline approach smoot.gc=0.25 (diamonds); spline approach smoot.gc=0.50 (triangles)

		Model				Spline (smooth.gc.=NULL)				Spline (smooth.gc.=0.25)				Spline (smooth.gc.=0.50)				Mixed parameters set			
		μ	λ	A	AUC	μ	λ	A	AUC	μ	λ	A	AUC	μ	λ	A	AUC	μ	λ	A	AUC
Model	**μ**	**1.000**	0.240	0.852	0.588	0.946	0.183	0.871	0.593	0.958	0.158	0.869	0.596	0.910	0.297	0.868	0.608	**1.000**	0.240	0.852	0.588
	λ		**1.000**	0.334	0.233	0.227	0.718	0.307	0.248	0.226	0.735	0.309	0.240	0.199	0.898	0.323	0.223	0.240	**1.000**	0.334	0.233
	A			**1.000**	0.796	0.893	0.258	0.995	0.797	0.914	0.232	0.995	0.799	0.929	0.398	0.997	0.805	0.852	0.334	**1.000**	0.796
	AUC				**1.000**	0.623	0.169	0.809	0.999	0.629	0.148	0.812	1.000	0.632	0.182	0.800	0.999	0.588	0.233	0.796	**1.000**
Spline (smooth.gc.=NULL)	**μ**					**1.000**	0.196	0.904	0.624	0.984	0.160	0.901	0.629	0.945	0.310	0.903	0.643	0.946	0.227	0.893	0.623
	λ						**1.000**	0.232	0.179	0.187	0.904	0.234	0.172	0.162	0.783	0.243	0.163	0.183	0.718	0.258	0.169
	A							**1.000**	0.811	0.923	0.203	1.000	0.813	0.931	0.364	0.998	0.819	0.871	0.307	0.995	0.809
	AUC								**1.000**	0.632	0.157	0.814	1.000	0.632	0.193	0.802	0.999	0.593	0.248	0.797	0.999
Spline (smooth.gc.=0.25)	**μ**									**1.000**	0.162	0.922	0.636	0.968	0.318	0.925	0.649	0.958	0.226	0.914	0.629
	λ										**1.000**	0.206	0.151	0.134	0.811	0.215	0.142	0.158	0.735	0.232	0.148
	A											**1.000**	0.816	0.930	0.366	0.998	0.822	0.869	0.309	0.995	0.812
	AUC												**1.000**	0.635	0.186	0.804	0.999	0.596	0.240	0.799	1.000
Spline (smooth.gc.=0.50)	**μ**													**1.000**	0.316	0.933	0.648	0.910	0.199	0.929	0.632
	λ														**1.000**	0.377	0.179	0.297	0.898	0.398	0.182
	A															**1.000**	0.810	0.868	0.323	0.997	0.800
	AUC																**1.000**	0.608	0.223	0.805	0.999
Mixed parameters set	**μ**																	**1.000**	0.240	0.852	0.588
	λ																		**1.000**	0.334	0.233
	A																			**1.000**	0.796
	AUC																				**1.000**

In yellow correlation indeces discussed in the results section of the chapter

Table 5.4. All-against-all correlations measured by Pearson's correlation between model fitting and/or spline fit.

Figure 5.2 depicts different types of growth curves associated to stress and standard conditions for strains ABT301^T and CBS 4837. A slight and linear increase of absorbance values over time was observed in several experiments with alkali metal cations (KCl/NaCl concentrations ranging from 1.75 to 3.00 M). In some cases (i.e., 2.50 and 3.0 M KCl), curves described an intrinsically negative reaction (no growth), in others (1.75 M KCl), a reduced growth. Ideally, no growth would result in a horizontal line, and hence, in no convergence of any modelling approach. However, these linearly increasing noises and/or low signals were apparently sufficient to obtain estimates from both model and spline fits. In particular, spline computations with smooth.gc=NULL and 0.50 extended the lag phase even before the beginning of the measurement (at 0 h), resulting in negative values. According to the Akaike criterion, the modeling approaches found Gompertz/logistic and Gompertz as the best models to describe the 2.50 and 3.0 M NaCl data, respectively. They yielded positive, but highly variable estimates of lag phase, leading to broad CI. This suggests that the procedure is over-fitting the data and could have poor interpolative

properties. Spline fit with smooth.gc=0.25 reduced the λ negative values, compared to other spline fitting, and displayed a reduced over-fitting behaviour with respect to the model computation, thus resulting in smaller CI. The estimate for μ was slightly positive both from the model and spline fitting, whereas the other parameters were rather similarly overestimated by the different methods.

Our results indicate that the model-fitting approach is comparable with spline smoother in parameter estimation for growth curves arisen from batch fermentations in low stress conditions, but it is more prone to bias due to the irregularities in curve shapes detected at high stress conditions compared to spline fits (NA cases). Moreover, to select the smoothness parameter value using the default *grofit* cross-validation is not an efficient approach when analysing datasets with relatively few data points. On the basis of these observations, growth parameters derived from spline fitting with smoothness parameter manually set to 0.25 were selected for strains comparisons.

5.3.3. Strain variation in stress response

The complete quantitative survey of strain variation in growth parameters within *Z. rouxii* complex is shown in **Figure 5.3**. Growth comparison in basal medium at 27 °C (referred to as control condition) confirmed that *Z. sapae* ABT301^T and ABT601 strains grew poorly in comparison with ATCC 42981 and related strains CBS 4837 and CBS 4838. Moreover, maximum growth rate data (μ) suggested significant strain variation also within *Z. sapae* in growth kinetics, with ABT301^T and ABT601 being fast- and slow-growing, respectively. Otherwise, A and AUC parameters indicated that *Z. sapae* strains have a growth efficiency similar to those of *Z. rouxii* CBS 732^T and aneuploid OUT7136.

	CBS 732[T]					ATCC 42981					CBS 4837					CBS 4838				
Fermentation of	**3[§]**	**7**	**14**	**21**	**Final profile**	**3**	**7**	**14**	**21**	**Final profile**	**3**	**7**	**14**	**21**	**Final profile**	**3**	**7**	**14**	**21**	**Final profile**
Glucose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Fructose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Maltose	-	+	+	+	+	-	-	+	+	+	-	+	+	+	+	-	-	+	+	+
Lactose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Galactose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sucrose	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-
Trehalose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Growth on																				
Glucose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Fructose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Maltose	-	+	+	+	+	-	+	+	+	+	-	+	+	+	+	-	+	+	+	+
Lactose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Galactose	-	-	+	+	+	-	-	+	+	+	-	-	+	+	+	-	-	+	+	+
Sucrose	-	-	-	w	**w/l**	-	-	-	w	**w/l**	-	-	+	+	+	-	-	+	+	+
Trehalose	-	-	+	+	+	-	+	+	+	+	-	+	+	+	+	w	+	+	+	+
Xylose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ribose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Arabinose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Rhamnose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Melebiose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Raffinose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sorbitol	-	w	+	+	+	-	-	w	w	**w**	-	-	w	w	**w**	-	-	w	w	**w**
Inositol	-	-	w	w	**w**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Erythritol	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethanol	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Glycerol	-	w	w	w	**w**	-	+	+	+	+	-	+	+	+	+	-	+	+	+	+
Growth on																				
50% Dextrose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
60% Dextrose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Cycloeximide resistance	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1% Acetic Acid tolerance	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16% NaCl/5% dex	-	+	+	+	+	w	+	+	+	+	-	+	+	+	+	-	+	+	+	+
3%Gly_1M NaCl	-	-	w	+	**w/+**	-	+	+	+	+	-	+	+	+	+	-	+	+	+	+

	ABT301[T]					ABT601					OUT7136					OUT7140				
Fermentation of	**3**	**7**	**14**	**21**	**Final profile**	**3**	**7**	**14**	**21**	**Final profile**	**3**	**7**	**14**	**21**	**Final profile**	**3d**	**7**	**14**	**21**	**Final profile**
Glucose	w	+	+	+	+	w	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Fructose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Maltose	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-
Lactose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Galactose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sucrose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+
Trehalose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Growth on																				
Glucose	w	+	+	+	+	w	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Fructose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Maltose	-	-	+	+	+	-	-	+	+	+	-	+	+	+	+	w	+	+	+	+
Lactose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Galactose	-	-	+	+	+	-	-	w	+	+	-	-	-	+	+	-	-	+	+	+
Sucrose	-	-	w	w	**w**	-	-	-	w	**w/l**	-	-	+	+	+	w	+	+	+	+
Trehalose	-	-	-	-	-	-	-	-	-	-	-	-	-	+	**l**	-	-	+	+	+
Xylose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ribose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Arabinose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Rhamnose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Melebiose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Raffinose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	w	**w**
Sorbitol	-	w	w	+	+	-	-	w	+	**w/+**	-	-	w	w	**w**	-	+	+	+	+
Inositol	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Erythritol	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	w	**w**
Ethanol	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	w	**w**
Glycerol	-	w	w	+	+	-	-	w	+	**w/+**	-	+	+	+	+	-	+	+	+	+
Growth on																				
50% Dextrose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	w	+	+	+	+
60% Dextrose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	w	w	+	+	+
Cycloeximide resistance	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1% Acetic Acid tolerance	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
16% NaCl/5% dex	-	-	w	+	+	-	-	w	w	**w**	+	+	+	+	+	-	-	-	-	-
3%Gly_1M NaCl	-	-	w	+	**w/+**	-	-	-	w	**w/l**	-	+	+	+	+	-	-	-	-	-

+ positive: rapid growth after 1-2 weeks; l, delayed positive (latent), rapid growth after more than 2 weeks; s, slow positive: a positive growth develops slowly over a period exceeding 2 weeks; w, weakly positive growth; -, negative; [§]Time in days

Table 5.5. Fermentation and assimilation tests.

Next, we considered growth conditions divergent from the optimal ones for *Z. rouxii*, including the presence of alkali metal cations or copper, extreme temperature (37°C), alternative carbons source to dextrose and their combinations. When the temperature was shifted from 27 to 37°C, strains displayed a range of sensitivities with the *Z. sapae* strains resulting the most temperature-sensitive (growth rate ~ 0.25 h^{-1}), while the allodiploid strain ATCC42981 and *Z. rouxii* CBS 732^T the most tolerant (growth rates of 0.84 h^{-1} and 0.61 h^{-1}, respectively). Since the *Z. rouxii* complex strains showed a different pattern of efficiency in drop assays and in glycerol assimilation test (**Figure 5.1 and Table 5.5, respectively**), we also considered glycerol both in presence of dextrose and as unique carbon source. In mixed medium with dextrose, increase in glycerol concentration reduces the environmental a_w (**Table 5.2**) and enhances the fraction of the carbon source catabolized by respiration. As shown in the growth rate plot (**Figure 5.3**), 3.5% (v/v) glycerol in mixed medium did not induce any major difference among strains, whereas 20.0% (v/v) glycerol partitions the strains in two groups, i.e., those more effective in glycerol consumption (and more low a_w resistant) from those more defective in glycerol (and more sensitive to low a_w). Growth efficiency plot indicates that under these conditions, allodiploid strains overcame *Z. sapae* and were comparable with *Z. bailii* OUT7140 (**Figure 5.3**). When testing the strains ability to respire 3% (v/v) glycerol as unique carbon source, all strains displayed reduced growth rates and efficiencies of proliferation compared with control conditions. The same effect is more pronounced on medium containing 4% (v/v) glycerol. Functionally defective mitochondria could result in reactive oxygen species-mediated inhibition of respiration. To test this hypothesis, we assessed glycerol respiration in presence of copper. Copper is an essential cofactor by several mitochondrial enzymes in yeast, but, if present in high concentration, it also elicits potential cytotoxic effects, by reacting with hydrogen peroxide to form the highly reactive hydroxyl radical. Since growth analysis in dextrose medium supplemented with copper showed a decrease in yeast growth rate at copper levels >300 μM (Gamberi et al., 2009), we selected an optimal copper concentration of 62.0 μM (Kirchman and Botta, 2007). Addition of $CuSO_4$ to glycerol-containing medium affected

significantly yeasts fitness and split the strains into two clusters (**Figure 5.3**). One group includes strains CBS 4837, CBS 4838, and ATCC 42981 which improve glycerol respiration when supplemented with copper and display both growth rate and efficiency similar to those of control conditions. The other group includes yeasts, such as *Z. rouxii* CBS 732^T, *Z. sapae* and *Z. bailii* OUT7140, which were insensitive to copper when they are forced to respire. Considering that two different types of active-transport systems for glycerol uptake have been described, Na+/glycerol and H+/glycerol symports, not inhibiting concentrations of NaCl (1.0 and 2.0 M, respectively) were finally tested in presence of glycerol as carbon source in spite of dextrose. These combined stress conditions increased growth rate in allodiploid strains, compared with glycerol condition without NaCl, whereas completely inhibited *Z. sapae* and *Z. rouxii* CBS 732^T (**Figure 5.3**). These results suggest differences in glycerol uptake mechanisms between allodiploid strains and *Z. sapae*. Finally, cells were exposed to increasing concentration of NaCl and KCl. Alkali metal cations affected most significantly yeast fitness: the higher the salinity, the longer the lag-phase, the slower the rate of growth and the lower the stationary phase OD increment. This normal salt dependence of growth curves was observed for all the strains, which declined in growth rate and efficiency of proliferation when increasing concentrations of NaCl/KCl were added. KCl reduced growth rates and efficiency of proliferation in all strains (as compared to control condition), although different strains showed slightly different sensitivities, with *Z. bailii* being the most tolerant to the highest KCl concentration followed by allodiploid strains, *Z. sapae* and *Z. rouxii* CBS 732^T, respectively. Similarly, 1.75 M NaCl did not inhibit growth completely in all strains; whereas growth rates declined considerably when 3.0 M NaCl was added. Plots of growth rate and efficiency of proliferation highlighted that *Z. sapae* strains ABT301^T and ABT601 exhibit less halo-tolerance to 1.75 and 2.75 M NaCl than allodiploid strains ATCC 42981, CBS 4837, and CBS 4838, whereas strains OUT7136 and CBS 732^T have intermediate phenotypes. As expected, *Z. bailii* OUT7140 was unable to grow when supplemented with 1.75 M NaCl (**Figure 5.3**).

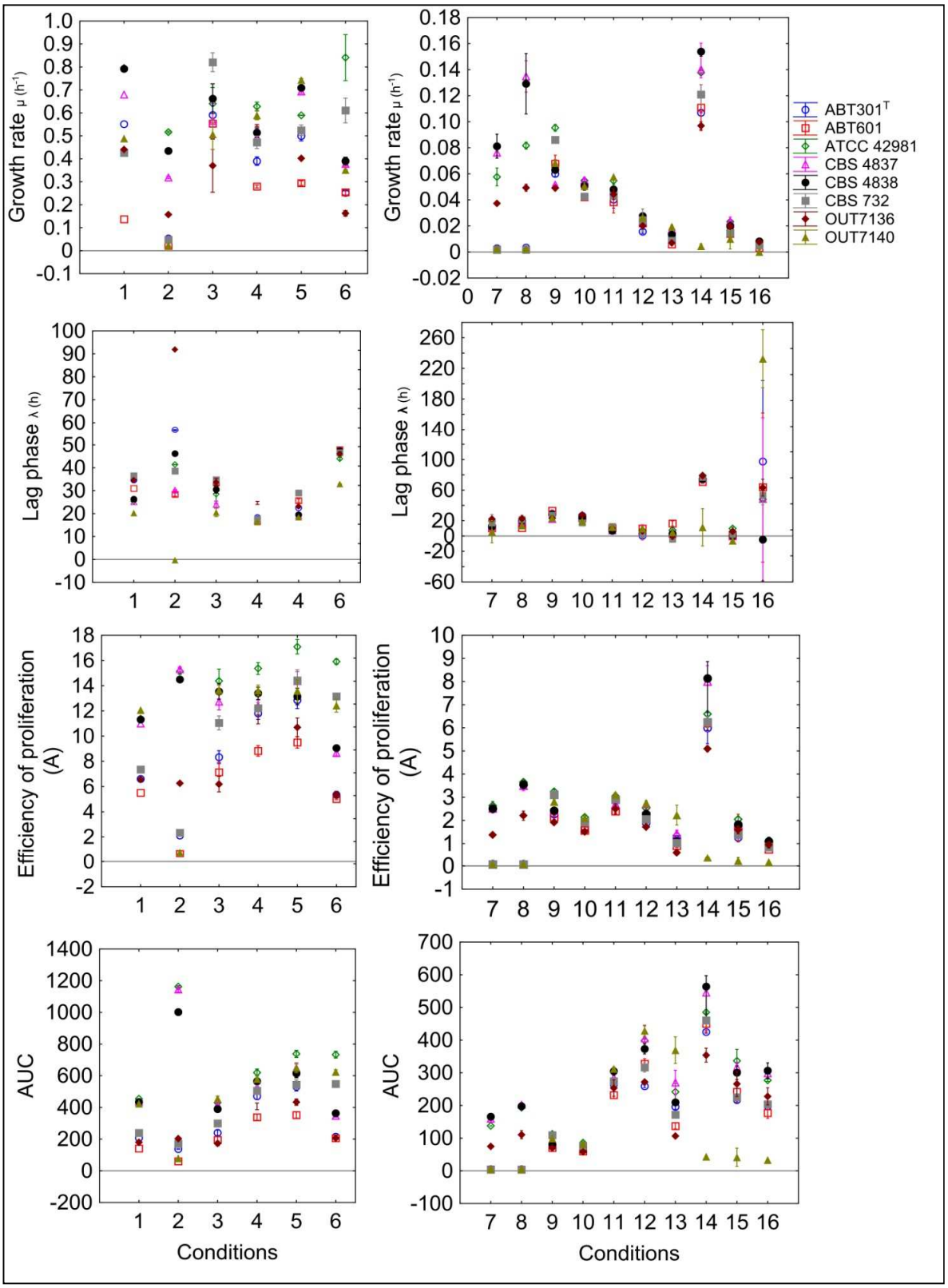
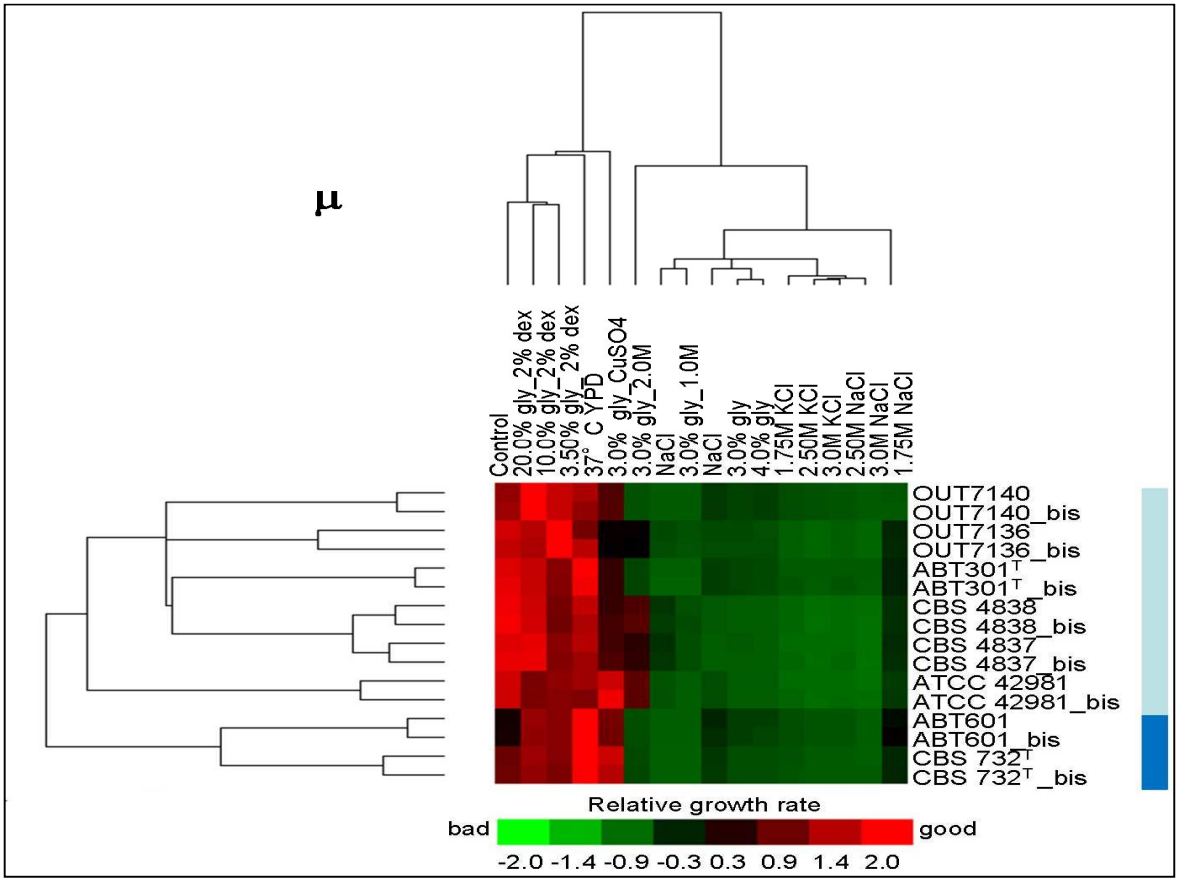
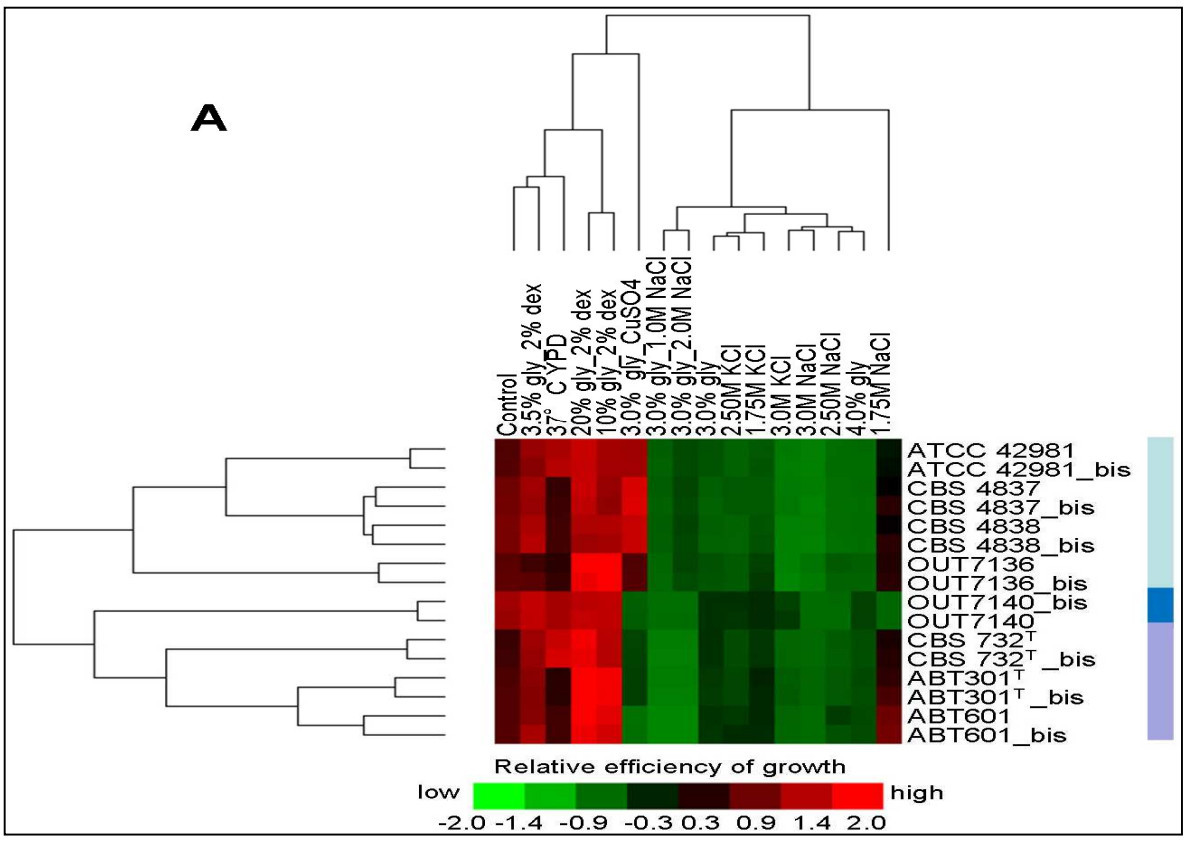


Figure 5.3. Growth traits variations within *Z. rouxii* complex. Plots of λ, μ, A, and AUC values and their 95% CI trains across 15 environmental perturbations were reported.

5.3.4. Clustering analysis and logarithmic phenotypic index

To underline the high degree of phenotypic variation among the various strains we mapped trait information and compared it with the phylogenetic characteristics. Specifically, strains of the *Z. rouxii* complex and growth conditions have been clustered using the quantitative variables, i.e., growth rate, efficiency of proliferation, adaptation time, and AUC. Grouping strains using μ values returned two main clusters, one composed of ABT601 and *Z. rouxii* CBS 732^T, the other including allodiploid/aneouploid strains and *Z. sapae* ABT301^T (**Figure 5.4**). Within the latter cluster, ATCC 42981 significantly differed from the other members due to its high temperature resistance. Furthermore, the low growth of ABT601 in basal medium contributed to diverge this strain from the conspecific ABT301^T. As previously mentioned, lag phase computation via model and spline methods in highly stress conditions results either in unreasonable negative estimates or over-fitting leading to broad CI range. A not entirely satisfactory fitting of the data for ABT301^T, ABT601 and CBS 4837 strains occurring at 3.0 M NaCl and KCl (where a linear increasing of absorbance was observed over time), resulted in an incongruent segregation of replicates (**Supplementary Figure. S5.1**). As such, these conditions were further omitted from the λ-based clustering. The resulting analysis segregated *Z. bailii* OUT7140 as a separate single strain, according to its taxonomical attribution, whereas the remaining strains were poorly discriminated based on their adaptation ability to environmental stress. The same strains were grouped in three main groups according to A and AUC values, namely *Z. sapae*/*Z. rouxii* CBS732^T, *Z. bailii* and allodiploid/aneuplouid one (**Figure 5.4**). These groups correspond with clusters of *Z. rouxii* complex previously defined analysing genetic markers and genomic features (Solieri et al., 2013b).


μ
Control
20.0% gly_2% dex
10.0% gly_2% dex
3.50% gly_2% dex
37° C YPD
3.0% gly_CuSO4
3.0% gly_2.0M NaCl
3.0% gly_1.0M NaCl
3.0% gly
4.0% gly
1.75M KCl
2.50M KCl
3.0M KCl
2.50M NaCl
3.0M NaCl
1.75M NaCl
OUT7140
OUT7140_bis
OUT7136
OUT7136_bis
ABT301T
ABT301T_bis
CBS 4838
CBS 4838_bis
CBS 4837
CBS 4837_bis
ATCC 42981
ATCC 42981_bis
ABT601
ABT601_bis
CBS 732T
CBS 732T_bis
Relative growth rate
bad
good
-2.0 -1.4 -0.9 -0.3 0.3 0.9 1.4 2.0


A
Control
3.5% gly_2% dex
37° C YPD
20% gly_2% dex
10% gly_2% dex
3.0% gly_CuSO4
3.0% gly_1.0M NaCl
3.0% gly_2.0M NaCl
3.0% gly
2.50M KCl
1.75M KCl
3.0M KCl
3.0M NaCl
2.50M NaCl
4.0% gly
1.75M NaCl
ATCC 42981
ATCC 42981_bis
CBS 4837
CBS 4837_bis
CBS 4838
CBS 4838_bis
OUT7136
OUT7136_bis
OUT7140_bis
OUT7140
CBS 732T
CBS 732T_bis
ABT301T
ABT301T_bis
ABT601
ABT601_bis
Relative efficiency of growth
low
high
-2.0 -1.4 -0.9 -0.3 0.3 0.9 1.4 2.0

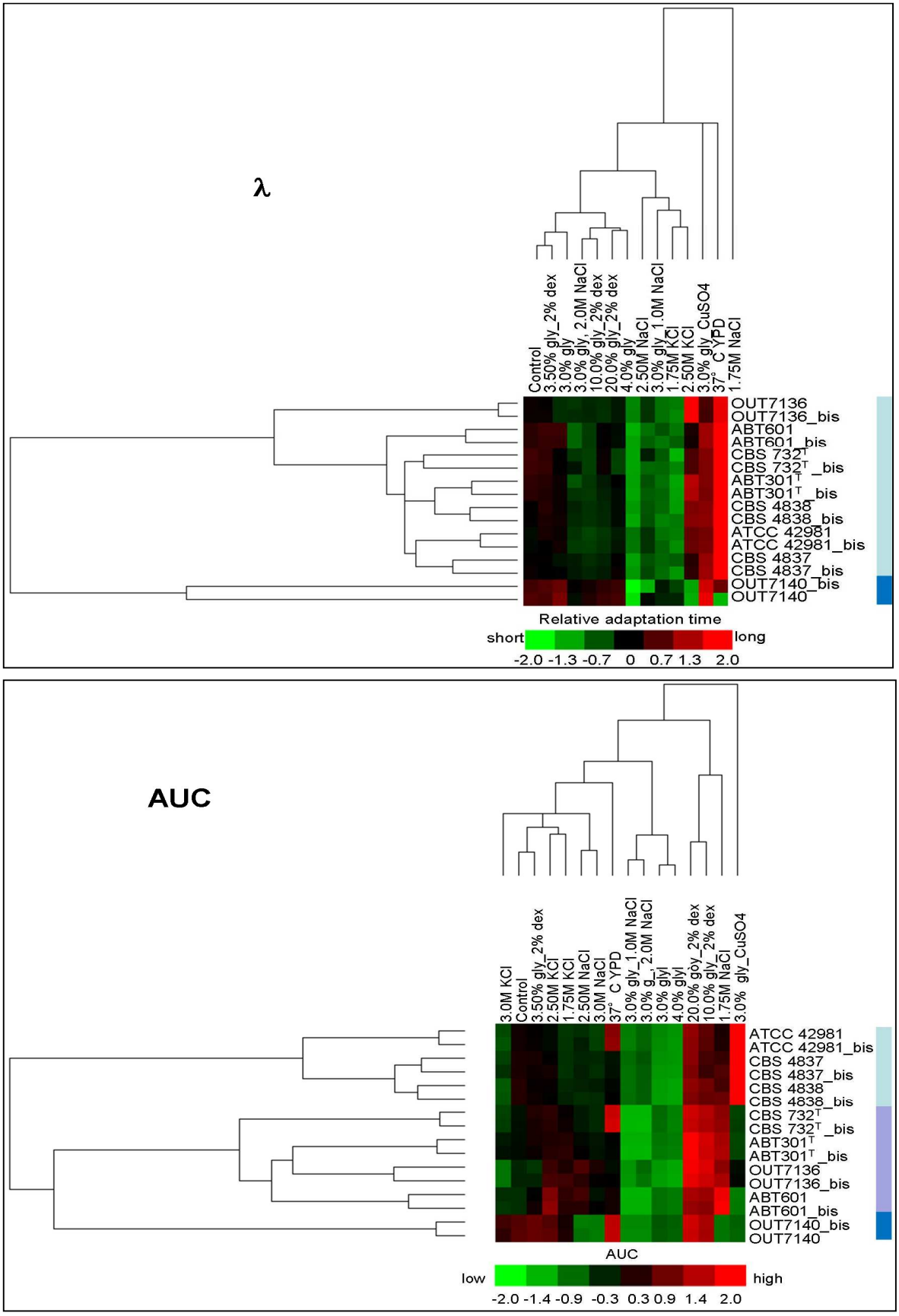


Figure 5.4. **Clustering of *Z. rouxii* complex strains based on μ (a), λ (b), A (c), and AUC (d) estimates.**

Dendrograms and heat-maps have been obtained using a hierarchical agglomerative clustering with eucledian and centroid as distance and linkage metrics, respectively. Green and red color shades represent: short and long relative adaptation time (λ), good and bad growth rate (μ), low and high efficiency of proliferation (A), and small to large areas under the curve (AUC). Highly stressful conditions 3.0 M NaCl and KCl were omitted from λ-derived clustering. Major strain groups are indicated as light blue, blue, and violet, respectively. Abbreviations: gly, glycerol; dex, dextrose.

Cluster analysis pointed out that, at least for growth rate, a more realistic comparison of the inter-strain differences requires normalizing the trait variables with respect to a reference strain and the behaviour in the basal medium. Therefore, $LSC_{[ctrl]}$ and $LSC_{[stress]}$ values for each strain were calculated normalizing the μ value in a particular environment to the corresponding average growth rate of reference strain *Z. rouxii* CBS 732^T (Warringer et al., 2003). The scatter plot of $LSC_{[ctrl]}$ and $LSC_{[stress]}$ (**Figure 5.5A**) indicates that *Z. sapae* ABT601 exhibit a growth defect both in control and stress conditions ($LSC_{[ctrl]}$ <0 and majority of $LSCs_{[stress]}$ ≤0). *Z. sapae* $ABT301^T$ did not show general growth defect in standard medium, but exhibits the same pattern of stress responses as strain ABT601 ($LSC_{[ctrl]}$ >0 and majority of $LSCs_{[stress]}$ ≤0). Significantly, allodiploid strains ATCC 42981, CBS 4838, and CBS 4837 have significant fitness in control ($LSC_{[ctrl]}$ >0) and stress conditions (majority of $LSC_{[stress]}$ values >0). Strain OUT7136 displays stress sensitivity similar to allodiploid strains (majority of $LSCs_{[stress]}$ >0), but growth fitness in standard condition similar to that of *Z. rouxii* CBS 732^T ($LSC_{[ctrl]}$ ~0). Subsequently, LSCs with and without stress were combined to distinguish the specific stress growth defect of a strain from a general growth defect (**Figure 5.5B**). The resulting LPI_{rate} gave a sensitivity measure for each strain to stress, in relation to the response of the reference strain CBS 732^T. Alteration in LPI_{rate} under many of the conditions confirmed only partially the previous results on LSC. When comparing the 15 stress environments, strain $ABT301^T$ resulted the most sensitive (highest number of negative LPI_{rate} values), followed by the conspecific strain ABT601. This latter strain displayed growth defects in standard condition, but stress environments did not proportionally affect its growth (LPI_{rate} > 0). Strains ATCC 42981, CBS 4837, CBS 4838, and OUT 7136 were confirmed to be a homogeneous group of stress-resistant yeasts that produced the same pattern of alterations in LPI_{rate} under stress conditions.

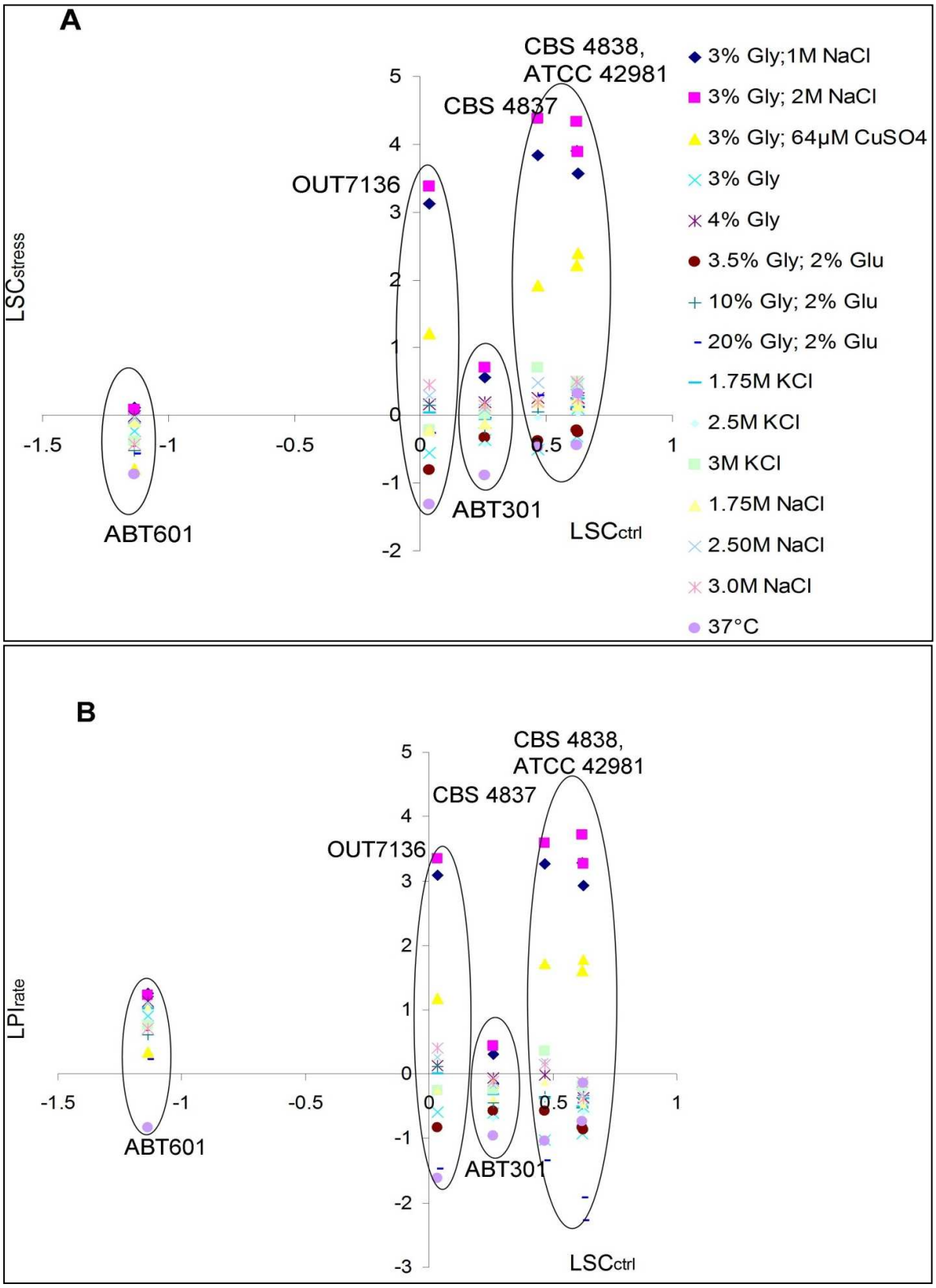


Figure 5.5. Scatter plots of LC and LP indices. Growth defects in standard condition (LSC_{ctrl}) were plotted *versus* growth defects in stress conditions LSC_{stress} (A) and LPI_{rate} (B), respectively.

5.4. Discussion

The capacity to adapt to adverse environmental conditions is an important trait for industrial yeasts, which should withstand sub-optimal conditions typical of industrial bioprocesses. *Z. rouxii* complex yeasts inhabit low a_w food-related environments and are subject to diverse selective pressures, as highly saline and sugary ones. A deeper understanding of the physiological and genomic variations within this complex would enhance the industrial exploitation of these yeasts and, somehow, direct the design of new strategies to overcome yeast spoilage in food environments. However, genomic information is limited, with a single genome sequence available for strain CBS 732^T (Souciet et al., 2009), whereas phenotypic studies have often considered a moderate number of phenotypes or have been limited to only a few individuals (Pribylova et al., 2007a, b, 2008). Another frequent limitation in these investigations is the ability to accurately quantify the phenotype of interest. Previous works highlighted that *Z. rouxii* complex is highly plastic, with diploid *Z. sapae* strains showing higher chromosome number and genome size than allodiploid/aneuploid ones (James et al., 2005; Gordon and Wolfe, 2008; Solieri et al., 2013b). Here, we investigated the differences of strains from the *Z. rouxii* complex in physiological response to industrial stress environments by quantitative phenotypic profiling using the *Grofit* package to quantify growth curves obtained across 15 stress environments. When comparing the spline and model approaches for curve description, it turned out that the spline smoother is more flexible than the model-fit to estimate growth parameters even for extreme curve shapes (highly stressful conditions). Therefore, based on the ability to describe also very irregular curves, we favour the spline-based approach for parameter estimation over the model-based one. However, both approaches presented some limitations to correctly compute the lag phase, i.e., the λ parameter. Our results indicate that the quantification of λ is more influenced by possibly uncommon shapes, induced by high stressful conditions, than the other parameters μ, A, and AUC, resulting both in negative values and broad CI. Modelling of bacterial lag time is complicated

because the mechanisms governing lag time are not fully understood (Swinnen et al., 2004; Kosekia and Nonaka, 2012). It is known that bacterial lag time is influenced not only by current environmental conditions but also by multiple other factors, such as previous growth conditions (Swinnen et al., 2004; Ryal et al., 2012), initial cell counts (Baranyi and Pin, 1999), and stochastic variability (McKellar and Hawke, 2006; Koutsoumanis, 2008; Ginovart et al., 2012). The parameters selection for the smoothing spline and the number of knots is the critical steps when using a smoothed cubic spline fitting (Eilers and Marx, 1996). In some studies where microbial growth data is modelled via cubic spline, the smoothing parameter has been either arbitrarily chosen or estimated via cross-validation technique (Oner et al., 1986; Seckse et al., 2012). Here, the high number of unreasonable negative λ values obtained when using the smoothing spline method suggested modifying the default smoothing parameter of *grofit* (Eilers and Marx, 1996). In a previous study, Vaas et al. (2012) observed a high number of spline-derived negative lambda values and broad CI in high-throughput datasets and indicated the need of an algorithmic improvement in *grofit* estimation of the default smoothing parameter. Despite the evaluation of suitable degree of smoothness was beyond the aim of this study, we empirically selected the smoothness parameter that reduced as possible the number of negative λ and narrowed the corresponding CIs. The manual selection of the smoothing parameter reduced the number of biologically unreasonable lag phase values, but only partially corrected the large CIs under limit growth conditions (negative reactions). These evidences suggests that further improvements of the *grofit* package could be focused on fitting of a single model or spline on a set of several repetitions of a dataset, implementing additional estimation methods for μ and λ, and using information from the second spline derivative (Vaas et al., 2012).

Analysis of different physiological aspects of growth responses to stress revealed a marked strain variation in *Z. rouxii* complex, which partially supports the previous segregation of strains in two different clusters based on molecular profiling, i.e., the allodiploid group and the *Z. sapae*.

Generally, *Z. sapae* is less responsive to saline adaptation than allodiploid strains. Otherwise, no differences were detected in response to osmotic pressure mediated by sugars. These results suggest that the two groups differ in mechanisms counteracting high ionic strength and low a_w and confirm that physiological response to sugar stress may not be as effective for salt stress (Lages et al., 1999). The modulation of membrane transport activity is a major mechanism that yeast cells exploit to compensate saline stress (Ramos et al., 2011). However, copy number variation and polymorphisms in *ZrSOD* gene coding Na+/H+ antiporter have been not yet clearly linked to variation in saline tolerance (Watanabe et al., 1995; Iwaki et al., 1998). Although strain ATCC 42981 possesses two divergent copies, namely *ZrSOD2* and *ZrSOD22,* which restore salt-tolerance in a *S. cerevisiae* salt-sensitive mutant, only *ZrSOD22* was reported to significantly contribute to salt tolerance in ATCC 42981 (Iwaki et al., 1998). Interestingly, *Z. sapae* ABT301^T and ABT601 also possess *ZrSOD22* in addition to the variant *ZrSOD2-22*, that occur in CBS732^T genome (Solieri et al., 2008a), but did not display a salt tolerance comparable with ATCC 42981. This evidence suggests that *ZrSOD22* could be irrelevant in conferring salt resistance. Recently SNPs in *GPD1* gene coding for glycerol-3-phosphate dehydrogenase and *FPS1* gene coding for a putative glycerol transporter were found responsible of improving salt-tolerance in a *Z. rouxii* mutant (Hou et al., 2013), confirming that production, accumulation, and retention of metabolically compatible glycerol trigger saline stress response (Lages et al., 1999; Watanabe et al., 2004; Tang et al., 2005). Since the presence of 1.0-2.0 M NaCl increases intracellular glycerol accumulation and the subsequent respiration in allodiploid group, but not in *Z. sapae* and *Z. rouxii* CBS 732^T, we can speculated that these yeasts differ in *FPS1* gene and/or, more in general, in glycerol uptake. Additionally, copper supplementation increases growth rate under conditions requiring respiratory metabolism only in allodiploid strains. Kirchman and Botta (2007) highlighted a similar copper-mediated effect in the life span of *S. cerevisiae* and hypothesized that copper may increase the functionality of cytochrome-c oxidase and reduce the generation of ROS. The insensitivity of *Z. sapae* to copper supplementation disagrees with the previously described respiration metabolism in

Z. rouxii (Merico et al., 2007) and suggests mitochondrial dysfunctions with the consequent respiration deficiency, although other factors could not be excluded. For example, auxotrophic mutations in *S. cerevisiae* has been reported to determine increasing sensitivity to copper and reduction of the stability of the mitochondrial DNA, due to a partial deletion of the promoter for the divergently transcribed *MRM1* gene, which encodes mitochondrial rRNA methyltransferase (Sirum-Connolly and Mason, 1993).

Clustering and LSC/LPI analysis provided groups that essentially resembled but not precisely overlap lineage genetic boundaries, delineating allodiploid strains and *Z. sapae* as separate clusters from CBS 732^T. In particular, a partial incongruence can be observed between molecular based segregation of allodiploid strains and *Z. sapae* and phenotypic profiling. Clustering based on μ and λ supported the segregation of $ABT301^T$ with allodiploid/aneuploid group, whereas clustering via A and AUC segregated $ABT301^T$ and ABT601 into the same group. Taking into account that the two strains resemble each other for genotypic and phylogenetic markers, but differ for genome content (Solieri et al., 2008a), this incongruence may reflect the inter-strain difference between $ABT301^T$ and ABT601. Accordingly, LC and LP indices showed that ABT601 is a slow-growing organism insensitive to stress perturbations, whereas strain $ABT301^T$ has a superior proliferative rate on rich medium, but is more sensitive to sub-optimal conditions.

Finally, the marked phenotypic variations underlined in this study could be linked to genetic variations previously reported for this pool of strains. Several lines of evidence showed that genetic and phenotypic variability may be induced by stress (reviewed by Berman and Hadany, 2012). Since it is known that environmental stress drives phenotypic evolution and mating (Badyaev, 2005), we speculate that stress conditions increase phenotypic and genetic variance in *Z. rouxii* complex. Variations in the ploidy state (haploid, aneuploid or diploid/allodiploid) and in chromosome number have been reported to be advantageous in fungi by enabling rapid adaptive

evolution (Pavelka et al., 2010). Aneuploidy facilitates adaptive evolution in yeast cells and chromosomal duplication may confer a selective advantage to *S. cerevisiae* under stress conditions by promoting genomic instability and mutation (Sheltzer et al., 2011; Yona et al., 2012). In a previous work, we proposed that a similar inbreeding system of *Zygosaccharomyces* could lead to aneuploidy and allodiploidy (Solieti et al., 2013b). We hypothesize that *Zygosaccharomyces* reproduction is a hypermutagenic process that generates genetic diversity aimed to produce variable progenies adapted to stress environments. Whether there is a link among genetic and phenotypic variation and inbreeding reproduction requires further investigations. The quantitative survey of fitness carried out in this work represents a step toward a better understanding of the relationships between phenotype and genome-level traits in non-conventional yeasts alternative to *S. cerevisiae*.

5.5. Conclusion

Our results showed that allodiploid/aneuploid strains ATCC 42981, CBS 4837, and CBS 4838 are markedly different from diploid *Z. sapae* and haploid CBS 732^{T}. The quantitative survey of fitness carried out in this work represents a step toward a proper exploitation of these yeasts as food starters or for preventing their growth as spoilage yeasts. Moreover, the study could contribute to better understand the relationships between phenotype and genome-level traits in poorly studied non-conventional yeasts. In a previous work, we proposed that the inbreeding system could lead to aneuploidy and allodiploidy in *Zygosaccharomyces* (Solieti et al., 2013b). Since genetic and phenotypic variability may be induced by stress (reviewed by Berman and Hadany, 2012), we hypothesize that *Zygosaccharomyces* reproduction is a hypermutagenic process that contribute to stress adaptation by generating progenies with different genetic and phenotypic outcomes. The generation of divergent lineages could be a successful strategy under stress to increase the probability to achieve descendants improved in adaptation to hostile environments. Whether there is a link among genetic and phenotypic variation and inbreeding reproduction in *Z. rouxii* complex requires further investigations.

Strain	Growth condition	Model-based fit								Spline-based fit - smooth.gc=NULL				Spline-based fit - smooth.gc=0.25				Spline-based fit - smooth.gc=0.5			
		reliability	*model*	*mu*	*lambda*	*A*	*AUC*	*mu STD*	*lambda STD*	*A STD*	*mu*	*lambda*	*A*	*AUC*	*mu*	*lambda*	*A*	*AUC*	*mu*	*lambda*	*A*	*AUC*
ABT 301	Control	TRUE	logistic	0.88	36.71	6.43	205.83	0.31	1.23	0.16	0.57	34.95	6.60	204.33	0.55	34.69	6.61	207.30	0.39	31.40	6.57	223.23
ABT 301	Gly_3%_1M_NaCl	TRUE	logistic	0.00	11.52	0.07	4.94	0.00	11.11	0.01	0.00	13.80	0.08	5.09	0.00	13.54	0.08	5.09	0.00	1.90	0.08	5.10
ABT 301	Gly_3%_2M_NaCl	TRUE	logistic	0.00	12.06	0.07	4.91	0.00	11.61	0.01	0.00	5.63	0.08	5.07	0.00	12.93	0.08	5.06	0.00	2.26	0.08	5.08
ABT 301	Gly_3%_CuSO4	TRUE	logistic	0.05	51.16	2.10	131.42	0.01	4.94	0.10	0.05	51.78	2.06	136.68	0.05	56.83	2.09	136.48	0.03	41.32	2.09	135.53
ABT 301	Gly_3%	TRUE	logistic	0.07	30.28	2.39	75.76	0.00	1.40	0.06	0.07	31.46	2.39	75.53	0.06	28.81	2.32	76.87	0.05	24.35	2.38	75.50
ABT 301	Gly_4%	TRUE	richards	0.06	30.46	1.69	66.56	0.01	1.99	0.03	0.06	28.11	1.75	67.39	0.05	26.90	1.74	67.39	0.04	19.03	1.74	66.37
ABT 301	Gly_3.5%_glu	TRUE	gompertz.exp	1.28	36.99	7.67	237.01	0.19	0.47	0.52	0.44	29.13	8.68	256.19	0.61	32.91	8.59	248.07	0.28	21.58	8.92	260.61
ABT 301	Gly_10%_glu	TRUE	logistic	0.82	21.75	10.90	467.77	0.45	2.13	0.55	0.36	10.58	12.09	477.90	0.40	18.16	12.04	482.39	0.34	13.63	11.27	472.40
ABT 301	Gly_20%_glu	TRUE	logistic	0.79	25.39	11.15	523.70	0.49	3.47	0.65	0.51	22.54	13.16	549.12	0.51	22.67	13.18	544.89	0.37	17.55	12.11	539.84
ABT 301	KCl 1_75M	TRUE	gompertz	0.04	13.61	3.03	267.98	0.01	6.53	0.19	0.03	0.46	3.19	258.27	0.04	8.57	3.03	261.68	0.03	-0.27	3.21	257.08
ABT 301	KCl 2_5M	TRUE	gompertz	0.01	0.74	1.91	255.85	0.00	9.32	0.11	0.02	-0.73	1.92	256.78	0.01	-0.31	1.90	256.99	0.01	-3.61	1.93	255.54
ABT 301	KCl 3M	TRUE	gompertz	0.01	13.10	1.15	191.69	0.00	6.63	0.04	0.01	1.37	1.18	191.70	0.01	6.50	1.15	191.80	0.01	-0.12	1.20	191.21
ABT 301	NaCl 1.75M	TRUE	logistic	0.22	88.61	5.71	401.35	0.11	4.92	0.47	0.14	80.16	5.57	441.86	0.11	71.38	5.65	426.75	0.05	47.29	5.95	392.48
ABT 301	NaCl 2.5M	TRUE	gompertz	0.02	7.36	1.24	211.18	0.00	0.43	0.00	0.01	1.57	1.24	211.20	0.01	0.62	1.24	211.29	0.01	-6.21	1.26	209.80
ABT 301	NaCl 3M	TRUE	gompertz	0.00	-3.08	0.94	201.91	0.00	16.40	0.09	0.00	-15.86	0.88	200.93	0.01	43.04	0.85	197.13	0.00	2.65	0.86	200.63
ABT 301	YPD_37	TRUE	logistic	0.35	49.52	5.24	221.80	0.06	0.99	0.11	0.24	47.29	5.37	218.96	0.26	47.61	5.41	217.91	0.16	43.00	5.30	222.89
ABT 301 bis	Gly_3%_1M_NaCl	TRUE	logistic	0.00	12.51	0.07	4.88	0.00	12.17	0.01	0.00	4.95	0.08	5.05	0.00	12.25	0.08	5.02	0.00	2.61	0.08	5.05
ABT 301 bis	Gly_3%_2M_NaCl	TRUE	logistic	0.00	20.90	0.07	4.68	0.00	5.83	0.01	0.00	19.78	0.08	4.87	0.00	19.31	0.08	4.88	0.00	8.72	0.08	4.89
ABT 301 bis	Gly_3%_CuSO4	TRUE	logistic	0.05	51.79	2.17	132.81	0.01	4.65	0.10	0.04	50.33	2.15	137.68	0.05	56.63	2.16	138.05	0.03	42.40	2.16	136.51
ABT 301 bis	Gly_3%	TRUE	logistic	0.06	31.08	2.30	71.50	0.00	1.45	0.06	0.07	32.05	2.30	71.26	0.06	29.47	2.23	72.58	0.05	25.00	2.29	71.21
ABT 301 bis	Gly_4%	TRUE	richards	0.06	29.78	1.62	63.57	0.01	2.36	0.03	0.05	27.85	1.68	64.48	0.05	26.14	1.67	64.33	0.04	18.79	1.66	63.30
ABT 301 bis	Gly_3.5%_glu	TRUE	gompertz.exp	1.16	37.03	7.09	220.74	0.15	0.43	0.43	0.41	29.43	8.16	238.17	0.57	33.01	8.07	230.86	0.26	21.82	8.42	242.35
ABT 301 bis	Gly_10%_glu	TRUE	logistic	0.80	22.09	10.41	444.79	0.46	2.21	0.54	0.35	11.15	11.61	456.69	0.38	18.47	11.56	460.09	0.33	13.87	10.78	449.60
ABT 301 bis	Gly_20%_glu	TRUE	logistic	0.74	25.14	10.53	497.05	0.46	3.49	0.62	0.49	22.46	12.56	521.78	0.49	22.47	12.51	517.04	0.34	17.28	11.46	512.61
ABT 301 bis	KCl 1_75M	TRUE	gompertz.exp	0.06	20.30	2.58	264.33	0.00	2.19	0.16	0.03	1.80	2.90	264.14	0.04	10.84	2.79	267.43	0.03	-1.28	3.01	258.81
ABT 301 bis	KCl 2_5M	TRUE	gompertz	0.01	0.67	1.85	259.48	0.00	9.15	0.10	0.02	-0.89	1.89	260.14	0.02	1.22	1.88	260.90	0.01	-3.75	1.90	258.78
ABT 301 bis	KCl 3M	TRUE	gompertz	0.01	13.98	1.24	200.14	0.00	5.08	0.03	0.01	1.99	1.25	200.71	0.01	-2.27	1.23	200.08	0.01	0.27	1.28	200.26
ABT 301 bis	NaCl 1.75M	TRUE	logistic	0.15	83.82	6.34	413.70	0.02	2.58	0.23	0.12	76.76	6.28	433.08	0.11	76.35	6.33	422.89	0.06	52.87	6.51	396.83
ABT 301 bis	NaCl 2.5M	TRUE	gompertz	0.02	4.30	1.27	221.12	0.00	4.63	0.03	0.02	-3.64	1.30	220.11	0.02	0.41	1.29	222.20	0.01	-8.09	1.31	218.23
ABT 301 bis	NaCl 3M	TRUE	gompertz	0.00	6.83	1.15	204.13	0.00	15.75	0.17	0.01	156.98	0.90	197.87	0.01	151.83	0.90	198.05	0.00	27.15	0.92	202.84
ABT 301 bis	YPD_37	TRUE	logistic	0.33	49.61	5.14	216.62	0.06	0.93	0.10	0.23	47.31	5.24	213.67	0.25	47.70	5.28	212.22	0.16	43.19	5.22	217.58
ABT 601	Control	TRUE	richards	0.23	37.40	5.57	140.31	0.04	2.36	0.20	0.14	31.24	5.50	141.09	0.14	30.99	5.50	141.14	0.14	30.30	5.59	139.52
ABT 601	Gly_3%_1M_NaCl	TRUE	richards	0.00	11.56	0.06	4.29	0.00	18.07	0.00	0.00	2.09	0.06	4.31	0.00	8.68	0.07	4.29	0.00	-0.50	0.06	4.31
ABT 601	Gly_3%_2M_NaCl	TRUE	logistic	0.00	9.00	0.06	4.14	0.00	5.63	0.00	0.00	12.99	0.07	4.10	0.00	10.28	0.07	4.18	0.00	0.89	0.06	4.21
ABT 601	Gly_3%_CuSO4	TRUE	logistic	0.02	23.70	0.61	58.66	0.00	3.23	0.02	0.03	33.74	0.66	59.59	0.02	28.92	0.64	59.50	0.01	8.99	0.62	59.74
ABT 601	Gly_3%	TRUE	richards	0.08	35.11	2.17	71.89	0.01	2.05	0.04	0.09	37.01	2.22	70.61	0.07	33.61	2.18	71.95	0.05	25.30	2.26	70.68
ABT 601	Gly_4%	TRUE	logistic	0.05	25.87	1.52	59.24	0.00	1.65	0.04	0.04	22.64	1.63	61.27	0.04	23.44	1.60	61.26	0.04	18.76	1.53	59.75
ABT 601	Gly_3.5%_glu	TRUE	logistic	1.29	37.23	6.28	187.76	0.47	0.78	0.20	0.84	36.27	6.79	176.27	0.55	33.77	6.73	185.69	0.22	21.28	6.31	199.98
ABT 601	Gly_10%_glu	TRUE	logistic	0.35	19.92	8.79	348.81	0.03	1.32	0.21	0.27	15.25	9.02	350.04	0.28	16.24	9.04	347.96	0.22	11.35	8.94	347.56
ABT 601	Gly_20%_glu	TRUE	logistic	0.42	30.30	8.79	344.95	0.07	2.37	0.35	0.31	25.81	9.85	357.63	0.30	25.24	9.74	361.28	0.24	21.15	8.97	349.81
ABT 601	KCl 1_75M	TRUE	gompertz.exp	0.06	22.82	2.32	233.50	0.00	0.87	0.06	0.03	2.00	2.56	233.48	0.04	13.07	2.43	237.09	0.03	-0.59	2.64	229.38
ABT 601	KCl 2_5M	TRUE	gompertz	0.03	13.70	1.96	317.80	0.00	5.93	0.06	0.02	0.77	2.00	318.68	0.02	7.61	2.01	319.50	0.02	-0.72	2.02	317.81
ABT 601	KCl 3M	TRUE	gompertz	0.01	16.67	1.02	140.92	0.00	11.96	0.08	0.01	3.17	1.00	141.78	0.01	14.55	0.95	141.82	0.01	3.17	1.00	141.78
ABT 601	NaCl 1.75M	TRUE	logistic	0.15	81.97	6.19	411.10	0.03	5.17	0.43	0.13	77.87	6.04	448.61	0.10	70.01	6.10	436.23	0.05	47.48	6.31	407.83
ABT 601	NaCl 2.5M	TRUE	gompertz	0.01	2.21	2.18	255.58	0.00	13.34	0.24	0.02	60.24	1.96	256.67	0.01	-0.50	1.97	257.41	0.01	-3.49	2.05	256.22
ABT 601	NaCl 3M	TRUE	logistic	0.00	43.66	0.84	168.82	0.00	10.87	0.05	0.00	57.11	0.77	168.88	0.00	113.88	0.76	168.74	0.00	50.71	0.78	168.86
ABT 601	YPD_37	TRUE	logistic	0.35	49.72	4.94	210.25	0.06	0.89	0.09	0.23	47.06	4.98	207.55	0.26	47.94	5.04	206.15	0.15	42.96	5.13	210.60
ABT 601 bis	Gly_3%_1M_NaCl	TRUE	logistic	0.00	10.63	0.06	4.04	0.00	5.17	0.00	0.00	14.08	0.07	4.01	0.00	11.78	0.07	4.07	0.00	2.31	0.06	4.11
ABT 601 bis	Gly_3%_2M_NaCl	TRUE	logistic	0.00	10.03	0.06	4.03	0.00	7.52	0.00	0.00	3.53	0.06	4.11	0.00	10.78	0.07	4.09	0.00	1.77	0.06	4.11
ABT 601 bis	Gly_3%_CuSO4	TRUE	richards	0.03	36.52	0.62	60.45	0.01	5.82	0.01	0.03	32.66	0.67	60.51	0.02	28.00	0.66	60.48	0.01	9.16	0.64	60.82
ABT 601 bis	Gly_3%	TRUE	richards	0.07	35.06	2.10	68.36	0.01	2.22	0.04	0.08	35.79	2.13	67.77	0.06	32.76	2.10	68.73	0.05	25.64	2.17	67.24
ABT 601 bis	Gly_4%	TRUE	logistic	0.05	25.55	1.45	56.15	0.00	1.74	0.04	0.04	23.77	1.55	57.93	0.04	23.70	1.53	57.99	0.03	18.64	1.45	56.57
ABT 601 bis	Gly_3.5%_glu	TRUE	logistic	1.05	36.63	6.95	185.32	0.28	0.95	0.24	0.77	35.70	7.45	196.97	0.55	33.15	7.54	205.05	0.24	21.65	6.95	217.74

ABT 601 bis	Gly_10%_glu	TRUE	logistic	0.34	20.55	8.36	327.99	0.03	1.34	0.20	0.26	16.16	8.59	329.73	0.28	17.17	8.61	328.01	0.21	12.06	8.50	327.07
ABT 601 bis	Gly_20%_glu	TRUE	logistic	0.39	30.19	8.31	325.11	0.07	2.56	0.35	0.31	27.31	9.39	337.48	0.29	26.04	9.27	341.67	0.22	21.28	8.47	330.26
ABT 601 bis	KCl 1_75M	TRUE	gompertz	0.05	16.20	2.35	225.97	0.00	1.77	0.03	0.03	1.37	2.44	224.27	0.03	8.40	2.36	226.47	0.02	-2.28	2.54	219.02
ABT 601 bis	KCl 2_5M	TRUE	gompertz	0.02	13.81	2.29	336.45	0.00	5.53	0.06	0.03	18.35	2.35	334.44	0.02	11.80	2.36	335.31	0.02	0.83	2.40	337.45
ABT 601 bis	KCl 3M	TRUE	gompertz	0.01	20.30	0.89	132.00	0.00	8.36	0.04	0.01	5.95	0.86	132.87	0.01	18.06	0.86	132.31	0.01	2.40	0.88	133.30
ABT 601 bis	NaCl 1.75M	TRUE	logistic	0.25	88.89	6.37	447.64	0.13	4.61	0.50	0.11	74.32	6.34	457.60	0.12	71.60	6.30	464.27	0.06	49.35	6.66	427.12
ABT 601 bis	NaCl 2.5M	TRUE	gompertz	0.02	9.46	1.36	225.71	0.00	2.15	0.02	0.01	2.28	1.36	225.99	0.01	2.28	1.36	225.99	0.01	-4.40	1.38	225.36
ABT 601 bis	NaCl 3M	TRUE	logistic	0.00	26.46	0.74	184.77	0.00	10.09	0.02	0.00	31.73	0.74	185.27	0.00	13.89	0.73	185.34	0.00	17.29	0.74	184.49
ABT 601 bis	YPD_37	TRUE	gompertz	0.28	48.43	5.03	204.72	0.03	0.75	0.12	0.21	46.86	5.02	205.67	0.25	47.83	4.98	205.61	0.15	43.07	5.15	207.30
ATCC 42981	Control	TRUE	richards	0.79	25.58	11.42	452.04	0.02	0.19	0.07	0.89	25.94	11.38	456.57	0.80	25.38	11.38	454.53	0.64	24.12	11.51	457.02
ATCC 42981	Gly_3%_1M_NaCl	TRUE	logistic	0.06	22.05	2.73	138.69	0.00	2.00	0.08	0.05	17.85	2.67	138.32	0.06	19.95	2.73	137.85	0.05	16.44	2.72	138.56
ATCC 42981	Gly_3%_2M_NaCl	TRUE	logistic	0.08	21.66	3.66	191.10	0.01	1.54	0.08	0.09	21.14	3.73	195.19	0.08	19.94	3.68	191.74	0.07	15.46	3.67	191.04
ATCC 42981	Gly_3%_CuSO4	TRUE	logistic	0.71	44.63	14.17	1149.39	0.13	1.78	0.25	0.60	43.01	15.21	1165.65	0.52	41.38	15.28	1165.68	0.28	28.53	14.45	1167.93
ATCC 42981	Gly_3%	TRUE	logistic	0.10	29.32	3.27	113.47	0.01	1.17	0.06	0.11	31.83	3.32	113.72	0.10	28.64	3.28	115.08	0.07	22.76	3.31	113.07
ATCC 42981	Gly_4%	TRUE	logistic	0.05	19.42	2.18	84.90	0.01	2.88	0.11	0.05	20.24	2.10	86.05	0.06	22.98	2.17	87.44	0.04	14.63	2.15	85.26
ATCC 42981	Gly_3.5%_glu	TRUE	logistic	0.56	26.69	15.23	411.11	0.09	2.64	0.80	0.60	28.29	14.81	427.83	0.68	30.52	14.86	429.20	0.40	20.53	15.20	426.66
ATCC 42981	Gly_10%_glu	TRUE	logistic	1.20	20.27	14.13	635.30	0.61	1.70	0.43	0.79	18.34	15.67	631.09	0.64	17.23	15.61	631.19	0.46	12.03	14.47	634.66
ATCC 42981	Gly_20%_glu	TRUE	gompertz	0.59	19.66	16.99	742.95	0.05	0.97	0.34	0.66	20.21	17.42	739.26	0.59	19.51	17.38	749.55	0.39	13.82	17.03	749.06
ATCC 42981	KCl 1_75M	TRUE	gompertz.exp	0.10	22.22	2.87	297.41	0.00	0.11	0.01	0.04	0.50	3.12	297.41	0.06	10.69	2.97	302.47	0.03	-2.07	3.21	292.36
ATCC 42981	KCl 2_5M	TRUE	gompertz	0.04	17.85	2.51	398.74	0.00	2.11	0.03	0.02	11.44	2.49	400.24	0.02	8.95	2.50	400.38	0.02	1.14	2.55	401.69
ATCC 42981	KCl 3M	TRUE	gompertz	0.03	17.07	1.31	255.24	0.01	3.92	0.03	0.01	9.99	1.41	253.93	0.02	7.41	1.41	254.77	0.01	-3.50	1.42	248.99
ATCC 42981	NaCl 1.75M	TRUE	gompertz.exp	NA	NA	NA	NA	NA	NA	NA	0.16	78.47	6.69	499.28	0.14	79.11	6.37	482.33	0.06	51.17	6.82	437.97
ATCC 42981	NaCl 2.5M	TRUE	richards	0.04	30.79	1.97	316.82	0.00	2.71	0.01	0.03	14.18	1.98	317.45	0.02	10.61	1.98	318.59	0.02	0.36	2.01	322.12
ATCC 42981	NaCl 3M	TRUE	gompertz	0.00	6.52	1.13	273.37	0.00	13.46	0.06	0.00	-9.09	1.12	272.67	0.01	47.12	1.10	275.38	0.00	6.11	1.11	273.24
ATCC 42981	YPD_37	TRUE	richards	1.16	47.47	15.50	740.51	1.84	14.05	0.23	0.87	44.44	16.25	703.45	0.79	43.64	15.79	722.21	0.43	35.65	15.91	740.25
ATCC 42981 bis	Gly_3%_1M_NaCl	TRUE	logistic	0.06	22.17	2.55	136.68	0.01	1.84	0.07	0.05	17.20	2.53	136.90	0.05	18.97	2.54	137.58	0.05	16.08	2.52	137.06
ATCC 42981 bis	Gly_3%_2M_NaCl	TRUE	logistic	0.09	20.65	3.62	195.27	0.01	2.00	0.10	0.08	18.06	3.60	195.38	0.08	17.84	3.63	194.86	0.07	15.43	3.57	195.80
ATCC 42981 bis	Gly_3%_CuSO4	TRUE	logistic	0.70	44.61	14.14	1145.09	0.11	1.50	0.21	0.59	43.06	14.97	1159.58	0.51	41.45	15.02	1159.91	0.28	28.55	14.39	1161.43
ATCC 42981 bis	Gly_3%	TRUE	logistic	0.10	29.72	3.18	109.32	0.01	1.21	0.06	0.11	32.24	3.23	109.45	0.09	28.94	3.19	110.80	0.07	23.13	3.22	108.83
ATCC 42981 bis	Gly_4%	TRUE	logistic	0.05	19.01	2.12	82.05	0.01	3.04	0.12	0.05	20.04	2.03	83.05	0.05	23.51	2.10	84.44	0.04	14.47	2.08	82.32
ATCC 42981 bis	Gly_3.5%_glu	TRUE	richards	0.73	28.97	13.89	429.02	0.04	0.77	0.15	0.60	26.41	13.91	432.54	0.60	26.42	13.91	432.50	0.41	18.71	14.52	438.57
ATCC 42981 bis	Gly_10%_glu	TRUE	logistic	1.19	20.44	13.65	612.67	0.65	1.74	0.43	0.61	17.28	15.14	609.17	0.62	17.40	15.13	608.90	0.44	12.16	14.00	611.95
ATCC 42981 bis	Gly_20%_glu	TRUE	gompertz	0.56	19.33	16.44	721.13	0.05	1.02	0.35	0.67	20.32	16.82	715.26	0.59	19.55	16.80	726.46	0.37	13.47	16.50	726.84
ATCC 42981 bis	KCl 1_75M	TRUE	gompertz.exp	0.08	21.62	2.99	301.38	0.00	0.87	0.08	0.04	2.01	3.27	301.33	0.05	12.08	3.12	305.60	0.03	-0.93	3.38	295.37
ATCC 42981 bis	KCl 2_5M	TRUE	gompertz	0.03	18.87	2.56	394.62	0.00	2.13	0.03	0.03	4.45	2.56	395.49	0.03	5.81	2.55	396.34	0.02	1.81	2.57	399.32
ATCC 42981 bis	KCl 3M	TRUE	gompertz	0.02	18.30	1.21	229.50	0.00	3.39	0.02	0.01	10.83	1.29	228.45	0.01	8.11	1.26	229.07	0.01	-1.47	1.24	226.65
ATCC 42981 bis	NaCl 1.75M	TRUE	logistic	0.36	90.96	6.60	474.83	0.34	4.84	0.30	0.16	78.84	6.61	501.45	0.14	78.97	6.60	484.86	0.06	51.73	6.99	442.42
ATCC 42981 bis	NaCl 2.5M	TRUE	gompertz	0.05	18.85	2.09	353.48	0.00	0.88	0.01	0.02	11.41	2.11	353.80	0.02	8.81	2.11	354.43	0.02	-0.99	2.16	354.63
ATCC 42981 bis	NaCl 3M	TRUE	gompertz	0.00	21.32	1.06	277.14	0.00	11.00	0.04	0.00	4.88	1.06	278.05	0.01	48.74	1.05	278.99	0.00	9.54	1.05	278.29
ATCC 42981 bis	YPD_37	TRUE	logistic	0.97	44.83	15.76	740.55	0.12	1.04	0.29	0.99	45.15	16.43	731.99	0.89	44.17	16.02	743.98	0.44	35.05	15.99	759.16
CBS 4837	Control	TRUE	richards	0.77	26.12	11.02	435.20	0.02	0.25	0.07	0.64	24.77	10.99	434.37	0.68	25.14	10.99	434.60	0.59	24.08	11.12	439.58
CBS 4837	Gly_3%_1M_NaCl	TRUE	gompertz	0.08	12.18	2.46	161.23	0.02	3.42	0.09	0.06	7.29	2.57	163.30	0.08	11.89	2.61	162.48	0.05	4.06	2.53	163.87
CBS 4837	Gly_3%_2M_NaCl	TRUE	gompertz	0.11	18.24	3.43	203.85	0.01	1.83	0.11	0.14	20.45	3.61	205.68	0.14	20.35	3.60	205.61	0.07	12.19	3.47	206.01
CBS 4837	Gly_3%_CuSO4	TRUE	logistic	0.47	36.60	13.53	1148.82	0.14	4.77	0.49	0.40	39.09	15.84	1157.61	0.32	30.28	15.40	1152.87	0.26	23.47	13.98	1161.87
CBS 4837	Gly_3%	TRUE	gompertz	0.06	25.80	2.81	82.24	0.00	0.94	0.09	0.05	18.76	2.51	83.76	0.05	21.55	2.47	83.84	0.05	22.76	2.57	82.61
CBS 4837	Gly_4%	TRUE	logistic	0.07	28.08	1.98	74.04	0.00	1.12	0.04	0.05	23.51	2.05	76.44	0.06	24.66	2.04	75.87	0.05	20.82	2.01	74.18
CBS 4837	Gly_3.5%_glu	TRUE	richards	0.79	28.39	12.66	419.18	0.07	1.16	0.15	0.72	27.26	13.11	421.24	0.60	24.72	13.07	426.14	0.39	16.38	12.98	434.39
CBS 4837	Gly_10%_glu	TRUE	gompertz	0.84	19.73	11.74	513.65	0.28	1.46	0.44	0.45	16.62	12.63	518.57	0.51	17.58	12.64	522.26	0.35	12.50	12.12	512.12
CBS 4837	Gly_20%_glu	TRUE	logistic	0.86	20.67	12.12	633.38	0.46	4.18	0.80	0.80	20.02	14.95	659.96	0.70	19.06	14.75	661.76	0.37	10.08	13.64	667.54
CBS 4837	KCl 1_75M	TRUE	logistic	0.08	18.48	2.60	279.92	0.00	0.69	0.02	0.04	-0.92	2.65	281.01	0.05	8.57	2.67	284.78	0.03	-5.94	2.78	272.15
CBS 4837	KCl 2_5M	TRUE	logistic	0.02	20.46	2.75	380.16	0.00	8.03	0.09	0.02	1.37	2.71	385.73	0.02	9.89	2.72	385.76	0.02	-1.46	2.78	385.35
CBS 4837	KCl 3M	TRUE	gompertz	0.02	4.69	1.29	250.80	0.00	3.58	0.02	0.02	-0.23	1.34	250.78	0.02	0.07	1.34	250.97	0.01	-8.60	1.36	244.59
CBS 4837	NaCl 1.75M	TRUE	gompertz.exp	NA	NA	NA	NA	NA	NA	NA	0.17	84.30	6.95	509.40	0.15	80.64	6.54	492.85	0.07	52.99	7.02	444.66
CBS 4837	NaCl 2.5M	TRUE	richards	0.03	3.38	1.83	308.38	0.00	2.33	0.03	0.02	-0.07	1.83	308.70	0.02	-0.64	1.84	308.71	0.02	-7.18	1.86	304.71
CBS 4837	NaCl 3M	TRUE	gompertz	0.00	-24.22	1.29	305.80	0.00	16.83	0.09	0.01	-6.44	1.18	305.14	0.01	-6.17	1.18	304.61	0.01	-12.56	1.20	304.52

CBS 4837	YPD_37	TRUE	logistic	0.46	49.22	8.23	340.43	0.06	0.83	0.16	0.33	46.96	8.55	341.31	0.37	47.69	8.55	342.46	0.25	43.41	8.42	342.28
CBS 4837 bis	Gly_3%_1M_NaCl	TRUE	logistic	0.09	14.77	2.26	155.44	0.02	3.07	0.06	0.07	11.23	2.40	155.84	0.08	12.01	2.42	155.62	0.05	4.80	2.36	157.27
CBS 4837 bis	Gly_3%_2M_NaCl	TRUE	gompertz	0.11	18.30	3.31	198.41	0.01	1.79	0.10	0.10	17.63	3.39	199.49	0.13	19.56	3.39	199.01	0.07	11.89	3.36	200.74
CBS 4837 bis	Gly_3%_CuSO4	TRUE	richards	0.48	38.04	13.23	1137.47	0.23	8.04	0.54	0.40	38.98	15.66	1142.10	0.32	30.02	15.22	1136.57	0.25	23.38	13.79	1144.02
CBS 4837 bis	Gly_3%	TRUE	gompertz	0.06	26.44	2.73	77.86	0.00	0.84	0.08	0.05	21.58	2.41	79.25	0.05	23.11	2.38	79.29	0.05	23.48	2.48	78.12
CBS 4837 bis	Gly_4%	TRUE	logistic	0.06	27.97	1.91	70.97	0.00	1.18	0.04	0.05	24.17	1.98	73.20	0.05	24.94	1.96	72.65	0.04	20.85	1.94	71.03
CBS 4837 bis	Gly_3.5%_glu	TRUE	richards	0.70	27.34	12.09	402.50	0.06	1.24	0.15	0.56	24.41	12.43	405.54	0.53	23.24	12.41	408.30	0.36	16.05	12.41	414.51
CBS 4837 bis	Gly_10%_glu	TRUE	logistic	0.71	20.31	11.24	487.09	0.22	1.70	0.43	0.42	16.92	12.14	494.56	0.48	17.85	12.16	498.35	0.34	12.83	11.63	487.83
CBS 4837 bis	Gly_20%_glu	TRUE	logistic	0.85	20.62	11.57	608.55	0.43	3.86	0.72	0.80	20.08	14.19	633.55	0.69	19.02	13.98	634.39	0.36	9.56	12.96	641.58
CBS 4837 bis	KCl 1_75M	TRUE	logistic	0.08	20.23	2.68	280.18	0.00	0.40	0.01	0.04	-0.41	2.77	280.29	0.05	6.25	2.71	284.18	0.03	-4.43	2.88	273.05
CBS 4837 bis	KCl 2_5M	TRUE	gompertz	0.03	5.85	2.76	421.95	0.00	3.59	0.05	0.03	-0.21	2.71	424.47	0.03	-0.19	2.71	424.47	0.02	-3.29	2.76	422.51
CBS 4837 bis	KCl 3M	TRUE	gompertz	0.04	15.77	1.45	288.45	0.01	2.32	0.02	0.02	6.78	1.50	288.95	0.02	5.15	1.50	289.14	0.02	-5.48	1.52	280.65
CBS 4837 bis	NaCl 1.75M	TRUE	logistic	0.22	85.58	8.01	532.82	0.03	2.78	0.34	0.17	78.45	7.91	560.83	0.14	76.97	7.99	545.27	0.07	51.96	8.27	507.85
CBS 4837 bis	NaCl 2.5M	TRUE	gompertz	0.03	5.28	1.84	324.20	0.00	2.54	0.02	0.03	0.60	1.87	324.86	0.03	-0.37	1.87	324.57	0.02	-7.97	1.89	319.09
CBS 4837 bis	NaCl 3M	TRUE	gompertz	0.00	-14.87	1.10	297.75	0.00	17.84	0.06	0.01	144.92	1.04	290.78	0.01	103.12	1.04	292.44	0.01	-11.77	1.05	296.07
CBS 4837 bis	YPD_37	TRUE	logistic	0.47	49.25	8.42	348.30	0.06	0.83	0.16	0.35	47.10	8.76	349.86	0.38	47.71	8.76	351.31	0.25	43.40	8.59	350.36
CBS 4838	Control	TRUE	richards	0.83	26.99	11.33	441.17	0.06	0.69	0.20	0.81	26.37	11.32	433.77	0.79	26.29	11.32	434.34	0.61	24.80	11.42	444.58
CBS 4838	Gly_3%_1M_NaCl	TRUE	logistic	0.09	13.48	2.41	164.92	0.01	2.17	0.04	0.09	13.38	2.52	164.22	0.08	11.52	2.52	164.48	0.05	4.39	2.51	166.06
CBS 4838	Gly_3%_2M_NaCl	TRUE	gompertz	0.08	14.91	3.52	198.88	0.01	3.00	0.18	0.17	21.41	3.67	197.66	0.14	20.14	3.64	199.35	0.07	10.60	3.48	200.37
CBS 4838	Gly_3%_CuSO4	TRUE	logistic	0.48	42.29	12.12	981.80	0.22	5.58	0.62	0.32	36.34	13.04	996.56	0.43	45.99	14.43	997.26	0.24	28.42	12.46	999.43
CBS 4838	Gly_3%	TRUE	logistic	0.07	29.63	2.54	82.48	0.00	1.23	0.06	0.06	27.39	2.45	83.35	0.06	27.44	2.47	83.42	0.05	23.67	2.55	82.19
CBS 4838	Gly_4%	TRUE	logistic	0.07	28.80	1.91	69.91	0.01	1.53	0.05	0.05	21.28	2.01	71.99	0.05	24.13	1.99	71.91	0.04	21.56	1.93	70.06
CBS 4838	Gly_3.5%_glu	TRUE	richards	0.78	32.89	12.72	378.30	0.10	1.83	0.28	0.79	32.86	13.33	380.63	0.69	31.36	13.24	382.13	0.38	20.84	13.36	383.83
CBS 4838	Gly_10%_glu	TRUE	logistic	0.82	20.50	13.48	578.47	0.04	0.33	0.10	0.47	17.35	13.68	576.85	0.52	17.65	13.64	575.32	0.39	12.72	13.91	574.15
CBS 4838	Gly_20%_glu	TRUE	logistic	0.87	21.00	11.64	609.12	0.38	3.17	0.62	0.81	20.28	13.57	622.00	0.71	19.51	13.48	628.81	0.36	9.89	12.80	638.39
CBS 4838	KCl 1_75M	TRUE	gompertz.exp	0.08	18.07	2.68	294.84	0.00	1.08	0.11	0.04	-0.69	2.96	294.67	0.05	8.49	2.87	298.26	0.03	-4.78	3.08	287.10
CBS 4838	KCl 2_5M	TRUE	gompertz	0.03	3.89	2.23	379.82	0.00	3.28	0.04	0.03	-0.77	2.30	379.21	0.03	0.22	2.30	380.67	0.02	-6.08	2.32	375.04
CBS 4838	KCl 3M	TRUE	gompertz	0.02	12.83	1.05	205.41	0.00	2.63	0.02	0.01	5.75	1.11	204.51	0.01	4.15	1.11	204.76	0.01	-5.55	1.11	200.44
CBS 4838	NaCl 1.75M	TRUE	gompertz.exp	NA	NA	NA	NA	NA	NA	NA	0.20	85.64	7.00	518.62	0.15	77.59	6.62	494.84	0.07	52.86	7.04	442.34
CBS 4838	NaCl 2.5M	TRUE	richards	0.03	5.79	1.86	310.41	0.00	2.01	0.02	0.02	1.47	1.87	310.42	0.02	0.74	1.87	310.61	0.02	-5.79	1.90	307.80
CBS 4838	NaCl 3M	TRUE	gompertz	0.00	-19.38	1.30	313.54	0.00	19.11	0.10	0.01	-4.08	1.18	319.35	0.01	-4.67	1.18	318.52	0.01	-13.91	1.20	313.22
CBS 4838	YPD_37	TRUE	logistic	0.48	49.04	8.47	353.18	0.08	1.02	0.20	0.48	48.83	8.99	366.01	0.38	47.39	8.95	358.31	0.26	43.10	8.56	355.75
CBS 4838 bis	Gly_3%_1M_NaCl	TRUE	logistic	0.10	14.61	2.43	167.55	0.01	1.82	0.04	0.08	11.82	2.50	167.22	0.09	12.73	2.50	166.69	0.06	4.64	2.52	169.11
CBS 4838 bis	Gly_3%_2M_NaCl	TRUE	gompertz	0.09	17.58	3.37	194.75	0.01	1.79	0.11	0.09	16.61	3.41	196.00	0.12	19.29	3.46	195.78	0.07	12.10	3.36	196.87
CBS 4838 bis	Gly_3%_CuSO4	TRUE	logistic	0.48	42.16	12.23	990.41	0.22	5.66	0.63	0.32	36.09	13.12	1006.12	0.44	46.34	14.56	1006.39	0.24	28.46	12.63	1008.87
CBS 4838 bis	Gly_3%	TRUE	logistic	0.07	30.32	2.45	78.18	0.00	1.25	0.06	0.06	28.02	2.36	79.06	0.06	28.05	2.38	79.09	0.05	24.29	2.46	77.84
CBS 4838 bis	Gly_4%	TRUE	logistic	0.06	28.71	1.84	66.93	0.01	1.59	0.05	0.05	24.20	1.94	69.03	0.05	23.99	1.92	68.80	0.04	21.67	1.87	66.98
CBS 4838 bis	Gly_3.5%_glu	TRUE	richards	0.69	31.20	13.42	394.25	0.09	2.07	0.39	0.61	29.36	13.80	397.36	0.63	29.73	13.85	397.57	0.38	20.56	13.95	396.48
CBS 4838 bis	Gly_10%_glu	TRUE	logistic	0.82	20.83	13.03	558.11	0.04	0.27	0.08	0.46	17.65	13.20	552.98	0.50	17.84	13.15	554.50	0.38	12.93	13.44	554.37
CBS 4838 bis	Gly_20%_glu	TRUE	logistic	0.87	21.02	11.02	581.19	0.38	3.04	0.59	0.81	20.35	12.97	591.32	0.71	19.51	12.81	600.29	0.34	9.43	12.15	611.35
CBS 4838 bis	KCl 1_75M	TRUE	gompertz.exp	0.08	17.20	2.93	308.04	0.00	0.56	0.06	0.04	0.21	3.10	307.99	0.05	5.19	3.02	311.18	0.03	-4.23	3.25	299.21
CBS 4838 bis	KCl 2_5M	TRUE	gompertz	0.04	14.51	2.13	362.68	0.01	3.68	0.05	0.03	8.94	2.26	365.15	0.03	6.58	2.25	365.07	0.02	-1.83	2.21	362.24
CBS 4838 bis	KCl 3M	TRUE	gompertz	0.02	9.07	1.08	213.62	0.00	2.00	0.01	0.01	2.48	1.10	214.09	0.01	1.66	1.10	213.98	0.01	-7.00	1.11	208.79
CBS 4838 bis	NaCl 1.75M	TRUE	logistic	0.28	89.43	8.12	550.30	0.06	2.87	0.44	0.20	82.40	8.04	584.59	0.15	74.71	8.14	563.86	0.08	52.66	8.46	518.54
CBS 4838 bis	NaCl 2.5M	TRUE	gompertz	0.02	5.67	1.76	289.44	0.00	1.96	0.02	0.02	-0.70	1.76	288.94	0.02	-0.54	1.76	289.41	0.02	-5.40	1.79	288.19
CBS 4838 bis	NaCl 3M	TRUE	gompertz	0.00	-23.13	1.01	296.91	0.00	19.56	0.04	0.01	-4.10	1.00	294.52	0.01	-4.62	1.00	294.38	0.01	-13.07	1.00	295.01
CBS 4838 bis	YPD_37	TRUE	logistic	0.50	49.16	8.71	363.63	0.08	0.95	0.18	0.38	47.13	9.15	366.98	0.40	47.51	9.16	368.60	0.26	43.13	8.81	365.94
CBS 732	Control	TRUE	richards	0.50	38.52	7.22	236.14	0.13	3.79	0.10	0.47	38.12	7.44	240.30	0.43	36.55	7.35	238.85	0.26	27.20	7.36	240.42
CBS 732	Gly_3%_1M_NaCl	TRUE	logistic	0.00	8.33	0.06	3.68	0.00	6.23	0.00	0.00	20.52	0.06	3.78	0.00	22.75	0.06	3.77	0.00	3.92	0.06	3.74
CBS 732	Gly_3%_2M_NaCl	TRUE	logistic	0.00	8.34	0.06	3.76	0.00	5.15	0.00	0.00	19.80	0.06	3.85	0.00	18.83	0.06	3.85	0.00	3.62	0.06	3.82
CBS 732	Gly_3%_CuSO4	TRUE	richards	0.06	44.18	2.24	168.45	0.01	4.63	0.05	0.05	40.06	2.39	167.96	0.05	38.46	2.30	168.16	0.03	29.46	2.27	167.64
CBS 732	Gly_3%	TRUE	logistic	0.10	29.39	3.17	109.65	0.00	0.93	0.05	0.10	29.44	3.17	110.11	0.09	27.00	3.15	111.02	0.07	22.76	3.23	109.09
CBS 732	Gly_4%	TRUE	logistic	0.05	22.53	1.91	76.16	0.00	1.32	0.04	0.04	18.73	1.98	77.76	0.04	17.28	1.97	77.67	0.04	15.39	1.92	76.33
CBS 732	Gly_3.5%_glu	TRUE	gompertz	1.29	36.84	11.09	275.42	0.25	0.85	0.41	0.71	33.42	11.34	312.61	0.80	34.39	11.33	308.05	0.37	24.19	11.59	317.87
CBS 732	Gly_10%_glu	TRUE	gompertz	0.66	19.05	12.47	519.09	0.03	0.37	0.13	0.47	17.29	12.47	516.94	0.49	17.49	12.45	516.98	0.34	12.51	12.72	516.43

CBS 732	Gly_20%_glu	TRUE	logistic	0.53	28.29	14.65	555.73	0.08	2.84	0.69	0.42	24.25	14.64	565.60	0.54	28.81	14.85	558.41	0.35	20.34	14.94	562.48
CBS 732	KCl 1_75M	TRUE	richards	0.06	20.91	2.84	273.14	0.00	0.65	0.01	0.05	16.83	2.84	273.11	0.04	13.39	2.85	271.76	0.03	-0.08	3.09	260.12
CBS 732	KCl 2_5M	TRUE	gompertz	0.02	-3.84	1.97	321.05	0.00	7.43	0.08	0.03	-0.88	2.02	323.78	0.03	-0.52	2.03	324.22	0.02	-8.68	2.07	317.70
CBS 732	KCl 3M	TRUE	gompertz	0.01	19.50	0.94	167.94	0.00	6.65	0.03	0.01	3.06	0.97	168.12	0.01	-2.32	0.99	168.93	0.01	0.72	0.97	167.32
CBS 732	NaCl 1.75M	TRUE	logistic	0.23	86.54	6.20	443.44	0.09	3.91	0.31	0.13	79.72	6.08	472.95	0.12	76.20	6.12	460.63	0.06	48.76	6.50	425.37
CBS 732	NaCl 2.5M	TRUE	gompertz	0.02	7.45	1.38	225.60	0.00	2.73	0.02	0.01	3.24	1.39	225.89	0.01	2.14	1.39	226.13	0.01	-4.87	1.41	224.79
CBS 732	NaCl 3M	TRUE	gompertz	0.00	36.78	0.88	207.45	0.00	7.65	0.03	0.00	51.37	0.86	205.04	0.00	47.15	0.86	205.70	0.00	27.28	0.86	207.81
CBS 732	YPD_37	TRUE	logistic	0.77	48.15	12.81	552.48	0.06	0.49	0.12	0.67	47.08	13.04	547.20	0.64	46.74	13.04	547.20	0.38	41.50	13.52	555.01
CBS 732 bis	Gly_3%_1M_NaCl	TRUE	richards	0.00	11.39	0.06	3.88	0.00	5.40	0.00	0.00	6.48	0.06	3.90	0.00	11.49	0.06	3.92	0.00	3.34	0.06	3.90
CBS 732 bis	Gly_3%_2M_NaCl	TRUE	richards	0.00	24.12	0.06	3.68	0.00	17.41	0.00	0.00	18.77	0.07	3.70	0.00	24.46	0.06	3.71	0.00	5.62	0.06	3.69
CBS 732 bis	Gly_3%_CuSO4	TRUE	richards	0.06	44.65	2.27	172.42	0.01	5.18	0.05	0.05	40.21	2.45	172.16	0.05	38.66	2.37	172.43	0.04	29.44	2.30	171.82
CBS 732 bis	Gly_3%	TRUE	logistic	0.10	29.83	3.09	105.43	0.01	1.05	0.05	0.10	30.12	3.08	106.01	0.08	27.42	3.06	106.93	0.07	23.21	3.14	104.89
CBS 732 bis	Gly_4%	TRUE	logistic	0.05	22.13	1.84	73.11	0.00	1.44	0.05	0.04	18.49	1.91	74.61	0.04	17.24	1.89	74.52	0.04	15.21	1.84	73.20
CBS 732 bis	Gly_3.5%_glu	TRUE	gompertz.exp	1.32	37.71	9.75	275.91	0.20	0.77	1.25	1.00	35.88	10.71	281.18	0.84	35.16	10.76	289.93	0.35	24.25	10.97	303.94
CBS 732 bis	Gly_10%_glu	TRUE	gompertz	0.63	19.36	11.99	495.20	0.03	0.39	0.13	0.44	17.53	12.00	493.39	0.46	17.74	11.98	493.48	0.33	12.79	12.23	492.44
CBS 732 bis	Gly_20%_glu	TRUE	logistic	0.48	27.90	13.93	525.70	0.08	2.94	0.68	0.40	24.77	13.85	535.42	0.51	29.24	13.99	528.23	0.33	20.30	14.21	532.24
CBS 732 bis	KCl 1_75M	TRUE	gompertz	0.06	18.34	2.94	274.92	0.00	1.55	0.04	0.03	3.15	3.04	273.04	0.04	11.83	2.95	276.15	0.03	-0.38	3.18	266.55
CBS 732 bis	KCl 2_5M	TRUE	gompertz	0.02	8.84	2.03	307.20	0.00	7.84	0.08	0.02	7.10	2.07	309.30	0.02	6.85	2.07	309.29	0.02	-1.00	2.12	306.65
CBS 732 bis	KCl 3M	TRUE	gompertz	0.01	17.84	0.97	175.32	0.00	5.72	0.03	0.01	-6.46	1.05	173.94	0.01	-5.97	1.05	174.59	0.01	-0.41	1.05	173.24
CBS 732 bis	NaCl 1.75M	TRUE	logistic	0.21	87.51	6.40	440.09	0.06	3.68	0.40	0.15	79.70	6.29	474.66	0.12	77.19	6.37	459.56	0.06	49.64	6.66	424.26
CBS 732 bis	NaCl 2.5M	TRUE	gompertz	0.02	7.01	1.38	223.75	0.00	3.18	0.02	0.01	4.32	1.38	224.98	0.01	2.41	1.38	224.78	0.01	-4.77	1.41	223.20
CBS 732 bis	NaCl 3M	TRUE	gompertz	0.00	45.13	0.82	199.63	0.00	10.72	0.03	0.00	50.53	0.85	200.57	0.01	60.53	0.85	201.22	0.00	37.54	0.85	200.26
CBS 732 bis	YPD_37	TRUE	logistic	0.69	47.91	13.12	554.80	0.09	0.88	0.25	0.53	45.76	13.34	552.60	0.58	46.39	13.25	549.56	0.38	42.00	13.64	558.39
OUT 7136 bis	Control	TRUE	gompertz.exp	0.62	37.34	5.70	167.75	0.16	1.63	1.48	0.40	34.00	6.53	181.36	0.44	34.58	6.57	179.56	0.26	28.00	6.51	183.39
OUT 7136	Gly_3%_1M_NaCl	TRUE	logistic	0.05	24.93	1.34	75.64	0.00	1.07	0.02	0.04	21.59	1.38	76.08	0.04	21.71	1.39	76.24	0.03	17.04	1.36	76.16
OUT 7136	Gly_3%_2M_NaCl	TRUE	logistic	0.06	26.92	2.28	116.26	0.00	1.13	0.04	0.05	22.56	2.30	116.63	0.05	21.43	2.31	116.94	0.05	20.87	2.27	116.58
OUT 7136	Gly_3%_CuSO4	TRUE	logistic	0.12	82.51	8.15	200.94	0.01	2.85	1.24	0.11	80.25	6.29	202.92	0.16	91.80	6.28	202.32	0.11	80.07	6.28	203.04
OUT 7136	Gly_3%	TRUE	logistic	0.06	28.46	1.91	70.20	0.00	1.30	0.04	0.05	26.61	1.99	73.37	0.05	22.84	1.96	72.35	0.04	21.22	1.94	70.46
OUT 7136	Gly_4%	TRUE	richards	0.07	32.28	1.46	57.80	0.01	2.92	0.03	0.05	28.82	1.54	59.30	0.05	27.42	1.54	59.53	0.04	19.86	1.48	58.28
OUT 7136	Gly_3.5%_glu	TRUE	logistic	0.35	33.37	5.86	148.74	0.09	2.58	0.35	0.29	31.66	5.88	165.58	0.31	32.77	5.88	165.24	0.17	22.78	6.02	165.80
OUT 7136	Gly_10%_glu	TRUE	gompertz.exp	NA	NA	NA	NA	NA	NA	NA	0.43	19.53	11.10	419.70	0.55	24.36	11.49	417.78	0.30	13.12	9.81	414.61
OUT 7136	Gly_20%_glu	TRUE	logistic	0.67	26.09	9.16	424.56	0.47	3.56	0.56	0.37	22.51	10.95	441.14	0.40	22.96	11.07	443.60	0.30	18.44	9.94	435.53
OUT 7136	KCl 1_75M	TRUE	gompertz.exp	0.07	21.13	2.24	237.06	0.00	0.67	0.05	0.03	0.25	2.50	237.03	0.04	9.46	2.39	240.55	0.03	-2.61	2.58	232.65
OUT 7136	KCl 2_5M	TRUE	logistic	0.08	21.26	1.53	270.43	0.96	31.60	0.05	0.02	9.53	1.71	268.58	0.02	7.37	1.72	268.58	0.02	-2.93	1.70	265.64
OUT 7136	KCl 3M	TRUE	gompertz	0.00	-8.10	0.61	104.86	0.00	15.94	0.05	0.01	0.44	0.63	107.14	0.01	-0.57	0.63	107.05	0.00	-10.51	0.65	103.94
OUT 7136	NaCl 1.75M	TRUE	richards	0.18	93.73	4.98	348.89	0.04	7.82	0.11	0.11	78.96	4.98	353.36	0.10	79.16	5.03	343.17	0.05	54.03	5.22	316.79
OUT 7136	NaCl 2.5M	TRUE	logistic	0.04	16.75	1.62	283.81	0.01	3.85	0.03	0.02	0.06	1.69	285.02	0.02	5.32	1.68	285.48	0.02	-3.69	1.69	283.20
OUT 7136	NaCl 3M	TRUE	gompertz	0.00	33.31	1.05	247.33	0.00	11.29	0.05	0.01	57.16	1.02	241.52	0.01	57.60	1.02	241.11	0.00	29.44	1.04	246.39
OUT 7136	YPD_37	TRUE	logistic	0.22	49.80	5.25	201.28	0.01	0.65	0.10	0.16	46.72	5.21	202.70	0.16	46.56	5.22	202.49	0.14	44.24	5.38	203.13
OUT 7136 bis	Gly_3%_1M_NaCl	TRUE	logistic	0.05	25.74	1.31	72.87	0.00	1.49	0.03	0.04	22.18	1.35	73.24	0.04	22.60	1.36	73.07	0.03	18.24	1.33	73.44
OUT 7136 bis	Gly_3%_2M_NaCl	TRUE	logistic	0.06	28.06	2.08	103.36	0.00	1.31	0.04	0.05	24.81	2.09	103.86	0.05	24.10	2.10	104.30	0.04	22.23	2.06	103.81
OUT 7136 bis	Gly_3%_CuSO4	TRUE	logistic	0.12	82.98	8.80	200.78	0.01	3.71	1.74	0.11	81.09	6.23	204.94	0.16	91.90	6.24	204.02	0.11	80.71	6.21	205.16
OUT 7136 bis	Gly_3%	TRUE	logistic	0.06	29.04	1.82	65.83	0.00	1.32	0.04	0.05	27.67	1.89	68.78	0.05	23.66	1.87	67.82	0.04	21.85	1.85	65.99
OUT 7136 bis	Gly_4%	TRUE	richards	0.06	32.41	1.39	55.10	0.01	3.11	0.03	0.06	30.21	1.48	56.34	0.05	27.74	1.48	56.67	0.03	19.80	1.41	55.52
OUT 7136 bis	Gly_3.5%_glu	TRUE	gompertz.exp	0.62	37.50	5.65	171.64	0.17	1.90	1.73	0.40	33.85	6.54	181.68	0.43	34.50	6.50	180.02	0.20	24.10	6.69	184.01
OUT 7136 bis	Gly_10%_glu	TRUE	gompertz.exp	NA	NA	NA	NA	NA	NA	NA	0.44	21.14	10.79	397.67	0.54	25.03	11.14	396.24	0.28	13.39	9.37	392.64
OUT 7136 bis	Gly_20%_glu	TRUE	logistic	0.65	26.04	8.66	403.57	0.45	3.50	0.53	0.37	22.61	10.29	419.75	0.41	23.26	10.32	423.27	0.29	18.28	9.42	414.19
OUT 7136 bis	KCl 1_75M	TRUE	logistic	0.07	22.08	2.53	255.20	0.01	2.06	0.04	0.03	1.57	2.56	257.82	0.05	10.65	2.61	266.25	0.03	-2.73	2.70	250.27
OUT 7136 bis	KCl 2_5M	TRUE	gompertz	0.03	13.44	1.64	275.55	0.00	3.74	0.03	0.02	7.05	1.71	274.64	0.02	5.89	1.72	275.19	0.02	-2.16	1.73	273.83
OUT 7136 bis	KCl 3M	TRUE	gompertz	0.01	3.50	0.54	105.31	0.00	7.57	0.02	0.01	-0.84	0.58	105.59	0.01	-0.13	0.58	106.11	0.01	-9.15	0.58	102.80
OUT 7136 bis	NaCl 1.75M	TRUE	logistic	0.18	88.55	5.15	355.58	0.04	2.62	0.24	0.12	79.73	5.09	376.62	0.10	78.49	5.16	364.75	0.05	51.83	5.39	335.79
OUT 7136 bis	NaCl 2.5M	TRUE	logistic	0.05	20.29	1.40	246.81	0.09	6.37	0.03	0.02	9.47	1.51	245.90	0.02	6.94	1.50	246.27	0.02	-4.00	1.48	244.45
OUT 7136 bis	NaCl 3M	TRUE	gompertz	0.01	49.11	0.88	223.97	0.00	16.86	0.05	0.01	67.94	0.87	214.94	0.01	68.83	0.87	214.61	0.00	35.79	0.87	223.01
OUT 7136 bis	YPD_37	TRUE	richards	0.25	53.12	5.39	207.96	0.03	2.90	0.15	0.17	45.36	5.48	210.28	0.17	45.59	5.50	210.17	0.15	44.32	5.53	210.17
OUT 7140 bis	Control	TRUE	logistic	0.55	20.78	12.27	430.04	0.03	0.75	0.18	0.54	21.97	12.06	420.93	0.49	20.21	12.06	423.79	0.37	14.76	12.30	431.86

OUT 7140	Gly_3%_1M_NaCl	TRUE	logistic	0.00	5.90	0.07	5.19	0.00	8.26	0.00	0.00	16.08	0.09	5.40	0.00	-1.66	0.09	5.33	0.00	-0.06	0.08	5.28
OUT 7140	Gly_3%_2M_NaCl	TRUE	logistic	0.00	6.32	0.07	5.02	0.00	7.47	0.00	0.00	-0.65	0.07	5.10	0.00	14.83	0.08	5.22	0.00	0.36	0.07	5.10
OUT 7140	Gly_3%_CuSO4	TRUE	gompertz	0.01	-4.59	0.75	80.09	0.00	6.76	0.03	0.02	-1.31	0.77	79.88	0.02	-0.30	0.77	80.18	0.01	-4.49	0.78	79.51
OUT 7140	Gly_3%	TRUE	logistic	0.09	28.97	2.85	99.29	0.00	1.13	0.05	0.07	23.48	2.87	100.49	0.07	23.58	2.85	100.70	0.06	22.27	2.90	98.68
OUT 7140	Gly_4%	TRUE	logistic	0.06	24.36	2.11	82.71	0.00	0.81	0.03	0.05	21.00	2.14	84.23	0.05	19.61	2.14	83.93	0.05	17.07	2.14	82.74
OUT 7140	Gly_3.5%_glu	TRUE	richards	0.72	26.01	13.30	455.80	0.17	3.51	0.43	0.61	23.72	14.14	455.00	0.54	21.43	13.86	463.27	0.39	14.55	13.60	466.05
OUT 7140	Gly_10%_glu	TRUE	gompertz	0.76	17.59	13.30	582.54	0.10	0.86	0.26	0.58	16.04	13.83	589.60	0.60	16.26	13.84	590.28	0.38	10.70	13.64	584.84
OUT 7140	Gly_20%_glu	TRUE	gompertz	1.13	21.10	12.03	634.26	0.47	2.32	0.64	0.69	17.79	13.99	669.62	0.75	18.41	14.03	666.82	0.36	8.13	13.34	678.87
OUT 7140	KCl 1_75M	TRUE	logistic	0.10	25.26	3.07	304.94	0.00	0.79	0.02	0.04	1.44	3.24	304.78	0.06	12.43	3.10	309.64	0.03	-1.29	3.33	299.20
OUT 7140	KCl 2_5M	TRUE	gompertz	0.04	14.66	2.64	418.29	0.00	4.70	0.06	0.03	2.25	2.71	420.11	0.03	8.53	2.70	419.99	0.03	0.16	2.74	418.79
OUT 7140	KCl 3M	TRUE	gompertz	0.03	24.20	2.40	389.87	0.01	10.91	0.12	0.02	5.90	2.42	392.66	0.02	5.31	2.45	389.95	0.02	4.59	2.41	392.23
OUT 7140	NaCl 1.75M	TRUE	logistic	0.00	-13.44	0.34	42.54	0.00	28.91	0.05	0.00	-34.43	0.33	42.66	0.01	23.78	0.38	42.73	0.00	-31.19	0.33	42.73
OUT 7140	NaCl 2.5M	TRUE	gompertz.exp	NA	NA	NA	NA	NA	NA	NA	0.01	-5.15	0.17	27.46	0.01	-7.96	0.15	27.37	0.00	-74.98	0.13	25.99
OUT 7140	NaCl 3M	TRUE	gompertz.exp	NA	NA	NA	NA	NA	NA	NA	0.00	228.48	0.17	32.69	0.00	251.94	0.16	32.58	0.00	182.26	0.17	32.37
OUT 7140	YPD_37	TRUE	logistic	0.38	33.26	12.79	631.89	0.04	1.97	0.35	0.34	32.22	12.66	632.54	0.35	32.81	12.70	631.24	0.28	26.81	12.59	639.78
OUT 7140 bis	Gly_3%_1M_NaCl	TRUE	logistic	0.00	7.24	0.07	4.92	0.00	8.30	0.00	0.00	16.76	0.08	5.14	0.00	13.12	0.08	5.07	0.00	0.92	0.07	5.02
OUT 7140 bis	Gly_3%_2M_NaCl	TRUE	logistic	0.00	6.54	0.07	5.06	0.00	8.28	0.00	0.00	16.42	0.08	5.27	0.00	12.63	0.08	5.20	0.00	0.42	0.08	5.15
OUT 7140 bis	Gly_3%_CuSO4	TRUE	gompertz	0.01	-5.69	0.73	77.72	0.00	7.74	0.03	0.02	-0.14	0.75	78.09	0.02	-0.31	0.75	77.91	0.01	-4.89	0.75	77.22
OUT 7140 bis	Gly_3%	TRUE	logistic	0.09	29.41	2.76	95.33	0.01	1.20	0.05	0.06	22.51	2.79	96.78	0.07	23.71	2.76	96.75	0.06	22.68	2.81	94.67
OUT 7140 bis	Gly_4%	TRUE	logistic	0.06	24.08	2.04	79.64	0.00	0.86	0.03	0.05	20.90	2.07	80.97	0.05	19.37	2.07	80.67	0.04	16.97	2.07	79.55
OUT 7140 bis	Gly_3.5%_glu	TRUE	logistic	0.54	21.15	13.35	447.94	0.04	1.23	0.31	0.52	21.62	13.62	440.21	0.47	19.60	13.48	447.18	0.37	14.30	13.51	450.73
OUT 7140 bis	Gly_10%_glu	TRUE	gompertz	0.72	17.64	12.85	560.54	0.10	0.89	0.27	0.54	15.96	13.38	567.31	0.58	16.39	13.40	568.62	0.36	10.80	13.18	562.65
OUT 7140 bis	Gly_20%_glu	TRUE	gompertz	1.20	21.33	11.32	601.00	0.48	2.03	0.58	0.72	18.20	13.12	636.18	0.74	18.41	13.13	635.10	0.34	7.46	12.56	647.51
OUT 7140 bis	KCl 1_75M	TRUE	gompertz.exp	0.09	21.97	3.01	309.67	0.00	0.48	0.05	0.04	0.97	3.28	309.66	0.06	11.02	3.12	314.57	0.03	-1.74	3.38	304.10
OUT 7140 bis	KCl 2_5M	TRUE	gompertz	0.04	18.98	2.78	434.91	0.00	5.02	0.07	0.03	3.58	2.83	438.92	0.03	10.98	2.80	437.31	0.03	2.13	2.86	438.08
OUT 7140 bis	KCl 3M	TRUE	gompertz	0.03	24.94	1.95	348.87	0.00	4.00	0.03	0.02	4.98	2.01	347.15	0.02	4.47	2.01	348.37	0.02	2.70	2.06	346.53
OUT 7140 bis	NaCl 1.75M	TRUE	richards	0.00	3.91	0.36	44.06	0.00	12.88	0.01	0.00	1.18	0.36	43.87	0.00	-1.02	0.36	43.99	0.00	-27.78	0.38	44.01
OUT 7140 bis	NaCl 2.5M	TRUE	gompertz.exp	NA	NA	NA	NA	NA	NA	NA	0.02	-2.39	0.32	56.53	0.01	-4.37	0.30	55.91	0.00	-59.37	0.30	52.35
OUT 7140 bis	NaCl 3M	TRUE	gompertz.exp	NA	NA	NA	NA	NA	NA	NA	0.00	152.27	0.16	33.77	0.00	213.30	0.16	33.62	0.00	154.13	0.16	33.74
OUT 7140 bis	YPD_37	TRUE	richards	0.40	35.34	12.19	618.08	0.04	1.72	0.19	0.35	32.82	12.16	617.65	0.35	32.96	12.16	617.59	0.27	26.62	12.25	623.60

Supplementary Table S5.1. Growth parameters obtained wit *grofit* package.

Chapter 6:

Comparative assessment of strain typing techniques for fingerprinting of *Zygosaccharomyces rouxii* complex strains

6.1. Introduction

The *Zygosaccharomyces* genus accommodates the most important yeasts in terms of spoilage ability (Stratford 2006). This genus includes osmotolerant, strongly fermentative yeasts that are able to resist weak-acid preservatives such as benzoic and sorbic acids, and to survive at low aw environments, thereby, enhancing their well-known ability to promote spoilage (Casas et al., 2004; Stratford, 2006). Among these yeasts are *Z. mellis*, isolated from honey, syrups and from low aw products in general (Stratford, 2006), and the *Z. rouxii* and *Z. bailii* species, commonly found in the food and beverages industries. Although the *Z. rouxii* species are involved in fermentative elaboration of food such as miso and traditional balsamic vinegar (Solieri and Giudici, 2008); however, they are frequently isolated from low aw spoiled foods such as marzipan or nougat (Casas et al., 2004 and Martorell et al., 2005). According to the zymological indicators defined by Sancho et al. (2000), they are considered to be some of the most dangerous yeasts for product stability in fruit pulps and concentrates.

In order to establish contamination sources during food processing and thus avoid economic losses, it is essential both to identify species that are present and also to discriminate them at strain level. Nowadays, molecular phylogenetic and systematic studies are getting much complicated than ever thought. As mentioned in previous chapter (**refer section 1.1, Chapter 1**) that genus *Zygosaccharomyces*, in addition to identified species, also comprises of strains of postulated hybrid

origin or of uncertain taxonomical affiliation. Conventional methods, in particular those based on evaluation of physiological and morphological characteristics, are impractical as they are laborious, time-consuming, sometimes lead to incorrect classification and identification (Kurtzman and Robnett, 1998), and often impeded by database limitations. It has been reported that the strains of *Z. mellis* (such as CBS 5499 and NCYC 2403^{T}) are notoriously difficult to distinguish from *Z. rouxii* using only chemotaxonomic based identification methods (James and Stratford, 2011). In fact, separation of these two species appears, to some extent, reliant upon the fact that *Z. rouxii* is more tolerant to sodium chloride (16% NaCl/5% Glucose) than *Z. mellis* (James and Stratford, 2011). While *Z. mellis* shows no growth in presence of sodium chloride, but *Z. rouxii* shows weak or positive response (James and Stratford, 2011). In spite of this, some strains of *Z. rouxii* and *Z. mellis* exhibit slight differences in sodium chloride tolerance making chemotaxonomic based yeast identification not only laborious and time-consuming, but misleading too as discussed above. As result, techniques based on similarity or dissimilarity of DNA, RNA or proteins achieved increasing attention for yeast identification and phylogenic reconstruction of evolutionary relationships among the species at a variety of taxonomic levels. In particular, comparative analysis of partial 26S rRNA gene (Kurtzman and Robnett, 1998) and 18S rRNA gene sequencing (James et al., 1994; 1996) for phylogenetic based studies have gained considerable attention in past decade. Despite the wide use of rRNA phylogenetic marker, there are several aspects that limit the interpretation of rRNA-derived results (**refer Chapter 7**). The most important fact is that rDNA is present in yeast genome in tandem array (**refer section 1.2.1, Chapter 1**) where its copy numbers per genome vary from 45 up to 200 or more, and these copies may be associated with some inter-strain and intra-strains variation.

Another problem for appropriate *Zygosaccharoymces* detection during food spoilage is the selection of a suitable fingerprinting technique to be used. In order for some of the *Zygosaccharomyces* species to be strain-typed, several molecular techniques have been studied.

Török et al. (1993) proposed an electrophoretic karyotyping, while Esteve-Zarzoso et al. (2003) suggested RFLP of mtDNA. Martorell et al. (2005) demonstrated that if the objective is to differentiate species belonging to the same genus, the best result is obtained by electrophoretic analysis. A combination of molecular techniques may be chosen for specific purpose. For instance, to characterize *Z. bailli* and *Z. rouxii* at strain level, Martorell et al. (2005) suggested the combination of RFLP and RAPD analysis. A combination of several typing techniques was therefore required (Maqueda et al., 2010). More recently, the intergenic region (IGS) of rDNA has been proposed as typing method because its locus is more variable than other existing loci investigated so far (Sugita et al., 2001). Several studies have exploited this region (Wrent et al., 2010). Sequence analysis of the IGS region permits the separation of clinical isolates of *Cryptococcus neoformans* into two varieties (Diaz and Fell, 2000; Diaz et al., 2005; Fan et al., 1995). Strain typing of *Pichia anomala* has also been achieved by analyzing the sequence of the intergenic region 1 (IGS1) (Bhardway et al., 2007). Other studies have focused on discriminating strains belonging to different yeast species, such as *Phaffia rhodozyma* and *Xanthophyllomyces dendrorhous* (Fell and Blatt, 1999). The restriction fragment length polymorphism (RFLP) of the IGS2 region of rDNA has made possible to differentiate between the physiologically similar dairy yeast species *Kluyveromyces marxianus* and *K. lactis* (Naumova et al., 2005b); among halotolerant *Debaryomyces* species (Romero et al., 2005; Quirós et al., 2006); and among strains belonging to *Z. bailii*, *Z. mellis* and *Z. rouxii* species (Wrent et al., 2010).

The aim of this investigation is to comparatively evaluate the usefulness of M13-RAPD, GTG5-PCR, and PCR-RFLP analysis of the IGS region of rDNA as typing methods for rapid discrimination at strain level of isolates retrieved from salty and sugary niches.

6.2. Materials and methods

6.2.1. Strains, medium and culture conditions

In this study, a total of 79 strains strains, collected from different ambient and different culture collections as listed in **Table 6.1**, were used. The list includes *Z. rouxii* type strain CBS 732^T and six *Z. rouxii* isolates from TBV (ABT802, ABT803, ABT807 to ABT810), six *Z. mellis* isolates from TBV (ABT401 to ABT403; ABT405 to ABT407), fourteen *Z. sapae* isolates (ABT301^T to ABT306; ABT601 to ABT608) (Solieri et al., 2007), and two *Z. bailli* strains DBVPG 6920 and OUT 7140. Rest 49 strains were uncharacterized that include the strains collected from sugary niche such as honey (five isolates of 1990 sampling and nineteen isolates of 2008 sampling) and TBV (six isolates of 1989 sampling), and fourteen strains associated with salty niche. The latter mentioned salty niche strains have been procured from Biological Resource Center (NBRC), Japan where they were putatively ascribed as *Z. rouxii* strains. Besides this, five strains ATCC 42981, CBS 4837, CBS 4838, OUT 7136 and NCYC 3042 of uncertain taxonomical affiliation that showed heterogeneity in rDNA phylogenetic markers, were also included in this study for comparison (**refer Chapter 3**). *S. cerevisiae* ABT507 strain was used as the control strain. The strains were maintained and routinely grown in medium containing 1% yeast extract, 2% peptone, 2% glucose (YPD) with or without 2% agar. All the strains were long term cryo-preserved at -80°C at the Unimore Microbial Culture Collection (www.umcc.unimore.it)

6.2.2. DNA manipulation

Genomic DNA (gDNA) isolation procedure utilizes physical disruption of the yeast cell wall using small glass beads and Phenol-Chloroform based extraction protocol as described by Hoffman &

Winston (1987) and Sambrook et al., (1989) and as detailed in previous chapter (**refer section 2.2.2, Chapter 2**). Genomic DNA pellets obtained after extraction procedure were washed with 70% Ethanol, air dried for 20 min, resuspended in 35µL sterile double distilled water containing 1.5µL of RNAse (10mg/mL) and incubated at 37°C for 1hr in water bath. Finally, the extracted DNAs were stored at -20°C. Resultant genomic DNAs were diluted to 100ng/µL and used for performing for further experiments.

6.2.3. PCR amplification and 5.8S ITS-RFLP screening

The entire internal transcribed spacer (ITS) region (ITS 1, ITS 2 and the intervening 5.8S rRNA gene) for cloning purpose was amplified using primers ITS1 and ITS4 (White et al., 1990) as described in previous chapter (**refer section 2.2.4, Chapter 2**). For RFLP-based screening of the strains, restriction digestion of 5.8S ITS region of rRNA region with *Hae*III, *Hinf*I and *Hha*I restriction endonucleases was performed in accordance with manufacturer's instruction. Digested product were electrophoretically separated in a 1.8% agarose gel (containing Ethidium bromide) in 0.5X Trisborate-EDTA buffer and visualized under UV illumination (BioDoc Analyze, Biometra, Göttingen, Germany). GenRuler 100bp plus molecular weight marker was included in each run (MBI Fermentas).

Domains 1 and 2 of the 26S rRNA region (D1-D2), for restriction digestion with *Ava*I and cloning purpose, were amplified using primers NL1 and NL4 (O'Donnell, 1993) as described in previous chapter (**refer section 2.2.4, Chapter 2**). For sequencing purpose, 26S D1-D2 PCR products were purified by the DNA Clean and Concentrate TM-500 Kit (Zymo Research), following the manufacturer's instructions, and delivered to commercial sequencing service provider (MWG, Germany) for sequencing.

6.2.4. Strain genotyping and IGS restriction analysis

Microsatellite-primed PCR (MSP-PCR) fingerprinting and random amplified of polymorphic DNA (RAPD) analysis were carried out with primers $(GTG)_5$ (5'-GTGGTGGTGGTGGTG-3') and M13 (5'-GAGGGTGGCGGTTCT-3'), respectively (Baleiras Couto et al., 1996; Solieri et al., 2013a) as detailed in previous chapter (**refer section 2.2.3, Chapter 2**). The PCR products were separated on 1.8% agarose gels with electrophoretic run for 150 min at 80 V, visualized under UV, and photographed using BioDoc Analyser (Biometra GmbH, Germany).

The IGS region of the rDNA was amplified using CNL12 (5'-CTGAACGCCTCTAAGTCAG-3') and CNS1 (5'-GAGACAAGCATATGACTACTG-3') forward and reverse primers, respectively (Appel and Gordon, 1995; Wrent et al, 2010). A set of PCR reactions were carried out, using Takara LA Taq GC kit (TaKaRa Bio Inc, Shiga, Japan), in microtubes containing a master mix with a final volume of 25 µl, containing target DNA (100 ng), 1X LA Taq GC buffer I, 0.4mM of dNTP (0.4mM each), 0.5µM of each primer (MWG, Germany), 1.25 U of *La Taq* (TaKaRa) and sterilized distilled water up to 25 µl. The thermal cycling parameters were as follows: an initial denaturation step at 94°C for 5 min, followed by 35 cycles of 35 s at 94°C (denaturation), 1 min at 55°C (annealing), 3 min extension, and a final extension at 72°C for 10 min (Romero et al., 2005; Quirós et al., 2006). In order to improve the results for strains unable to amplify with above protocol, amplification was carried out in both GC Buffer I and GC Buffer II with concurrent decrease in the annealing temperature from 55 to 52°C and increasing extension time from 3 min to 5 min. Restriction fragments were separated on 2.5% (w/v) agarose gels, stained with 0.05% (v/v) ethidium bromide and visualized under UV light. The GenRuler 100bp Plus DNA ladder (MBI Fermentas) was used as a molecular size marker.

PCR amplification products from the IGS region of DNA (20 µl) were digested without further purification using *Hap*II, *Hha*I and *Mbo*I endonucleases (Amersham Pharmacia Biotech, Buckinghamshire, UK) (Romero et al., 2005 and Quirós et al., 2006). Restriction fragments were separated on 2.0% (w/v) agarose gels in 0.5×TBE (89 mM Tris-borate, 2 mM EDTA pH 8) buffer, stained with 0.05% (v/v) ethidium bromide and visualized under UV light Fragment lengths were estimated by comparing them with a GenRuller 100bp Plus DNA ladder (MBI Fermentas) as size marker. The gel was photographed using BioDoc Analyze (Biometra GmbH, Germany).

6.2.5. Image processing and clustering analysis

Gel images were processed using the BioNumerics (version 3.0; Applied Maths). For MSP-PCR and GTG5-PCR fingerprints, the dendrogram was generated using the Pearson's correlation similarity coefficient and the unweighted-pair group method using the arithmetic means (UPGMA) clustering method, with zero percent optimization and curve smoothening. For IGS RFLP-PCR patterns, we used both curve-based (Pearson's correlation coefficient) and band-based (Jaccard's coefficient) approach. For the curve-based algorithm, the comparison setting was kept as mentioned above, while for the band based method, the comparison setting with 1.0% tolerance and 1.0% optimization was used.

Category	Strains	Genetic background, thallism, ploidy, mating type and genetic markers	Source	Reference/Culture Collection/Strain Provider
Sugary niche	CBS 732	Haploid, Homothallic, *MATa*	Must, Italy	Yarrow, 1984
	ABT507	-	TBV, Italy	Solieri et al, 2006
	ABT301	Diploid	TBV, Italy	Solieri et al, 2006; Solieri et al., 2007; Solieri et al., 2008
	ABT302	-	TBV, Italy	Solieri et al, 2006; Solieri et al., 2007
	ABT303	-	TBV, Italy	Solieri et al, 2006; Solieri et al., 2007
	ABT304	-	TBV, Italy	Solieri et al, 2006; Solieri et al., 2007
	ABT305	-	TBV, Italy	Solieri et al, 2006; Solieri et al., 2007
	ABT306	-	TBV, Italy	Solieri et al, 2006; Solieri et al., 2007
	ABT601	Diploid	TBV, Italy	Solieri et al, 2006; Soleiri et al., 2007; Solieri et al., 2008
	ABT602	-	TBV, Italy	Solieri et al, 2006; Soleiri et al., 2007
	ABT603	-	TBV, Italy	Solieri et al, 2006; Soleiri et al., 2007
	ABT604	-	TBV, Italy	Solieri et al, 2006; Soleiri et al., 2007
	ABT605	-	TBV, Italy	Solieri et al, 2006; Soleiri et al., 2007
	ABT606	-	TBV, Italy	Solieri et al, 2006; Soleiri et al., 2007
	ABT607	-	TBV, Italy	Solieri et al, 2006; Soleiri et al., 2007
	ABT608	-	TBV, Italy	Solieri et al, 2006; Soleiri et al., 2007
	NCYC 3042	Hybrid	Sugar, UK	James et al., 2005
	ABT401	-	TBV, Italy	Solieri et al, 2006
	ABT402	-	TBV, Italy	Solieri et al, 2006
	ABT403	-	TBV, Italy	Solieri et al, 2006
	ABT405	-	TBV, Italy	Solieri et al, 2006
	ABT406	-	TBV, Italy	Solieri et al, 2006
	ABT407	-	TBV, Italy	Solieri et al, 2006
	ABT802	-	TBV, Italy	Solieri et al, 2006
	ABT803	-	TBV, Italy	Solieri et al, 2006
	ABT807	-	TBV, Italy	Solieri et al, 2006
	ABT808	-	TBV, Italy	Solieri et al, 2006
	ABT809	-	TBV, Italy	Solieri et al, 2006
	ABT810	-	TBV, Italy	Solieri et al, 2006
	M21	-	Honey-1990 sampling	UMCC, Italy
	M22	-	Honey-1990 sampling	UMCC, Italy
	M23	-	Honey-1990 sampling	UMCC, Italy
	M24	-	Honey-1990 sampling	UMCC, Italy
	M25	-	Honey-1990 sampling	UMCC, Italy
	2	-	Honey-2008 sampling	UMCC, Italy
	3	-	Honey-2008 sampling	UMCC, Italy
	4	-	Honey-2008 sampling	UMCC, Italy
	7	-	Honey-2008 sampling	UMCC, Italy
	9	-	Honey-2008 sampling	UMCC, Italy
	12	-	Honey-2008 sampling	UMCC, Italy
	23	-	Honey-2008 sampling	UMCC, Italy

Category	Strains	Genetic background, thallism, ploidy, mating type and genetic markers	Source	Reference/Culture Collection/Strain Provider
Sugary niche	24	-	Honey-2008 sampling	UMCC, Italy
	27	-	Honey-2008 sampling	UMCC, Italy
	35	-	Honey-2008 sampling	UMCC, Italy
	40	-	Honey-2008 sampling	UMCC, Italy
	41	-	Honey-2008 sampling	UMCC, Italy
	56	-	Honey-2008 sampling	UMCC, Italy
	68	-	Honey-2008 sampling	UMCC, Italy
	70	-	Honey-2008 sampling	UMCC, Italy
	76	-	Honey-2008 sampling	UMCC, Italy
	1CF	-	Honey-2008 sampling	UMCC, Italy
	5CF	-	Honey-2008 sampling	UMCC, Italy
	6C	-	Honey-2008 sampling	UMCC, Italy
	B8911	-	TBV, Italy-1989 sampling	UMCC, Italy
	B8932	-	TBV, Italy-1989 sampling	UMCC, Italy
	B8933	-	TBV, Italy-1989 sampling	UMCC, Italy
	B8941	-	TBV, Italy-1989 sampling	UMCC, Italy
	B8943	-	TBV, Italy-1989 sampling	UMCC, Italy
	B89221	-	TBV, Italy-1989 sampling	UMCC, Italy
Salty niche	ATCC 42981	Allodiploid	Miso, Japan	Kiuchi et al. 1980; Gordan and Wolfe, 2008; Solieri et al., 2008
	CBS 4837	Aneuploid, Heterothallic, *MATa*	Miso, Japan	Wickerham & Burton 1960; James et al., 2005; Solieri et al, 2013b
	CBS 4838	Aneuploid, Heterothallic, *MATα*	Miso, Japan	Wickerham & Burton 1960; Solieri et al., 2013b
	OUT 7136	Aneuploid	Soy-moromi, Japan	Prof. Y. Kaneko (Osaka University, Japan); Solieri et al, 2013b
	NBRC0495	-	Shoyu-moromi or Shoyu-koji, Japan	Takahashi and Yukawa, 1915
	NBRC0505	-	Shoyu yeast, Japan	Seuzawa et al., 2008
	NBRC0506	*MATa*	Shoyu yeast, Japan	Seuzawa et al., 2008
	NBRC0521	*MATa*	Shoyu-moromi, Japan	Seuzawa et al., 2008
-	NBRC0523	*MATa*	-	Seuzawa et al., 2008
	NBRC0525	-	Miso, Japan	Seuzawa et al., 2008
	NBRC0845	-	Mash of Tamari soya, Japan	Seuzawa et al., 2008
	NBRC0846	*MATa*	Mash and koji of tamari-soya, Japan	Ohara and Nonomura, 1954
-	NBRC10652	Segregant of hybrid between NRRL2547(a) & 2548(α); *MATa*; Arg1	-	Mori, 1973; Mori and Windisch, 1982
-	NBRC10655	Segregant of hybrid between Lys-M5(a) & Lys-M12(α); ***MATα/MATα***	-	Mori, 1973
	NBRC10668	*MATα*	Shoyu-mash, Japan	Mori and Onishi, 1967; 1969
	NBRC10669	*MATa*	Shoyu-mash, Japan	Mori and Onishi, 1967; 1970
-	NBRC10670	*MATα*; ***MATα/MATα***	-	Mori, 1973
	NBRC10672	*MATα*; ***MATα/MATα***	Shoyu yeast, Japan	-
	DBVPG 6920	Diploid	Salad dressing, USA	Industrial Yeast collection (DBVPG, Italy)
	OUT 7140	Diploid	-	Prof. Y. Kaneko (Osaka University, Japan)

Table 6.1. ***Zygosaccharomyces*** **strains from sugary and salty-niche, their strain codes, source of isolation, and genetic background.**

6.3. Results and discussion

6.3.1. Ribosomal DNA restriction analysis reveals rDNA heterogeneity

In this study, we used 49 uncharacterized strains isolated from sugary (honey or TBV) and salty niche (soy, miso, and similar food niches) (**Table 6.1**). In order to assign a proper taxonomical and phylogenetic position to these strains, we performed a series of experiment as depicted in **Figure 6.1**. Firstly, we performed 5.8S ITS PCR-RFLP using three restriction endonucleases *Hae*III, *Hinf*I and *Hha*I. Following this, the representative strains were chosen from each group that represents strains with identical 5.8S ITS PCR-RFLP pattern and were subjected to 26S D1-D2 domain sequencing. Besides this the strains that remain single were also subjected to 26S D1-D2 domain sequencing. Firstly, we amplified the 5.8S ITS region and the PCR amplicons were subjected to RFLP analysis. **Tables 6.2** shows the sizes of the PCR products and the restriction fragments obtained using the restriction endonucleases *Hae*III, *Hinf*I and *Hha*I. Fragments smaller than 50 bp, could not be reproducibly visualized, and therefore, not included in the **Table 6.2**. Based on the RFLP-fingerprint pattern obtained for each strain, we put strains with identical banding pattern in a group and those strains that not share banding patter with any of the strain studied were kept single. We, on the basis of 5.8S ITS PCR-RFLP, resolved 79 strains (comprising 29 identified, 49 unidentified and 1 *S. cerevisiae*) into 9 restriction pattern types. Out of 79 strains, 54 strains that showed shared restriction pattern with one or more strains were put together and a group was assigned to them (from G1 to G9); however, rest 14 strains remained as singlet along with *S. cerevisiae* strain ABT 507 (**Table 6.2**).

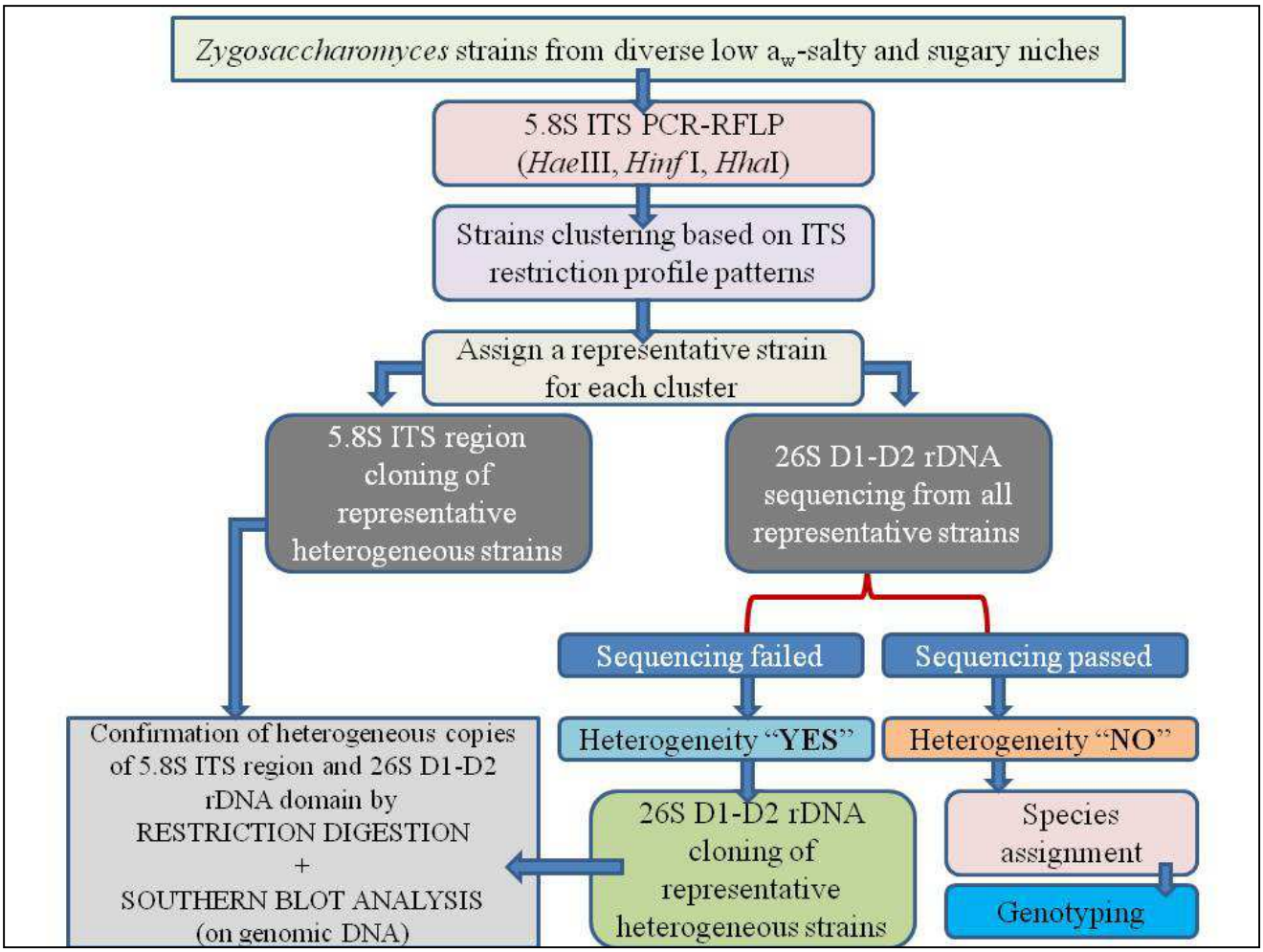


Figure 6.1. A schematic diagram for assigning a proper phylogenetic status to uncharacterized strains. Strain were retrieved from different ecological niche.

Group G1 to G5 and G7 accommodate strains with homogeneous 5.8S ITS PCR-RFLP pattern. Our 26S D1-D2 domain sequencing effort with the representative strains from these groups was successful for all strains, except strain NBRC0505 representing group G7. This suggest that strains belonging to group G1 to G5 are homogeneous with respect to both 5.8S ITS region and 26S D1-D2 domain region. However, strains belonging to group G7 are homogeneous with respect to their 5.8S ITS region but not with respect to their 26S D1-D2 domain.

Strains	Amplicon size	5.8S ITS gene			Grouping based on 5.8S ITS-RFLP profile	26S D1-D2 domain Sequencing	Species assignment	References
		*Hae*III	*Hinf*I	*Hha*I				
ABT 507	880	320-240-180-150	370-370-140	380-360-150			*S. cerevisiae*	Solieri et al., 2006
OUT 7140	800	690-100	360-220-170	300-270-100			*Z. bailii*	This study
DBVPG 6920	800	700-100	360-230-180	310-280-100			*Z. bailii*	This study
ABT401 ABT402 ABT403 ABT405 ABT406 ABT407	700	520-180	320-280-120	280-220	G-1	Passed	*Z. mellis*	Solieri et al., 2006
2 3 12 24 56 68 76	900	520-200-80	380-280-180	300-220-190-100	G-2	Passed	*Z. mellis*	This study
4 7 23 35 40 1CF 6C	900	680-170	480-290-140	310-190-100	G-3	Passed	*Z. mellis*	This study
9 27 41	900	390-250-170	480-290-140	300-190-100	G-4	Passed	*Z. mellis*	This study
70	900	680-170	480-290-140	380-320-130		Passed	*Z. mellis*	This study
NCYC 3042	700	480-170-90	360-250-220-180	320-290-100		Passed	Mosaic	James et al., 2005; Solieri et al., 2013b
CBS 732^{T}	700	390-210-90	360-250-180	290-200-170-100	G-5	Passed	*Z. rouxii*	Solieri et al.,2006
OUT 7136	700	390-210-90	360-250-180	290-200-170-100	G-5	Passed	Mosaic	Solieri et al., 2013b
ABT802 ABT803 ABT807 ABT808 ABT809 ABT810	700	390-210-90	360-250-180	290-200-170-100	G-5	Passed	*Z. rouxii*	Solieri et al., 2006
B8911 B8932 B8933 B8941 B89221	700	390-210-90	360-250-180	290-200-170-100	G-5	Passed	*Z. rouxii*	This study
ATCC 42981 CBS4837 CBS 4838	700	480-390-250-210-170-90	380-360-250-230-180	360-290-200-170-100	G-6	Failed	Mosaic and uncharacterized	Solieri et al., 2013b
NBRC0495	700	480-440-220-170	360-250-220-110	320-290-200-170-100		Failed	uncharacterized	This study
NBRC0505….NBRC0506.. NBRC0521…. NBRC0523	700	480-170-90	360-220-180	320-290-100	G-7	Failed	uncharacterized	This study
NBRC0525	700	480-170-90	360-220-180	320-290-100		Not performed	Not performed	This study
NBRC0845 .. NBRC0846	700	390-210-90	360-250-180	290-200-170-100	G-5	Passed	*Z. rouxii*	This study
NBRC10652 NBRC10655	700	390-210-170-90	360-250-220-180	320-290-200-170-100	G-8	Failed	uncharacterized	This study
NBRC10668	700	390-170-90	360-250-220-180	320-290-100		Passed	uncharacterized	This study
NBRC10669	700	480-390-210-90	360-280-250-220-180	290-200-170-100		Failed	uncharacterized	This study
NBRC10670	700	480-390-210-90	360-250-180	290-200-170-100		Failed	uncharacterized	This study
NBRC10672	700	480-390-250-210-170-90	360-250-220-150	290-200-170-100		Failed	uncharacterized	This study
AB 301 ABT302 ABT303 ABT304 ABT305 ABT306	700	510-480-390-250-210-170-90	380-360-260-250-220-180	360-290-200-170-100	G-9	Passed	*Z. sapae*	Solieri et al., 2007; Solieri et al., 2013a
ABT601 ABT602 ABT603 ABT604 ABT605 ABT606 ABT607 ABT608	700	510-480-390-250-210-170-90	380-360-260-250-220-180	360-290-200-170-100	G-9	Passed	*Z. sapae*	Solieri et al., 2007; Solieri et al., 2013a
M21 M22 M23 M25	700	510-480-390-250-210-170-90	380-360-260-250-220-180	360-290-200-170-100	G-9	Passed	*Z. sapae*	This study
5CF	700	700	360-270-250-220-180	360-290-200-170-100		Passed	*D. hansenii*	This study
B8943	650	420-150	450-280	300-290		Not performed	Not performed	This study

Strains in bold are subjected to D1-D2 sequencing where the strains that showed polymorphism were later cloned. Strains in bold and underlined were subjected to D1-D2 sequencing and their species assignment was done as mentioned in listed reference.

Table 6.2. Strain characterization and species assignment based on 5.8S ITS-RFLP and 26S D1-D2 sequencing.

In group G6, G8 and G9 strains the sum of bands obtained from at least one of the enzyme used for 5.8S ITS PCR-RFLP based screening exceeded the size of 5.8S ITS PCR amplicon, and therefore, the strains belonging to these groups were considered heterogeneous with respect to their 5.8S ITS region. The group G6 accommodates three strains (ATCC 42981, CBS 4837 and CBS 4838) that have shown to possess three haplotypes for their 5.8S ITS region and two haplotypes for their 26S D1-D2 region as previously (**refer Chapter 3**) and reported in Solieri et al. (2013b). These strains along with singlet strain OUT 7136 have been referred as mosaic strains (**refer Chapter 3, Solieri et al., 2013b**). However, 26S D1/D2 domain sequencing showed that the representative strain of G8 (NBRC10652) and G9 (M21) are not heterogeneous with respect to 26S D1/D2 domain region as sequencing was successful.

6.3.2. Strain identification

Based on 5.8S ITS-RFLP profile followed by 26S D1/D2 sequencing, G1 strains isolated from TBV (ABT401-ABT 403, ABT405-ABT 407) have been characterized as *Z. mellis* (Solieri et al., 2006). Sequencing of the 26S D1/D2 domain of the strains isolated from honey ascribed groups G2-G4 strains as *Z. mellis* (2008 sampling) and group G9 strains as *Z. sapae* (1990 sampling). One strain 5CF that have been isolated from honey (2008 sampling) was characterized as *Debaryomyces hansenii*. Strain isolated from honey and identified as *Z. sapae* (1990 sampling) showed complex 5.8S ITS PCR-RFLP similar to what obtained for other fourteen *Z. sapae* isolates obtained from TBV (Solieri et al, 2006). Six strains of group G5 obtained from TBV (ABT802, ABT803, ABT807 to ABT810) were previously characterized as *Z. rouxii* strains (Solieri et al., 2006). In current study, the remaining five strains (out of six strain of G5 group) also retrieved from TBV (1989 sampling) were also identified as *Z. rouxii* strains leaving B8943 as uncharacterized. Group G5 strains have 5.8S ITS PCR-RFLP pattern similar to that of the *Z. rouxii* type strain CBS 732^{T}. The group G6 accommodates

three mosaic strains (ATCC 42981, CBS 4837 and CBS 4838) as discussed earlier. We obtained fourteen strains associated with fermentative elaboration of Japanese soy and miso based condiments from Biological Resource Center, Japan (NBRC). These strains have been designated as *Z. rouxii* strains in NBRC culture collection. However, based on 5.8S ITS-RFLP fingerprint profile, only two strains NBRC0845 and NBRC0846 (belonging to group G5) actually showed typical *Z. rouxii* CBS 732^{T} like 5.8S ITS PCR-RFLP banding pattern. Remaining 12 strains these strains showed complex banding pattern (sum of bands greater that size of 5.8S ITS PCR amplicon) and of these 12 strains some were grouped in two groups G7 and G8. The strains of these two groups showed heterogeneity in their rDNA region quite similar to that we previously obtained with mosaic strains of *Z. rouxii* complex as discussed earlier in this section and previous chapter (**refer Chapter 3; Solieri et al., 2013b**). The mosaic strains have shown considerable heterogeneity in their rDNA region (Solieri et al., 2013b). Besides this, mosaic strains harbor different copies for 5.8S ITS region and D1-D2 rDNA region in their genome. Therefore, in order to interpret of rRNA-derived results (**refer Chapter 3**), we studied the rDNA heterogeneity in mosaic strains as well as fourteen uncharacterized salty-niche strain in more depth using Southern blot analysis and by prediction of RNA structure of sequence arising from divergent copies of 26S D1-D2 domain (**refer Chapter 7**).

In summary, *Z. mellis* strains were assigned four groups from G1-G4. Twenty three out of twenty five TBV and honey (2008 sampling) isolates resulted in four groups each comprising strains with different restriction pattern type. Two strains 70 and 5CF, however, remained as singlet out of which the strain 70 was characterized as *Z. mellis* and the strain 5CF was characterized as *D. hansenii*. Thirteen strains of *Z. rouxii*, retrieved from both salty-niche and TBV, were analyzed and put together in group G5 along with *Z. rouxii* type strain CBS 732. The strain OUT 7136 of group G5, however, showed 5.8S ITS PCR-RFLP profile similar to *Z. rouxii* type strain CBS 732, but it possess 26S D1-D2 domain of *Z. pseudorouxii*-like (Solieri et al., 2013b). G6 housed mosaic strains. Twelve

strain (out of fourteen strains) obtained from NBRC culture collection remained uncharacterized. Out of these twelve strains seven were assigned group (G7 and G8) and six other strains (NBRC0495, NBRC0525, NBRC10668, NBRC10669, NBRC10670, NBRC10672) remained as singlet. Eighteen strains, four from honey (1990 sampling) and fourteen from TBV (2005 sampling) showed same restriction profiles and were grouped in one single group G9 as *Z. sapae* strains. The *Z. pseudorouxii* strain NCYC 3042 putatively ascribed as conspecific with *Z. sapae* (James et al., 2005) has given different restriction profile, and therefore, kept as singlet. Two *Z. bailii* strains OUT 7140 and DBVPG 6920 showed slight, but visible difference in size of restriction fragments that make us to keep both strains also as singlet.

6.3.3. Strain typing

For strain-typing of the collection of 78 osmo- and halo-tolerant strains plus *S. cerevisiae* strain ABT507 (**Table 6.1**), three genotyping techniques, RAPD-M13-PCR, microsatellite primed $(GTG)_5$-PCR, and IGS-PCR, were employed and compared. Since the M13-RAPD and GTG_5-PCR are based on the amplification of random DNA segments with single primers of arbitrary nucleotide sequence, we firstly tested the reproducibility of these fingerprinting techniques on three independent DNA preparations from randomly selected strains. The results showed a good reproducibility for both techniques, with only the low intensity bands being not consistently reproducible (data not shown). Therefore, these bands were not taken into consideration when the patterns were compared and clustered.

6.3.3.1 M13-RAPD typing analysis

In M13-RAPD typing, we obtained from 6 to 17 amplified electrophoretic bands, with size from 250 to 2920 bp (**Table 6.3**). PCR fingerprints displayed intra and interspecies variability

although PCR profiles obtained from different strains of the same species were far more similar than those derived from different *Zygosaccharomyces* species. The UPGMA dendrogram resulting from M13-RAPD fingerprinting shows two major clusters, branching at a similarity value of 20.9% (**Figure 6.2**). At above mentioned comparison settings, we distinguished eight clusters at similarity percent cut-off of 83.2 (cluster 1), 76.3 (cluster 2), 70.6 (cluster 3), 70.9 (cluster 4), 79.0 (cluster 5), 59.7 (cluster 6), 54.3 (cluster 7), and 55.4 (cluster 8).

Clusters	Strains	M13-RAPD band profile
Cluster 1	NBRC10668	1510, 1290, 1160, 1000, 950, 870, 810, 700, 670, 620, 590, 470, 420
	NBRC10669	1510, 1300, 1010, 810, 690, 670, 610, 470, 370
	NBRC10670	1570, 1490, 1210, 1160, 1110, 1040, 950, 700, 670, 620, 580, 460
	NBRC10672	1580, 1490, 1270, 1160, 980, 700, 660, 620, 580, 460
Cluster 2	ATCC 42981	2920, 1920, 1500, 1230, 920, 680, 650, 610, 580, 460
	CBS 4837	2920, 1960, 1480, 1230, 940, 690, 650, 610, 580, 470
	CBS 4838	2810, 1920, 1480, 1220, 920, 690, 650, 610, 580, 470
	OUT 7136	2020, 1660, 1460, 1190, 1100, 1060, 940, 780, 680, 650, 610, 580, 520, 470, 370
	NBRC0495	1510, 1230, 940, 670, 640, 600, 570, 450
	NBRC0505	1500, 1230, 950, 680, 650, 600, 570, 460
	NBRC0506	1530, 1240, 950, 680, 650, 610, 580, 460
	NBRC0521	1550, 1240, 960, 650, 610, 460
	NBRC0523	1500, 1220, 930, 680, 640, 610, 570, 460
	NBRC0845	1510, 1220, 930, 680, 640, 600, 570, 450
	NBRC0846	1530, 1230, 950, 680, 650, 610, 580, 520, 460
	NBRC10652	1510, 1220, 940, 690, 650, 610, 580, 460
	NBRC10655	1550, 1250, 1180, 1140, 960, 650, 610, 580, 460
Cluster 3	301	1800, 1620, 1280, 1220, 980, 680, 640
	302	1730, 1530, 1230, 1170, 950, 670, 630
	303	1710, 1550, 1220, 1170, 960, 670, 640
	304	1700, 1560, 1230, 1170, 950, 670, 630
	305	1720, 1560, 1240, 1190, 970, 670, 630
	306	1710, 1570, 1240, 1190, 960, 680, 630
	601	1700, 1580, 1250, 1190, 960, 680, 640
	602	1710, 1570, 1240, 1190, 960, 680, 640
	603	1690, 1560, 1240, 1190, 960, 670, 640
	604	1680, 1540, 1250, 1180, 960, 670, 640
	605	1700, 1540, 1250, 1180, 970, 660, 640
	606	1690, 1540, 1250, 1180, 960, 660, 630
	607	1690, 1530, 1240, 1170, 960, 670, 640
	608	1680, 1510, 1240, 1160, 960, 670, 630
	NBRC0525	1500, 1210, 930, 630, 600, 570, 450
Cluster 4	CBS 732	2010, 1580, 1300, 1190, 1080, 1030, 930, 810, 660, 620, 590, 530, 470, 380
	802	2010, 1470, 1270, 1080, 780, 630, 590, 560, 500, 360
	803	2030, 1090, 990, 900, 780, 630, 590, 560, 450
	807	2100, 1520, 1100, 1060, 1000, 790, 640, 590, 570, 450, 360
	808	2090, 1110, 1000, 910, 800, 640, 600, 570, 450, 360
	809	2080, 1110, 1000, 910, 790, 640, 600, 570, 450, 360
	810	2070, 1110, 1070, 1000, 920, 870, 800, 650, 630, 610, 570, 460, 410
	M21	960, 610, 460
	B8911	1940, 1450, 1280, 1090, 1050, 980, 890, 780, 630, 590, 570, 460
	B8932	1960, 1090, 1050, 980, 900, 780, 630, 600, 570, 420
	B8933	1940, 1320, 1100, 1060, 1000, 910, 640, 610, 580
	B8941	1950, 1320, 1110, 1060, 1000, 920, 650, 610, 580
	B89221	1950, 1300, 1110, 1070, 1010, 920, 650, 620, 590

Cluster	Strains	M13-RAPD band profile
Cluster 5	401	1640, 1490, 1390, 1200, 1130, 900, 630, 600, 480, 460, 400
	402	1600, 1450, 1360, 1170, 1100, 870, 620, 580, 550, 470
	403	1620, 1490, 1400, 1170, 1120, 880, 630, 590, 560, 470
	405	1950, 1490, 1390, 1190, 1130, 890, 810, 630, 590, 560
	406	1640, 1490, 1390, 1200, 1120, 980, 890, 630, 590, 560, 480
	407	1660, 1190, 1140, 980, 900, 640, 590, 480, 390
	M22	1660, 1540, 1430, 1230, 1160, 930, 610, 490, 460
	M23	2130, 1660, 1560, 1420, 1230, 1160, 930, 610, 490, 470
	M25	1700, 1520, 1430, 1230, 1160, 930, 610, 490, 460
Cluster 6	OUT 7140	2080, 1520, 1340, 1090, 970, 730, 620
	DBVPG 6920	2080, 1520, 1340, 1090, 970, 710, 660, 610
	B8943	1980, 1460, 1060, 950, 700, 660, 630, 600, 580, 530, 490, 470, 410, 380, 250
Cluster 7	2	1990, 1530, 1410, 1170, 1120, 1020, 940, 780, 700, 630, 590, 520, 440, 340, 270
	3	1990, 1510, 1400, 1150, 1080, 1040, 940, 900, 760, 690, 660, 570, 510, 480, 360, 270
	12	1990, 1570, 1420, 1150, 1000, 940, 900, 770, 700, 670, 520, 490, 390, 280
	24	2060, 1440, 1170, 1030, 970, 780, 720, 690, 610, 530, 490, 290
	56	2130, 1600, 1500, 1210, 1070, 1010, 970, 820, 750, 720, 650, 620, 550, 520, 400, 300
	68	2247, 1730, 1480, 1170, 1020, 820, 760, 550, 310
	76	2040, 1570, 1450, 1180, 1110, 980, 800, 720, 690, 530, 510, 390
Cluster 8	4	2920, 2000, 1490, 1370, 1110, 1060, 950, 770, 720, 640, 460, 440, 340, 270
	7	2910, 2080, 1520, 1400, 1130, 1070, 960, 770, 720, 650, 460
	9	2030, 1550, 1430, 1160, 1100, 1020, 970, 770, 720, 640, 560, 350
	23	2060, 1580, 1460, 1160, 1100, 990, 940, 790, 720, 660, 470, 440, 350
	27	2060, 1440, 1170, 1030, 970, 780, 720, 690, 610, 530, 490, 290
	35	1570, 1450, 1130, 720, 670, 610, 490, 370
	40	1600, 1480, 1150, 730, 680, 490, 380
	41	2160, 1600, 1460, 1170, 1120, 1030, 820, 770, 690, 500, 370
	70	2250, 1660, 1540, 1240, 1090, 840, 780, 700, 570, 500, 450
	1CF	1660, 1480, 1160, 730, 650, 480, 370
	6C	2010, 1610, 1450, 1330, 1180, 1120, 1080, 1010, 970, 810, 750, 720, 680, 540, 490, 460, 370
	NCYC 3042	2920, 2150, 1620, 1420, 1220, 950, 690, 620, 500
Outgroup	5CF	710, 640, 560, 460

Table 6.3. RAPD-PCR fingerprint profile obtained from different *Zygosaccharomyces* strains with M13 primer.

Cluster 1 and 2 comprise of strains isolated from salt-rich habitats. In addition, cluster 2 comprises of strain ATCC 42981, which is supposed to possess an allodiploid genome, harboring p- and t-subgenomes, respectively. (Gordan and Wolfe, 2008). Cluster 2 also housed strain CBS 4837 (NCYC 1682) and its sister strain CBS 4838; and the former has been found to possess two highly divergent copies of the nuclear-encoded *ADE2*, *HIS3* and *SOD2* genes indicating its postulated hybrid nature (James et al., 2005). Moreover, both strains possessed aneuploid genome content packed in

eight chromosomes amounting for total genome size of 21.7±0.33Mb and 22.5±0.20, respectively for strain CBS 4837 and strain CBS 4838. Thirty-six isolates obtained from Traditional balsamic vinegar grouped as Cluster 3, 4, and 5. Fourteen isolates belonging to cluster 3 represent a recently identified novel *Zygosaccharomyces* yeast species, *Z. sapae* (Solieri et al., 2007; Solieri et al., 2013a). Cluster 4 comprises mainly of TBV isolates characterized as *Z. rouxii*, and the type strain of *Z. rouxii* CBS 732^{T}; whereas, cluster 5 accommodated TBV isolates characterized as *Z. mellis*. Cluster 6 contains three strains, of which strain B8943 is uncharacterized and was isolated from TBV; while other two strains DBVPG 6920 and OUT 7140 are *Z. bailii* strains. Cluster 7 and 8 comprises of seven and eleven strains, respectively. These strains were recovered from honey clustered together with strain NCYC 3042, which was also recovered from sugary niche and has been supposed to be implicated in allodiploid nature of strain ATCC 42981. Another strain 5CF also recovered from honey was characterized as *D. hansenii*, which served as an outgroup for RAPD-M13 genotyping analysis (**Figure 6.2**).

Different patterns were generated for strains belonging to each of the four *Zygosaccharomyces* species tested. Most of the strains tested for these four species showed a similarity higher than 75%. At the strain level, different polymorphisms were observed in the different species. However, the RAPD PCR of strains also yielded a high number of shared bands between strains, of them most can be easily distinguished. Strains of cluster 1 showed six identical amplified fragment whose size were approximated around 460bp, 610bp, 660bp, 690bp, 1160 (except in NBRC10669), and 1490bp. Likewise, there were six identical fragments in all the amplified profiles of strains isolated from salty niche (cluster 2). The length of the identical fragments were approximated around 450bp, 600bp, 640bp, 670bp (except in NBRC0521), 920bp, and 1220bp (except in OUT 7136 which instead contain 1190 bp fragment). Three strain of this cluster ATCC 42981, CBS 4837 and CBS 4838 showed highly consistent band pattern that coincided with the same species and common ecological

niche (Solieri et al., 2013b). These results indicated that the strains that have been isolated from the same source tend to show a high degree of similarity between fingerprint fragments. The RAPD profiles of cluster 3 strains that putatively represent a novel yeast species were exceptionally consistent and showed almost identical band sizes and band number. The low degree of genetic diversity in these strains indicates the primary clonal nature of the population consistent with our previous finding where they rarely reproduce sexually and form ascospores (Solieri et al., 2013a). On the other hand, it has been observed that the strains of the same species that have been isolated from different sources tend to give a high degree of diversity between fingerprint fragments. It was clearly seen in case of *Z. rouxii* strains (cluster 1, 2, and 4) where the strains isolated from TBV samples and that from salty niche showed considerable difference in band sizes, band number and overall banding profile. Although, *Z. mellis* strains isolated from TBV (cluster 5) showed high degree of consistency, strains isolated from honey (cluster 7 and 8) showed unusually inconsistent RAPD profile and showed high degree of diversity among different isolates. Similar intra-specific diversity was evident in *Z. mellis* strains isolated from TBV (cluster 5) and honey (cluster 7 and 8). The RAPD-M13 PCR analysis allowed us to differentiate two different patterns in the *Z. bailii* strains. **Table 6.3** also shows two patterns corresponding to strains *Z. bailii* DBVPG 6920 and OUT 7140 (cluster 6) that showed considerable difference from the *Z. rouxii*, *Z. mellis* and *Z. sapae* specific patterns of cluster 4 indicate the discriminatory power of the RAPD method on the species level with the primer M13.

RAPD-PCR analysis also yielded some conflicting results. Strains M21, M22, M23, M25, which were clustered together with fourteen *Z. sapae* TBV isolates, however, on the basis of 5.8S-ITS RFLP, in the UPGMA dendrogram clustered them with either *Z. rouxii* strain ABT803 (cluster 4) or *Z. mellis* TBV isolates (cluster 5).

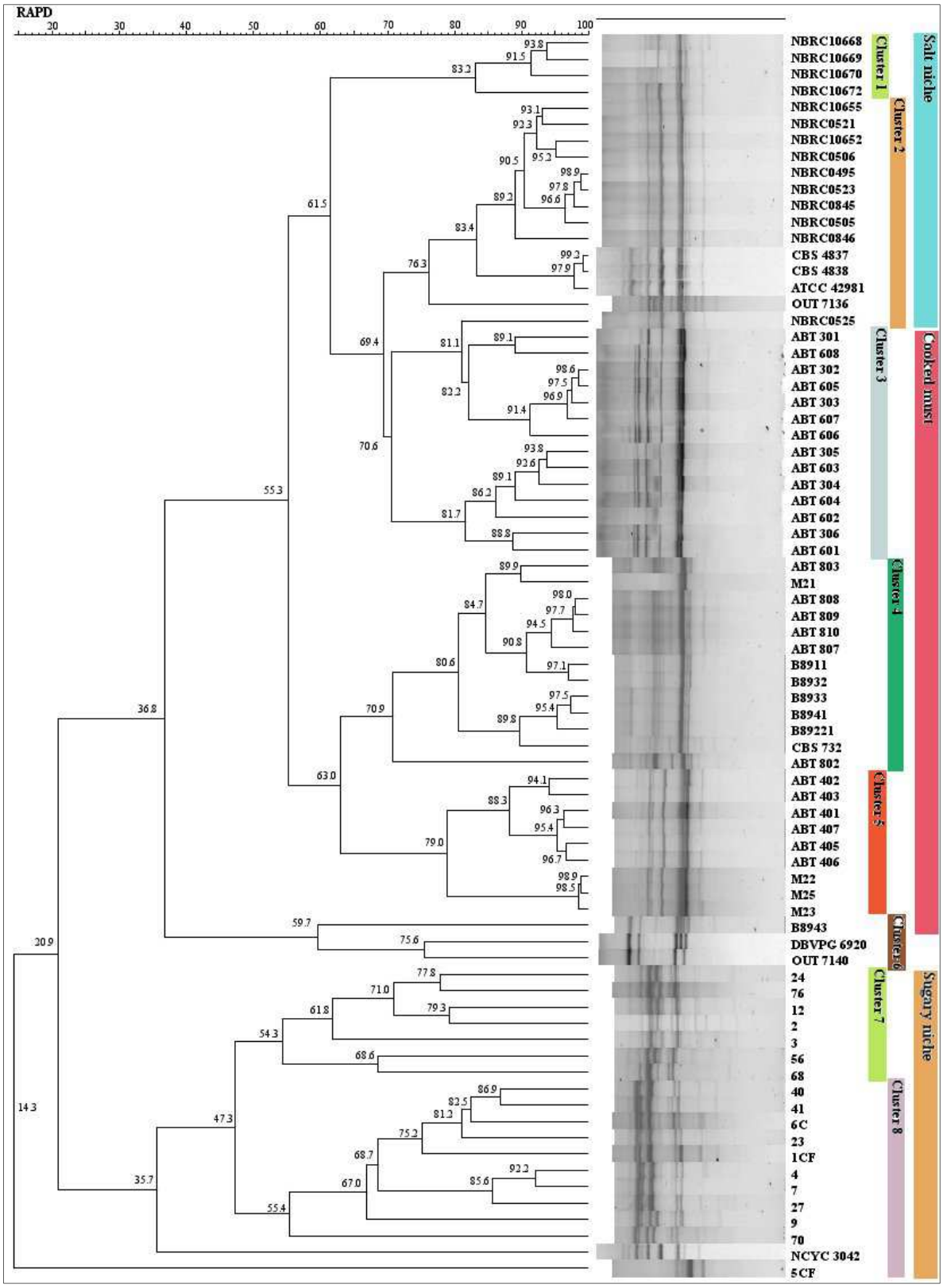


Figure 6.2. Pearson's coefficient based UPGMA dendrogram for *Zygosaccharomyces* strains (M13-RAPD-PCR).

Robustness of the tree was estimated with bootstrap values on 1000 replicates indicated as a percentage.

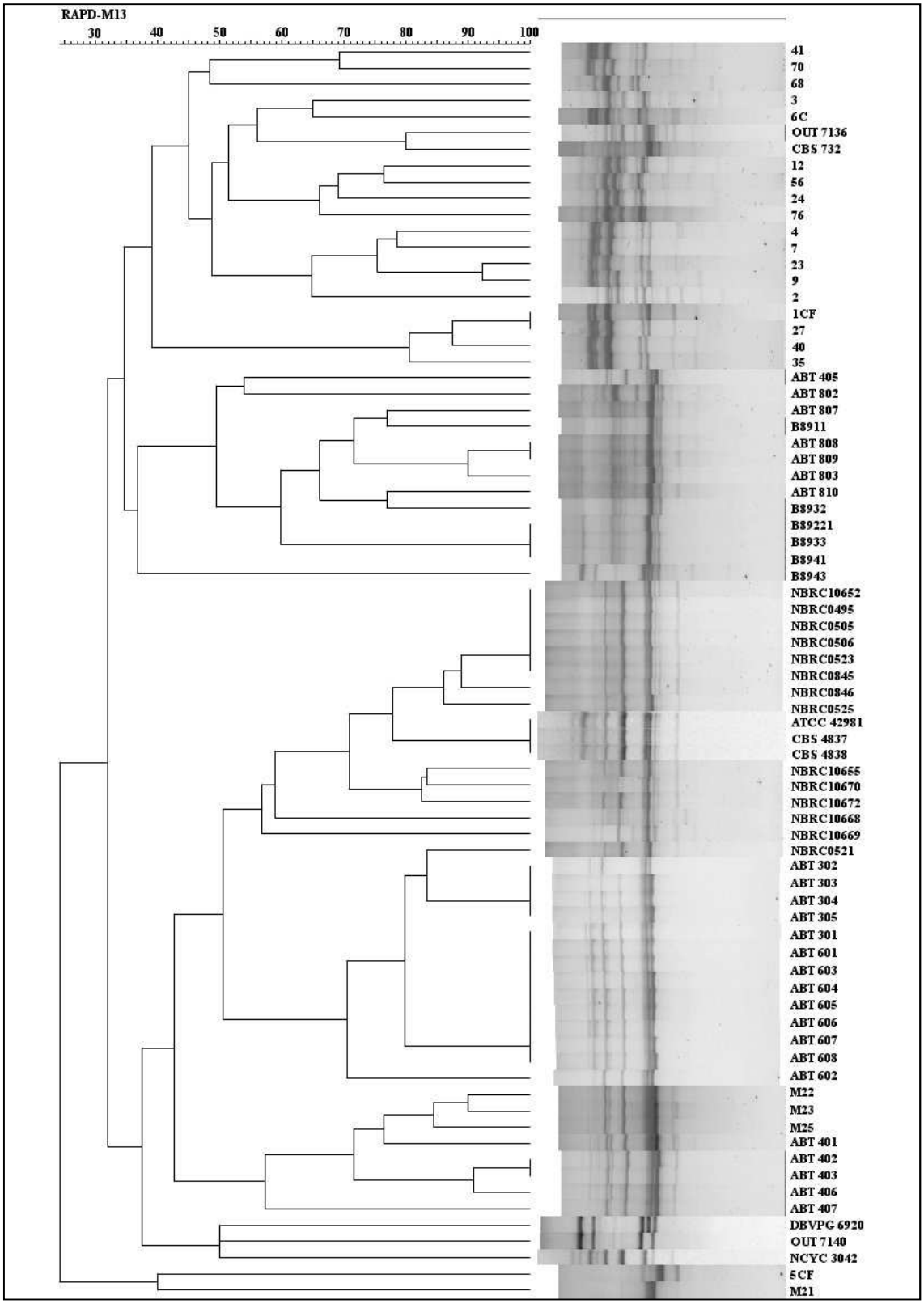


Figure 6.3. Jaccard's coefficient based UPGMA dendrogram for *Zygosaccharomyces* strains (M13-RAPD-PCR).

Robustness of the tree was estimated with bootstrap values on 1000 replicates indicated as a percentage.

We also computed the relationships among the yeast strains based on their RAPD profiles supported by band based method using Jaccard's coefficient (**Figure 6.3**). In comparison settings, we set optimization and band filtering tolerance at one percent. We found 100% similarity between strain NBRC0495, NBRC0505, NBRC0506, NBRC0523, NBRC10652 (cluster 2), three miso strains ATCC 42981, CBS 4837 and CBS 4838 (cluster 2), thirteen isolates out of fourteen representing putative *Z. sapae* (cluster 3), B8933, B8941, and B89221 (cluster 3 and 5), ABT 808 and ABT 809 (cluster 4), ABT 402 and ABT 403 (cluster 5), 27 and 1CF (cluster 8). The high degree of genetic similarity in these strains indicates the primary clonal nature of the population.

Our results indicated the potential of M13 primer for use in RAPD analysis would allow differentiation between *Z. bailii, Z. rouxii*, *Z. mellis* and *Z. sapae* at the species level and for differentiating and typing strains of the same species obtained different ecological niche. All these data suggest that the major monomorphic bands obtained by RAPD analysis are useful for the differentiation of *Zygosaccharomyces* species. Yeast also undergoes a number of changes in accordance with the environment. Altogether, it is clearly proved in the dendrogram that yeast isolated from different sources are very much different from each other and may represent different genetically divergent strains of the same species or different species.

6.3.3.2. Microsatellite primed $(GTG)_5$ PCR

Microsatellite primer $(GTG)_5$-PCR was performed for a total of 75 strains of *Zygosaccharomyces* genus. The resulting fingerprinting consisted of 8 to 17 electrophoretic bands with sizes ranging from 290 to 3940 bp (**Table 6.4**). By UPGMA analysis of $(GTG)_5$-fingerprinting, the strains clustered into seven clusters at similarity percent cut-off of 48.5 (cluster 1), 58.9 (cluster 2), 70.4 (cluster 3), 76.2 (cluster 4), 66.9 (cluster 5), 59.3 (cluster 6), and 30.4 (cluster 7) (**Figure 6.4**).

Clusters	Strains	MSP-GTG5 band profile
Cluster 1	ATCC 42981	2830, 2030, 1670, 1290, 1090, 870, 830, 790, 770, 600, 540, 310
	CBS 4837	2810, 2020, 1670, 1290, 1100, 880, 840, 790, 770, 600, 540, 310
	CBS 4838	2810, 2030, 1680, 1320, 1110, 890, 840, 790, 770, 610, 540, 310
	OUT 7136	2660, 1960, 1650, 1270, 1080, 870, 840, 800, 780, 700, 620, 560, 420, 330
	NBRC10668	3360, 2130, 1710, 1280, 1070, 960, 830, 800, 750, 560, 520, 300
	NBRC10669	3320, 2120, 1710, 1260, 1070, 840, 790, 760, 560, 520, 300
	76	3600, 2850, 2070, 1870, 1600, 1490, 1360, 1080, 840, 710, 570
	1CF	3360, 2130, 1710, 1280, 1070, 960, 830, 800, 750, 560, 520, 300
	6C	3320, 2120, 1710, 1260, 1070, 840, 790, 760, 560, 520, 300
Cluster 2	3	2610, 2100, 1740, 1590, 1390, 1260, 1220, 1040, 860, 800, 750, 650, 610, 360, 340
	4	1810, 1600, 1240, 1050, 860, 810, 670, 610
	7	2860, 2390, 1790, 1590, 1240, 1040, 870, 810, 750, 660, 620
	9	2620, 2290, 1780, 1580, 1240, 1040, 870, 810, 750, 620
	12	2620, 2150, 1790, 1600, 1390, 1060, 880, 810, 750, 660, 620
	23	2440, 1830, 1620, 1250, 1050, 880, 820, 660, 620, 450
	24	2490, 2080, 1710, 1570, 1390, 11220, 1050, 870, 810, 760, 660, 620
	27	2730, 2350, 1760, 1230, 1040, 900, 880, 850, 720, 620, 560
	35	1810, 1250, 1050, 910, 880, 850, 720, 630, 570
	40	1810, 1240, 1050, 910, 880, 850, 720, 630, 570, 300
	41	1820, 1610, 1530, 1260, 1070, 870, 820, 780, 670, 630
	56	2580, 2100, 1790, 1620, 1520, 1400, 1070, 880, 830, 760, 670, 630, 520
	68	2610, 2140, 1800, 1630, 1560, 1400, 1070, 870, 830, 770, 670, 630
	70	1790, 1590, 1250, 1060, 890, 830, 770, 630
Cluster 3	CBS 732	2500, 1920, 1580, 1240, 1060, 920, 880, 830, 780, 760, 680, 620, 550, 460, 390, 320
	301	2790, 2030, 1590, 1240, 1070, 910, 860, 820, 300
	302	2790, 2030, 1590, 1240, 1080, 920, 860, 830, 310
	303	2810, 2050, 1600, 1250, 1080, 920, 870, 830, 310
	304	2830, 2050, 1600, 1260, 1090, 920, 870, 830, 320
	305	2810, 2070, 1590, 1260, 1080, 920, 870, 830, 320
	306	2760, 2030, 1600, 1250, 1080, 920, 870, 830, 320
	601	2710, 1940, 1590, 1230, 1060, 910, 860, 820, 320
	602	2680, 1970, 1570, 1220, 1060, 910, 860, 820, 320
	603	2710, 1990, 1570, 1220, 1060, 910, 860, 820, 320
	604	2650, 1950, 1560, 1210, 1060, 910, 860, 820, 310
	605	2630, 1930, 1570, 1210, 1050, 910, 850, 810, 310
	606	2590, 1930, 1560, 1210, 1050, 900, 850, 810, 310
	607	2620, 1940, 1560, 1210, 1050, 900, 850, 810, 310
	608	2570, 1930, 1550, 1210, 1040, 900, 840, 810, 310
	B8941	1970, 1620, 1270, 1080, 900, 810, 600, 540, 440, 310
	B89221	2490, 1840, 1540, 1230, 1060, 920, 880, 830, 780, 760, 730, 700, 610, 550, 460, 390, 320

Clusters	Strains	MSP-GTG5 band profile
Cluster 4	NBRC0495	3250, 2040, 1630, 1230, 1040, 830, 790, 750, 560, 510, 300
	NBRC0505	3360, 2030, 1680, 1240, 1050, 830, 790, 750, 550, 510, 300
	NBRC0506	3220, 2030, 1650, 1230, 1050, 830, 760, 560, 300
	NBRC0521	3150, 2000, 1610, 1220, 1040, 830, 790, 750, 300
	NBRC0523	3100, 1990, 1620, 1230, 1040, 870, 830, 790, 750, 560, 520, 300
	NBRC0525	3050, 1970, 1600, 1230, 1040, 830, 790, 750, 550, 300
	NBRC0845	3100, 2000, 1640, 1230, 1040, 870, 830, 790, 740, 570, 510, 300
	NBRC0846	3120, 2000, 1640, 1220, 1030, 870, 780, 740, 570, 510, 300
	NBRC10652	2900, 1970, 1590, 1205, 1030, 830, 790, 750, 550, 510, 300
	NBRC10655	3000, 1980, 1590, 1200, 1030, 840, 790, 750, 560, 520, 300
	NBRC10670	3100, 2020, 1620, 1220, 1050, 840, 800, 760, 560, 520, 310
	NBRC10672	2960, 1940, 1600, 1200, 1030, 840, 790, 530, 560, 520, 310
	NCYC 3042	1650, 1230, 1050, 860, 780, 590, 310
Cluster 5	401	2370, 1840, 1510, 1150, 980, 840, 790, 760, 730, 710, 500, 300
	402	2380, 1760, 1490, 1150, 980, 920, 850, 800, 760, 730, 710, 500, 300
	403	2420, 1870, 1500, 1170, 990, 930, 850, 800, 770, 740, 720, 500, 470, 300
	405	2470, 1880, 1500, 1200, 1010, 940, 860, 810, 740, 730, 510, 300
	406	2530, 1900, 1540, 1190, 1010, 870, 810, 780, 750, 730, 510, 300
	407	2580, 1560, 1200, 1020, 870, 820, 790, 750, 740, 510, 300
	B8911	2430, 1800, 1500, 1190, 1010, 860, 790, 720, 580, 520, 420, 360, 290
	B8932	2510, 1820, 1520, 1200, 1030, 870, 800, 730, 580, 520, 300
	B8933	2440, 1820, 1510, 1210, 1030, 890, 810, 740, 670, 590, 540, 440, 370, 300
Cluster 6	802	3030, 2060, 1680, 1320, 1110, 940, 630, 570, 320
	803	2970, 2040, 1660, 1320, 1110, 940, 620, 550, 320
	807	3080, 2090, 1710, 1340, 1110, 940, 620, 550, 310
	808	3140, 2100, 1670, 1340, 1120, 940, 620, 560, 310
	809	3080, 2080, 1690, 1340, 1120, 930, 610, 550, 310
	810	1090, 900, 700, 580, 530, 310
	M21	2920, 2150, 1680, 1270, 1090, 930, 870, 800, 620, 560, 320
	M22	2920, 2140, 1740, 1290, 1100, 940, 890, 840, 630, 560, 520, 320
	M23	2960, 2130, 1690, 1340, 1120, 960, 890, 850, 630, 320
	M25	2960, 2130, 1690, 1340, 1120, 960, 890, 850, 630, 320
Cluster 7	OUT 7140	2970, 2300, 1180, 1100, 820, 690
	DBVPG 6920	3450, 2300, 1770, 1450, 1240, 1100, 920, 800, 730, 680
Outgroup	B8943	2950, 2570, 2050, 1410, 1280, 1220, 1070, 930, 890, 800, 690, 620, 500, 440, 410
	2	3310, 2600, 1650, 1370, 1160, 1070, 820, 790, 660, 620, 550, 470
	5CF	3940, 2240, 1840, 1070, 780, 570, 490, 410

Table 6.4. MSP-PCR fingerprint profile obtained from different *Zygosaccharomyces* strains with GTG_5 primer.

Cluster 1 comprises of strains isolated from both salt-rich habitats and honey as mentioned earlier. Allodiploid strain ATCC 42981, and aneuploid strains CBS 4837, CBS 4838 and OUT 7136 were also housed in cluster 1. This is congruent with our previous results, with showed that they are similar with respect to their genome content (ranging between 21.7±0.33 to 22.5±0.20) and number of

chromosome (n=8) (Solieri et al., 2013b). Cluster 2 comprises of *Z. mellis* isolates obtained from honey; while cluster 4 comprises mainly of strains from salt-rich habitats. Cluster 3, 5, and 6 accommodated different strains belonging to *Z. rouxii*, *Z. mellis* and 14 isolates of *Z. sapae* species, all recovered from TBV. In particular all fourteen TBV isolates belonging to cluster 3 possessed eight bands with almost similar band patterns and sizes. The approximate size of bands were 2790, 2030, 1590, 1240, 1070, 910, 860, 820, 300 bp. All isolates grouped into a homogeneous cluster well separated from B8941, B89221 and the type strain of *Z. rouxii* CBS 732^T that clustered together with the fourteen TBV isolates. At a similarity threshold of 70.4 %, the TBV isolates could be further subdivided into two subclusters. These subclusters are related with the source of isolation of these TBV isolates collected from sample A and B during a survey conducted in order to explore the indigenous yeast population of TBV (Solieri et al., 2006). Type strain of *Z. rouxii* CBS 732^T clustered together with cluster 3; while postulated hybrid strain NCYC 3042 within cluster 4. Cluster 7 contains two *Z. bailii* strains, DBVPG 6920 and OUT 7140. Strain B8943 and strain 2 did not fall within any of the cluster. Strain 5CF recovered from honey was characterized as *D. hansenii*, which served as an outgroup for MSP-PCR genotyping analysis (**Figure 6.4**).

Like conflicting result obtained from RAPD-PCR analysis for strain M21, M22, M23 and M25, the same discrepancy has been found with MSP-PCR analysis also. Here these strains get clustered along with *Z. rouxii* strain isolates of TBV samples (cluster 6). Similarly strain B8911, B8932 and B8933 that have been characterized as *Z. rouxii* strains based on 5.8S ITS-RFLP analysis get clustered along with *Z. mellis* strains in cluster 5. However, the failure of primers $(GTG)_5$ to discriminate *Z. sapae*, *Z. mellis* and *Z. rouxii* strains is evidence of their close phylogenetic proximity. As exhibited in the dendrograms obtained from the band patterns (**Figure 6.4**), the similarity among strains of the different species was always lower than 75.0%.

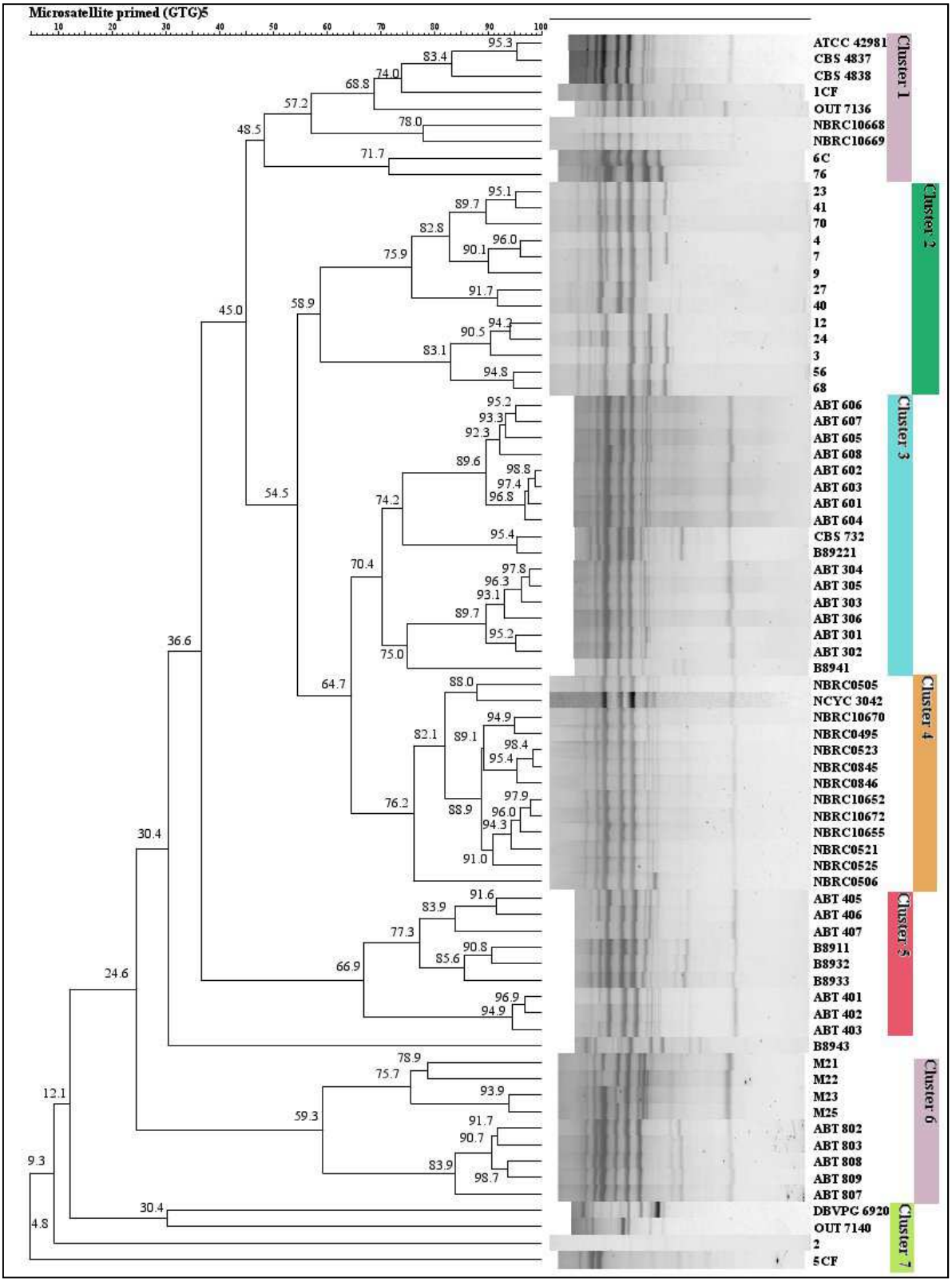


Figure 6.4. Pearson's coefficient based UPGMA dendrogram for *Zygosaccharomyces* strains (MSP-PCR). Robustness of the tree was estimated with bootstrap values on 1000 replicates indicated as a percentage.

6.3.3.3. IGS Genotyping

IGS regions were PCR-amplified with CNL12 and CNS1 primers, which span from base position 3046–3064 on the 26SrDNA of *S. cerevisiae* strain NRRL Y-12632 (GenBank Accession number-AY048154) to base position 37-17 on the 18S rDNA (GenBank Accession number-J01353). Previous work demonstrated that strains of the same *Zygosaccharomyces* species have IGS region identical in length (Wrent et al., 2010). *Zygosaccharomyces* strains considered here displayed IGS PCR amplicons of 4200 and 6400 bp, respectively (**Table 6.5**). We, using the protocol mentioned in Materials and methods, were able to amplify most of the *Zygosaccharomyces* strains (few *Z. mellis* honey isolates, *Z. mellis* TBV isolates, *Z. rouxii* TBV isolates, *Z. sapae* TBV isolates, *Z. bailii* strains, *Z. rouxii* CBS 732 and mosaic strains ATC 42981, CBS 4837, CBS 4838, OUT 7136, and NYCY 3042). The strains left unamplified include, seven honey isolates (2008 sampling) that represent Cluster 7 of Pearson's coefficient based UPGMA tree of M13-RAPD profile (**refer Figure 6.2**), five honey isolates of *Z. sapae* (1990 sampling), five TBV isolates of *Z. rouxii* plus strain uncharacterized B8943 (1989 sampling), and all fourteen strains procured from NBRC culture collection. For strains that were unable to amplify, we tested different Buffer (GC Buffer II) with concurrent decrease in the annealing temperature from 55 to 52°C and increasing extension time from 3 min to 5 min. Amplification of five honey isolates of *Z. sapae* (1990 sampling) and five TBV isolates of *Z. rouxii* (1989 sampling) was successful using the standard IGS-PCR protocol with GC Buffer I, however, by increasing the extension time from 3 min to 5 min. Seven honey isolates (2008 sampling) did not amplify with either by increasing extension time or decreasing annealing temperature in both GC Buffer I and GC Buffer II based standard PCR mix. However, a faint band for strain 2 was seen when GC Buffer I based standard PCR mix was used and the annealing temperature was decreased from 55 to 52°C.

The number of patterns varies depending on the endonuclease used. The digestion of the IGS amplicons with *Hap*II endonuclease yielded 26 different profiles, while 22 different profiles were obtained with both *Hha*I and *Mbo*I, respectively (**Table 6.5 and Table 6.6**). It was fairly difficult to discriminate strains and species using a single endonuclease. In general, *Zygosaccharomyces* strains were best discriminated using *Hap*II endonuclease (**Table 6.5**). However, in case of *Z. rouxii* the best discrimination was obtained using the *Mbo*I endonuclease (**Table 6.5**). The restriction patterns obtained with all three endonucleases, as illustrated in **Table 6.6**, enable us to differentiate all the tested *Zygosaccharomyces* strains at species level by examining the pool of species-specific conserved bands. We observed a high degree of polymorphism for each species with the three enzymes assayed (**Table 6.4 and Table 6.6**). The potential of an endonuclease to discriminate strains belonging to a species have shown, to some extent, relevance with their source of isolation. For example, the strains of *Z. mellis* obtained from TBV have shown no shared IGS restriction profile with strains that have been isolated from honey, except ABT401. In contrast, strains of a species isolated from same source environment showed high degree of genetic similarity as they possessed high number of shared bands. Fourteen putative *Z. sapae* isolates recovered from two different samples revealed two, three and five profiles respectively with *Mbo*I, *Hha*I and *Hap*II.

Species and Strains studied	Restriction fragments (bp)		
	Profiles *Hap*II (A)	Profiles *Hha*I (B)	Profiles *Mbo*I (C)
***S. cerevisiae* (S)**			
ABT 507	SA1-750, 700, 600, 350	SB1-3000, 800, 250, 80	SC1-1600, 650, 630, 460, 270, 150
***Z. bailii* (B)**			
OUT 7140	BA1-1050, 950, 740, 540, 440, 230, 190, 170, 150, 130, 110	BB1-1700, 620, 440, 420, 350, 330, 300, 200, 180	BC1-2200, 900, 650, 410, 330, 260, 160, 70
DBVPG 6920	BA2-1500, 1400, 1200, 1100, 620, 600, 550, 480, 460, 340, 300, 230, 180, 170	BB2- 2200, 1700, 680, 650, 490, 420, 340, 250	BC2-1200, 1000, 630, 510, 480, 270, 170
***Z. mellis* (M)**			
ABT401, ABT402, ABT403	MA1-1000, 930, 480, 460, 320, 300, 280, 190, 170, 120, 90	MB1-1600, 800, 510, 380, 370, 270	MC1-1300, 900, 800, 430, 350, 270, 160, 120, 100
ABT405, ABT406	MA1-1000, 930, 480, 460, 320, 300, 280, 190, 170, 120, 90	MB1-1600, 800, 510, 380, 370, 270	MC2-1350, 920, 820, 430, 350, 270, 160, 130, 100
ABT407	MA1-1000, 930, 480, 460, 320, 300, 280, 190, 170, 120, 90	MB1-1600, 800, 510, 380, 370, 270	MC3-1300, 900, 820, 430, 350, 270, 160, 130, 100
2	MA2-1050, 750, 500, 380, 320, 230, 190, 170, 120, 90	MB2-1100, 800, 750, 650, 460, 440	-
4	MA3-900, 320, 290, 190, 170, 140	MB3-1700, 900, 800, 390, 270	MC1-1300, 900, 800, 430, 350, 270, 160, 120, 100
7, 23	MA4-900, 320, 280, 190, 170, 140	MB3-1700, 900, 800, 390, 270	MC4-1300, 900, 800, 530, 350, 270, 160, 130, 100
9	MA5-1000, 850, 750, 310, 290, 230, 200, 180, 170	MB4-1250, 550, 520, 440, 390, 340, 240	MC5-1250, 1050, 480, 360, 340, 270, 250, 180, 110
27	MA6-1050, 950, 310, 290, 190, 170, 140, 120	MB5-1700, 850, 520, 390, 380, 270	MC1-1300, 900, 800, 430, 350, 270, 160, 120, 100
35, 40	MA7-1200, 1000, 310, 290, 190, 170, 140, 120	MB6-1700, 900, 770, 390, 270	MC6-1250, 850, 750, 530, 350, 270, 160, 130, 100
41	MA4-900, 320, 280, 190, 170, 140	MB5-1700, 850, 520, 390, 380, 270	MC1-1300, 900, 800, 430, 350, 270, 160, 120, 100
70	MA8-1050, 880, 750, 380, 290, 230, 190, 170, 140, 120	MB4-1250, 550, 520, 440, 390, 340, 240	MC5-1250, 1050, 480, 360, 340, 270, 250, 180, 110
1CF	MA7-1200, 1000, 310, 290, 190, 170, 140, 120	MB6-1700, 900, 770, 390, 270	MC6-1250, 850, 750, 530, 350, 270, 160, 130, 100
6C	MA4-900, 320, 280, 190, 170, 140	MB3-1700, 900, 800, 390, 270	MC1-1300, 900, 800, 430, 350, 270, 160, 120, 100
***Z. pseudorouxii* (P)**			
NCYC 3042	PA1-1000, 950, 900, 330, 310, 190, 170	PB1-1550, 900, 850, 520, 390, 380, 260, 150	PC1-1300, 890, 850, 770, 430, 350, 270, 170, 150
***Z. rouxii* (R)**			
CBS 732^T	RA1-1000, 710, 510, 400, 330, 310, 290, 220, 190, 170, 150, 120, 100	RB1-980, 700, 660, 640, 460, 440, 220, 200	RC1-770, 700, 680, 530, 500, 430, 350, 300, 270, 170, 80
ATCC 42981	RA2-1400, 730, 720, 420, 380, 340, 230, 190, 170, 100	RB2-1500, 900, 680, 400, 360, 350, 200, 160	RC2-1700, 800, 680, 520, 420, 270, 170, 150
CBS 4837, CBS 4838	RA2-1400, 730, 720, 420, 380, 340, 230, 190, 170, 100	RB3-1500, 900, 880, 680, 400, 360, 350, 200, 160	RC3-1700, 800, 770, 680, 520, 420, 270, 170, 150

OUT 7136	RA2-1400, 730, 720, 420, 380, 340, 230, 190, 170, 100	RB4-1600, 880, 680, 400, 350, 200, 160	RC4-1900, 770, 730, 520, 420, 270, 170, 150
ABT802	RA3-1000, 710, 510, 490, 400, 390, 310, 290, 170, 150, 120, 100	RB5-980, 770, 750, 700, 640, 460, 440, 250, 200, 190	RC5-1200, 780, 550, 530, 450, 360, 260, 170, 80
ABT803, ABT808	RA4-1000, 710, 510, 420, 400, 310, 290, 190, 170, 150, 120, 100	RB6-980, 770, 700, 640, 460, 440, 200	RC6-1200, 690, 550, 530, 380, 360, 260, 170, 80
ABT807, ABT809, ABT810	RA4-1000, 710, 510, 420, 400, 310, 290, 190, 170, 150, 120, 100	RB6-980, 770, 700, 640, 460, 440, 200	RC6-1200, 690, 550, 530, 380, 360, 260, 170, 80
B8911	RA5-1050, 730, 520, 500, 400, 390, 310, 290, 280, 230, 170, 150, 100	RB7-1000, 760, 720, 640, 620, 460, 440, 230, 200	RC7-1000, 650, 460, 410, 360, 310, 300, 260, 170, 80
B8932, B8941	RA5-1050, 730, 520, 500, 400, 390, 310, 290, 280, 230, 170, 150, 100	RB7-1000, 760, 720, 640, 620, 460, 440, 230, 200	RC8-1100, 680, 510, 500, 360, 300, 260, 170, 80
B8933	RA6-1050, 730, 510, 490, 400, 310, 290, 280, 170, 150, 100	RB8-1000, 700, 670, 640, 620, 460, 440, 210, 190	RC9-730, 510, 500, 410, 360, 310, 260, 170, 80
B89221	RA7-1050, 730, 490, 400, 310, 280, 170, 150, 100	RB8-1000, 700, 670, 640, 620, 460, 440, 210, 190	RC9-730, 510, 500, 410, 360, 310, 260, 170, 80
***Z. sapae* (S)** **ABT301^T**	SA1-1050, 900, 720, 480, 400, 340, 320, 270, 210, 180, 170	SB1-1700, 980, 700, 690, 620, 520, 460, 450, 390, 380, 260, 210, 180	SC1-1100, 800, 770, 530, 430, 350, 270, 170, 150, 120
ABT302, ABT303	SA2-1050, 900, 720, 480, 400, 380, 320, 270, 180, 170	SB2-1700, 980, 720, 700, 620, 520, 460, 450, 390, 380, 180	SC2-1300, 890, 770, 530, 430, 350, 270
ABT304, ABT305, ABT306	SA3-1050, 720, 480, 400, 380, 320, 270, 180, 170	SB2-1700, 980, 720, 700, 620, 520, 460, 450, 390, 380, 180	SC2-1300, 890, 770, 530, 430, 350, 270
ABT601, ABT604, ABT606	SA4-1050, 900, 850, 720, 480, 400, 380, 320, 280, 190, 170	SB3-1700, 980, 970, 720, 700, 620, 520, 460, 450, 390, 380, 260, 180	SC2-1300, 890, 770, 530, 430, 350, 270
ABT602, ABT603, ABT605	SA5-1050, 900, 720, 480, 400, 380, 320, 280, 190, 170	SB3-1700, 980, 970, 720, 700, 620, 520, 460, 450, 390, 380, 260, 180	SC2-1300, 890, 770, 530, 430, 350, 270
ABT607	SA4-1050, 900, 850, 720, 480, 400, 380, 320, 280, 190, 170	SB3-1700, 980, 720, 700, 620, 520, 460, 450, 390, 380, 260, 180	SC2-1300, 890, 770, 530, 430, 350, 270
ABT608	SA5-1050, 900, 720, 480, 400, 380, 320, 280, 190, 170	SB3-1700, 980, 720, 700, 620, 520, 460, 450, 390, 380, 260, 180	SC2-1300, 890, 770, 530, 430, 350, 270
M21, M25	SA6-1200, 1100, 950, 750, 650, 500, 440, 320, 290, 190, 170, 140, 120	SB4-1700, 1000, 680, 650, 520, 470, 450, 410, 390, 290	SC3-1450, 980, 850, 700, 560, 450, 360, 270
M22, M23	SA7-1100, 950, 750, 650, 500, 440, 320, 290, 190, 170, 140, 120	SB4-1700, 1000, 680, 650, 520, 470, 450, 410, 390, 290	-
M24	-	SB4-1700, 1000, 680, 650, 520, 470, 450, 410, 390, 290	-

Table 6.5. IGS-RFLP profile obtained for different *Zygosaccharomyces* strains with *Hap*II, *Hha*I and *Mbo*I endonucleases.

Species and Strains studied	Source of isolation	IGS band size	Restriction fragments (bp)		
			Profiles *Hap*II (A)	Profiles *Hha*I (B)	Profiles *Mbo*I (C)
***S. cerevisiae* (S)** **ABT 507**	-	-	SA1	SB1	SC1
***Z. bailii* (B)** **OUT 7140**	Unknown	6300	BA1	BB1	BC1
DBVPG 6920	Salad-dressing	6300	BA2	BB2	BC2
***Z. mellis* (M)** **ABT401, ABT402, ABT403**	TBV	4200	MA1	MB1	MC1
ABT405, ABT406	TBV	4200	MA1	MB1	MC2
ABT407	TBV	4200	MA1	MB1	MC3
2	Honey	4200	MA2	MB2	-
4	Honey	4200	MA3	MB3	MC1
7, 23	Honey	4200	MA4	MB3	MC4
9	Honey	4200	MA5	MB4	MC5
27	Honey	4200	MA6	MB5	MC1
35, 40	Honey	4200	MA7	MB6	MC6
41	Honey	4200	MA4	MB5	MC1
70	Honey	4200	MA8	MB4	MC5
1CF	Honey	4200	MA7	MB6	MC6
6C	Honey	4200	MA4	MB3	MC1
***Z. pseudorouxii* (P)** **NCYC 3042**	Sugar	4200	PA1	PB1	PC1
***Z. rouxii* (R)** **CBS 732^T**	Cooked Must	4200	RA1	RB1	RC1
ATCC 42981	Miso	4200	RA2	RB2	RC2
CBS 4837, CBS 4838	Miso	4200	RA2	RB3	RC3
OUT 7136	Soy-moromi	4200	RA2	RB4	RC4
ABT802	TBV	4200	RA3	RB5	RC5
ABT803, ABT808	TBV	4200	RA4	RB6	RC6

ABT807, ABT809, ABT810	TBV	4200	RA4	RB6	RC6
B8911	TBV	4200	RA5	RB7	RC7
B8932, B8941	TBV	4200	RA5	RB7	RC8
B8933	TBV	4200	RA6	RB8	RC9
B89221	TBV	4200	RA7	RB8	RC9
***Z. sapae* (S)** **ABT301^T**	TBV	4200	SA1	SB1	SC1
ABT302, ABT303	TBV	4200	SA2	SB2	SC2
ABT304, ABT305, ABT306	TBV	4200	SA3	SB2	SC2
ABT601, ABT604, ABT606	TBV	4200	SA4	SB3	SC2
ABT602, ABT603, ABT605	TBV	4200	SA5	SB3	SC2
ABT607	TBV	4200	SA4	SB3	SC2
ABT608	TBV	4200	SA5	SB3	SC2
M21, M25	TBV	4200	SA6	SB4	SC3
M22, M23	TBV	4200	SA7	SB4	-
M24	TBV	4200	-	SB4	-

Table 6.6. Strains, their origin, source of isolation, size of the amplified IGS region, and code of the RFLP pattern (obtained with *Hap*II, *Hha*I and *Mbo*I endonucleases).

Two major reasons could be attributed to this cause, where the first could be the primary clonal nature of the population and second possibility could be that the strains arise from colonies of the same strain. The latter possibility has been explained by some other authors also (Wrent et al., 2010) and it has been clearly evident from profiles of *Z. rouxii* isolates from TBV samples, where some strains of a species isolated from different samples of TBV resulted in diverse banding profile.

When the strain patterns belonging to different species were analyzed, together with the dendrogram, some conflicting results were observed. Therefore, we assessed the reliability of clusters obtained from different clustering algorithms implemented in the analysis. For this, we created and compared the dendrogram for IGS-PCR fingerprints (obtained using all three enzymes) using both the curve-based and the band-based clustering algorithms.

6.3.1.4. Selection of best clustering algorithm for constructing dendrogram

When the strain IGS-PCR fingerprinting patterns were analyzed by Pearson's coefficient-based UPGMA dendrogram, some conflicting results were observed. Therefore, we created the dendrogram for IGS-PCR fingerprints (obtained using all three enzymes) using band-based clustering algorithms (**refer section 6.3.3.3**) and then compared it with the dendrogram created using curve-based clustering. From a wide range of clustering algorithms available under the category curve-based and band-based in Bionumerics module version 6.0, we used Pearson's (abbreviated as "PC") and ranked Pearson's (abbreviated as "RPC") correlation under former while Jaccard's (abbreviated as "JA") under the latter. For curve based methods, we adjusted curve smoothing (abbreviated as (CS) and optimization (abbreviated as "OP") to zero percent; while in case of band-based metrix, the optimization (abbreviated as "OP") and band tolerance (abbreviated as "BT") settings were set to one percent.

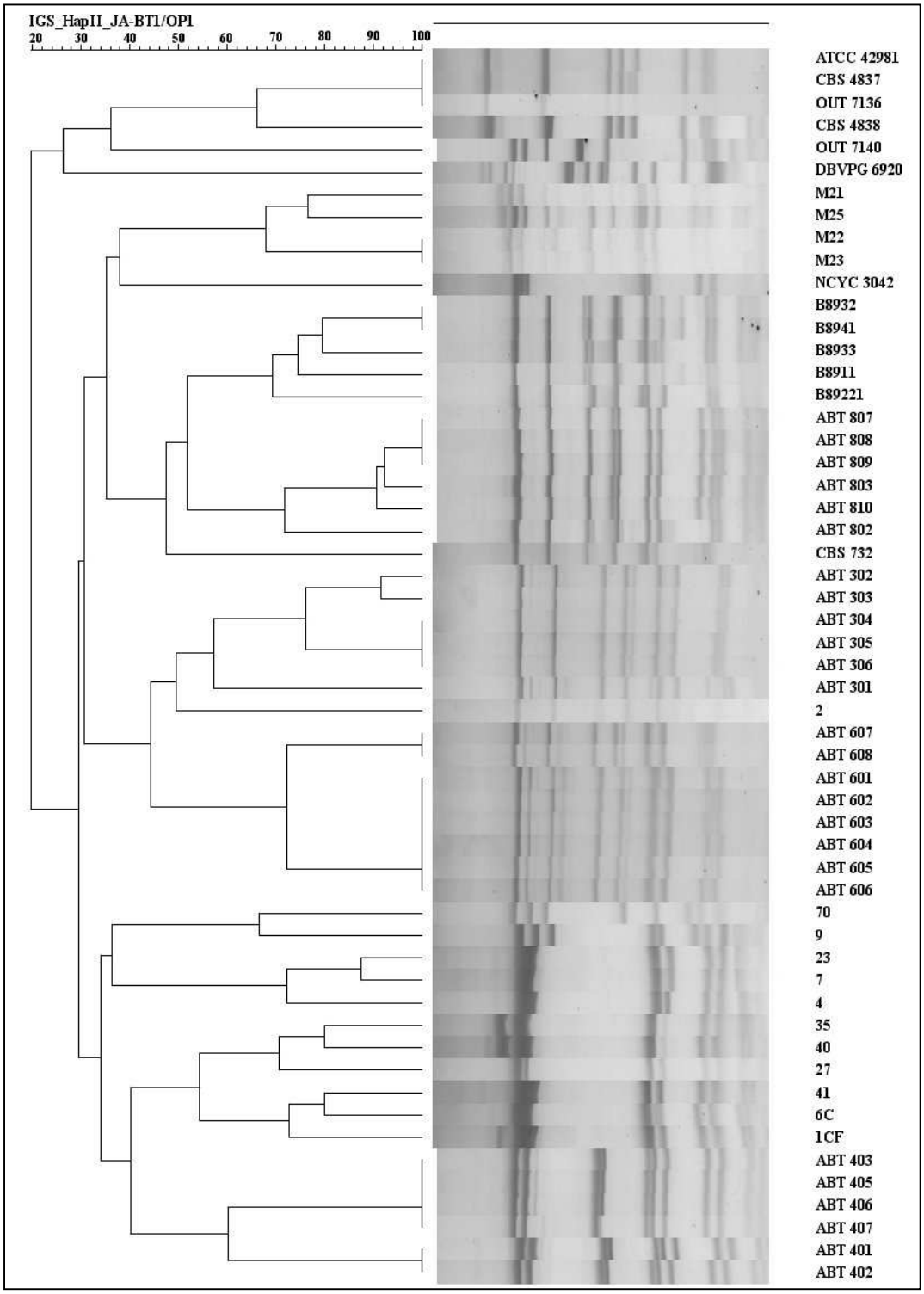


Figure 6.5. Jaccard's coefficient based UPGMA dendrogram for *Zygosaccharomyces* strains (*Hap*II IGS-PCR). Robustness of the tree was estimated with bootstrap values on 1000 replicates indicated as a percentage.

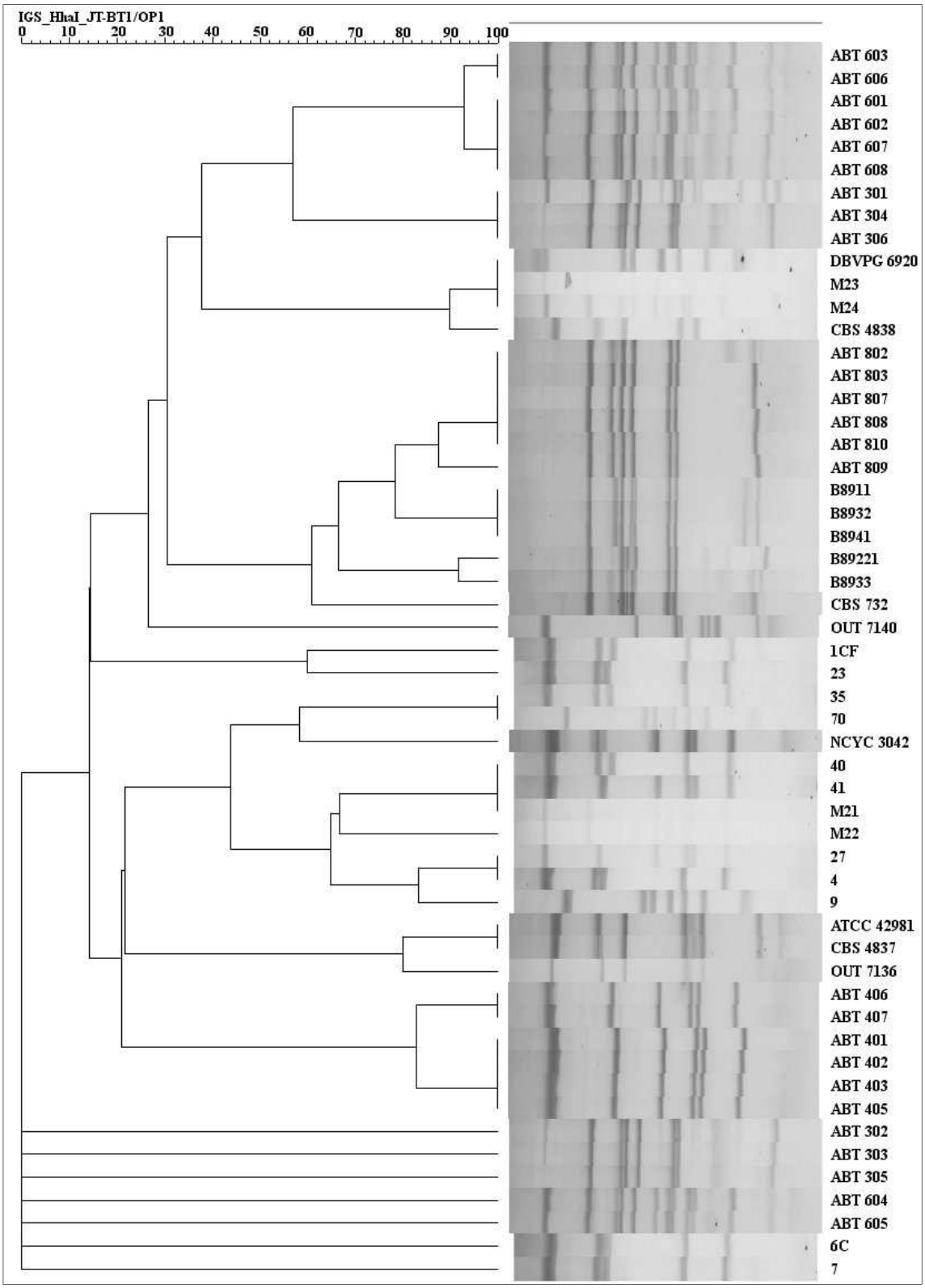


Figure 6.6. Jaccard's coefficient based UPGMA dendrogram for *Zygosaccharomyces* strains (*Hha*I IGS-PCR). Robustness of the tree was estimated with bootstrap values on 1000 replicates indicated as a percentage.

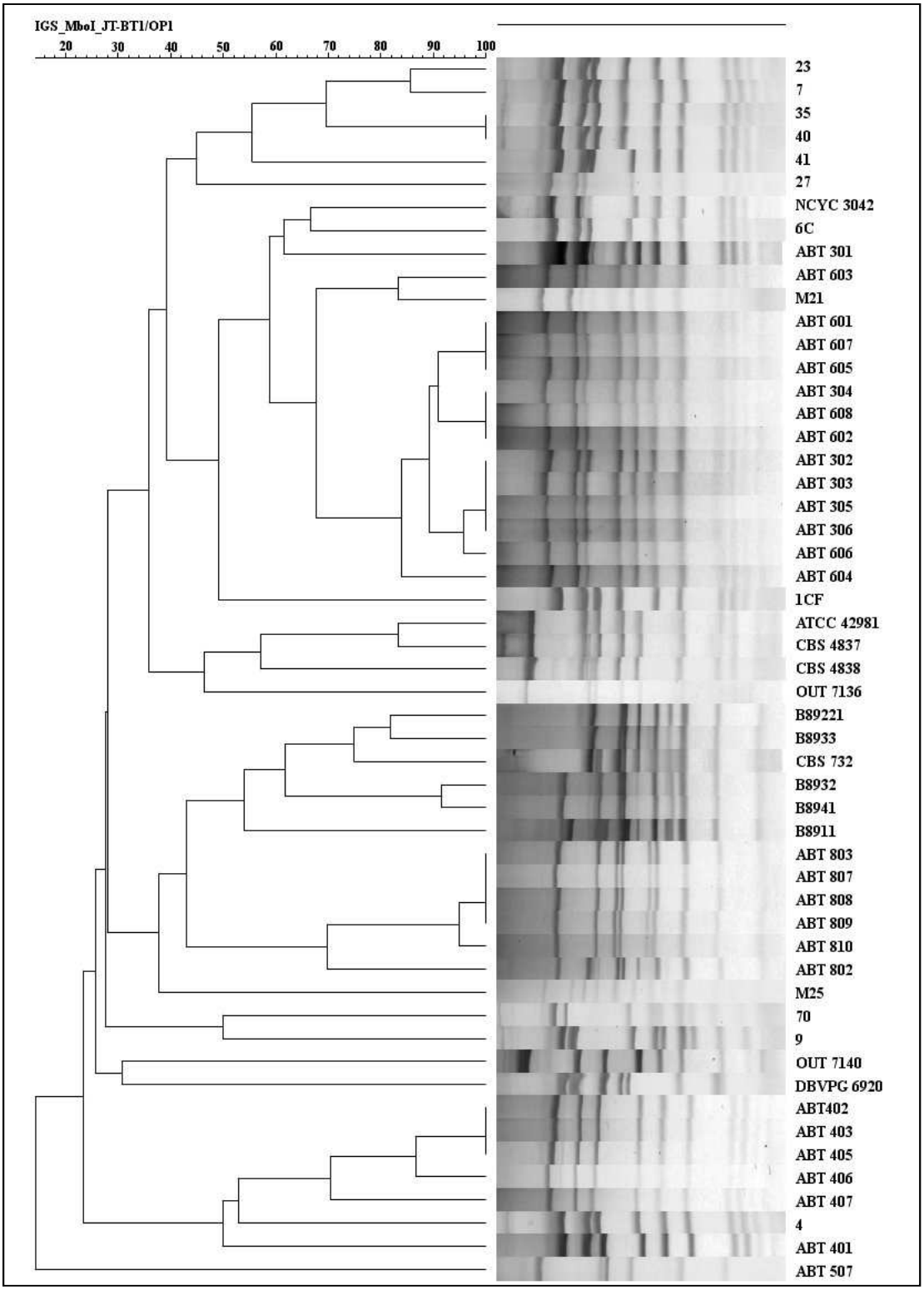


Figure 6.7. Jaccard's coefficient based UPGMA dendrogram for *Zygosaccharomyces* strains (*Mbo*I IGS-PCR). Robustness of the tree was estimated with bootstrap values on 1000 replicates indicated as a percentage. *Saccharomyces cerevisiae* ABT 507 was used as the outgroup species.

In *Hap*II PC-CS0/OP0 clustering, two *Z. bailii* strains OUT 7140 and DBVPG 6920 respectively appeared as an outgroup for the main clusters of *Z. rouxii* TBV isolates and *Z. rouxii* salty-niche strains (**Supplementary Figure S6.1**). The restriction profiles for these *Z. bailii* strain DBVPG 6920 appeared in this cluster showed only three common bands (1400, 340, and 230bp) with *Z. rouxii* miso and soy-moromi strains with *Hap*II endonucleases. While *Z. bailii* OUT 7140 strain that appeared as outgroup of *Z. rouxii* TBV strains showed no common bands with size more than 200bp. Both outgroup *Z. bailii* strains didn't cluster together as they showed completely diverse banding profiles with each other. However, both outgroup strains of *Z. bailii* clustered together when ranked Pearson's correlation based algorithm was used instead of Pearson's correlation based algorithm (data not shown); however, the comparison setting parameters were kept same while calculating both clustering comparisons. The type strain of *Z. rouxii* CBS 732^{T}, always appeared with the cluster of *Z. sapae* TBV isolates (either inside the main cluster or clustered) in the both curve-based approach used in this study; however clustering supported by band-based approach resolved this conflict and clustered the type strain of *Z. rouxii* CBS 732^{T}.with other *Z. rouxii* isolates from TBV samples (strain ABT605, ABT606, ABT607 and ABT608). The curve- based clustering of strain NCYC 3042 always showed inconsistency where the strain sometimes appeared inside the main cluster of *Z. mellis* honey isolates (in *Hap*II_PC-CS0/OP0) or cluster with M21, M22, M23, M25 strains ((in *Hap*II_RPC-CS0/OP0). It has also been found that some TBV isolates (strain ABT605, ABT606, ABT607 and ABT608) of putative *Z. sapae* did not appear in the main cluster that accommodated other putative *Z. sapae* isolates. However, in band-based Jaccard's supported dendrogram all the TBV isolates of putative *Z. sapae* appeared as defined cluster (**Figure 6.5**). In addition, this band–based clustering approach also resulted in the most well-defined separate clusters for strains or isolates of *Z. rouxii*, *Z. mellis* and *Z. sapae*. Our results indicated that band-based Jaccard's

clustering algorithm is adequately reliable in resolving various conflict and incongruent results we obtained with clustering algorithm that use curve-based coefficient.

In *Hha*I_PC-CS0/OP0 clustering, the major incongruence was regarding the placement of *Z. rouxii* miso strain CBS 4838 and strain NCYC 3042. The former strain based on fingerprints obtained from other restriction endonucleases always tends to cluster with its sister strain CBS 4837 that has also been isolated from same ecological environment. However, in *Hha*I PC-CS0/OP0 and *Hha*I_RPC-CS0/OP0 clustering it appeared in the main cluster of *Z. mellis* honey isolates. We ruled out that may be the clustering artifacts or the faint band intensity have resulted in incongruent placement of strain CBS 4838. Unlike in case of dendrogram obtained from fingerprints related to *Hap*II restriction endonucleases, the placement of strain NCYC 3042 showed reasonable congruent between banding size (as reported in **Table 6.5**) and dendrogram (**Supplementary Figure S6.2**) in *Hha*I related dendrogram. In case of *Hha*I, strain NCYC was found to be clustered *Z. rouxii* salty-niche strains using both curve-based clustering approaches. The NCYC 3042 shared five common bands (out of total eight bands obtained) with RFLP-fingerprints other *Z. rouxii* salty-niche strains. The shared band size were approximated as 1500, 900, 850, 390, and 150bp. However, by choosing band-based clustering approach, the strain NCYC 3042 get clustered inside the main cluster for *Z. mellis* honey isolates and showed high incongruence based on band size as it shares only one (900bp) or two bands (900 and 390bp) respectively with *Z. mellis* honey strains 35 and 70 that clustered with strain NCYC 3042 in *Hha*I_JT-BT1/OP1 clustering dendrogram (**Figure 6.6**).

Major incongurencies observed in *Mbo*I_PC-CS0/OP0 is related to placement of putative *Z. sapae* strains ABT301^{T} with *Z. mellis* honey isolates, and placement of two *Z. mellis* isolates, 27 (isolated from honey) and ABT407 (isolated from TBV), with putative *Z. sapae* isolates

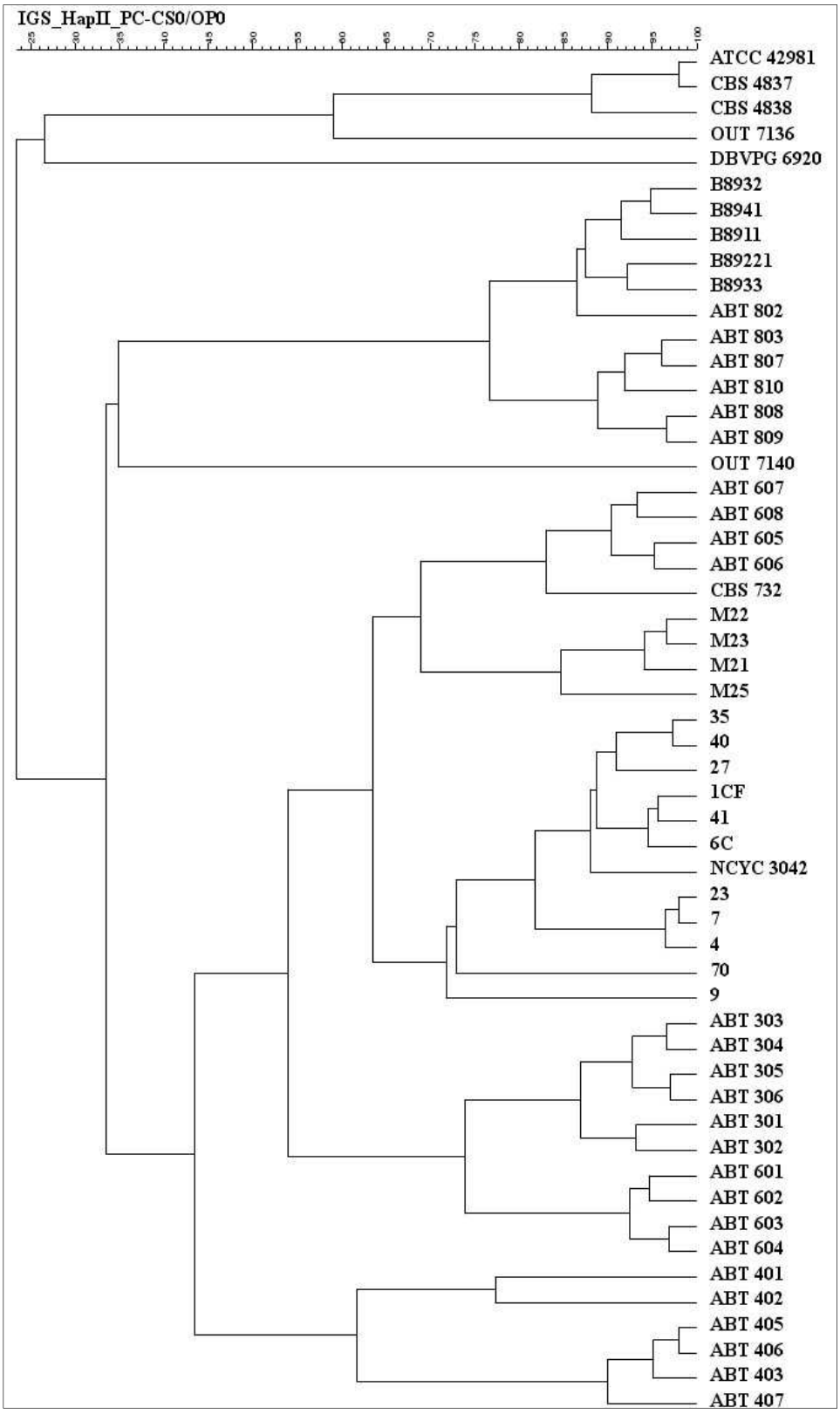
(**Supplementary Figure S6.3**). Besides this, in *Mbo*I related dendrogram obtained from PCR fingerprints supported by curve-based clustering algorithm, the strain OUT 7136 served as an outgroup instead of *S. cerevisiae* strain ABT507. However, band-based Jaccard's clustering algorithm the *S. cerevisiae* strain ABT507 was an outgroup strain. However, using band-based Jaccard's clustering approach, both incongruencies were resolved (**Figure 6.7**).

6.4. Conclusions

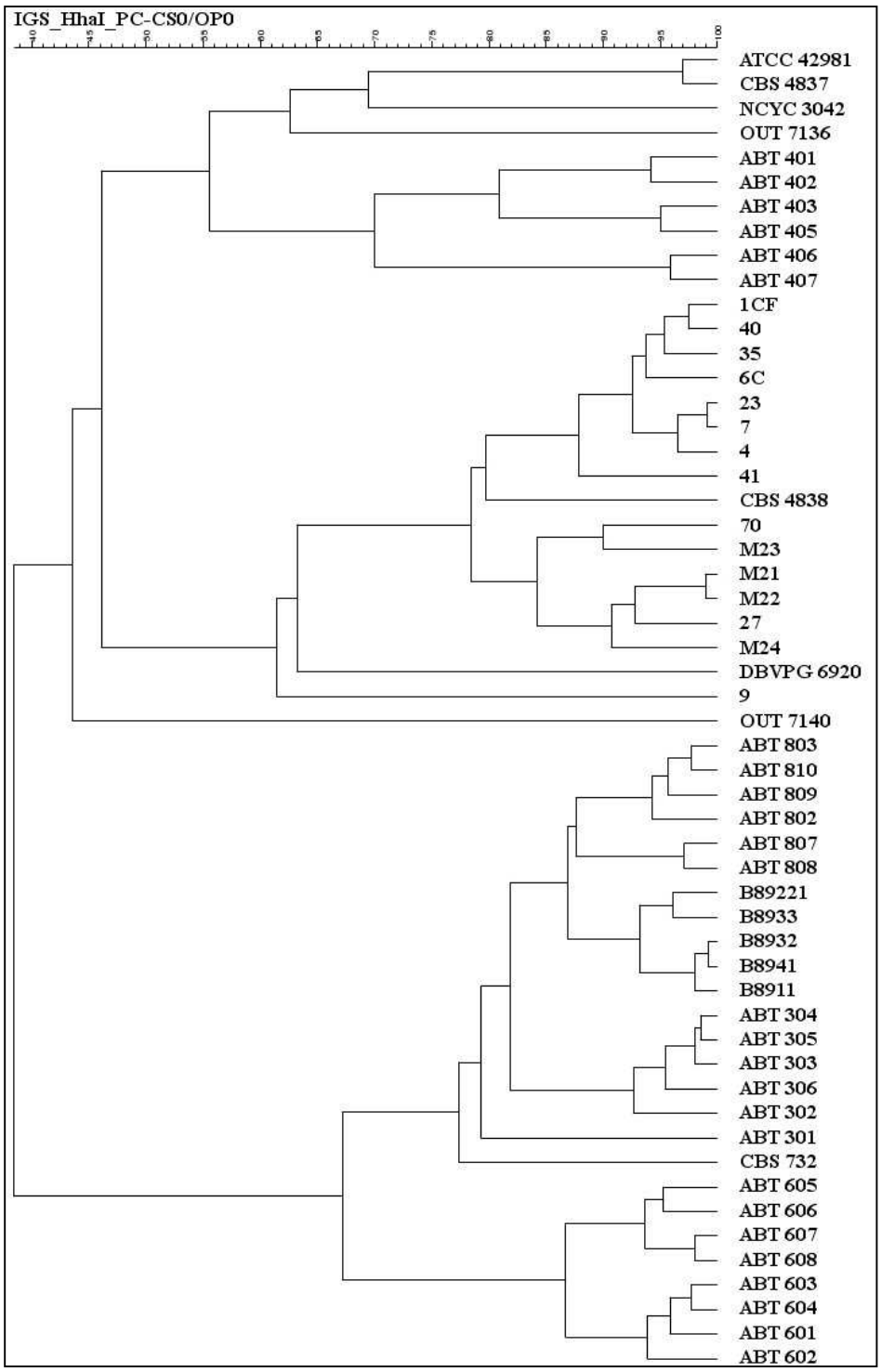
In food industries, strain typing can be considered from two perspectives: one for discriminating between different biotypes with specific properties, such as the production of volatile compounds for improving taste, ripening etc. (Andrade et al., 2009), as well as for monitoring the behaviour of a strain, for example, during the fermentation process (Suezawa et al., 2008). Besides this, another important reason for strain typing and is to ascertain the source of contamination in a food product in order to solve spoilage related problems associated with the production chain in a specific industrial process (Martorell et al., 2005).

Molecular typing based studies conducted in this study, to a large extent, were congruent with their assigned phylogenetic affiliation. Comparing the discriminatory power of the three techniques used, microsatellite primed-PCR and RAPD unveiled less variation than RFLP analysis of 5.8S ITS gene using *Hae*III, *Hha*I and *Hinf*I. Althought, using rDNA gene sequences (for instance, 26S rDNA D1/D2 domain) to identify yeasts is recognized; however, it has some limitations for the identification of postulated hybrids (James et al., 2005). During our study, we faced similar problem, where a pool of strains remained uncharacterized owing to wide spread heterogeneity in either 5.8S ITS region or 26S D1-D2 domain region or both. In order interpret the rDNA data obtained from these strains after sequencing and to assign these strains a proper

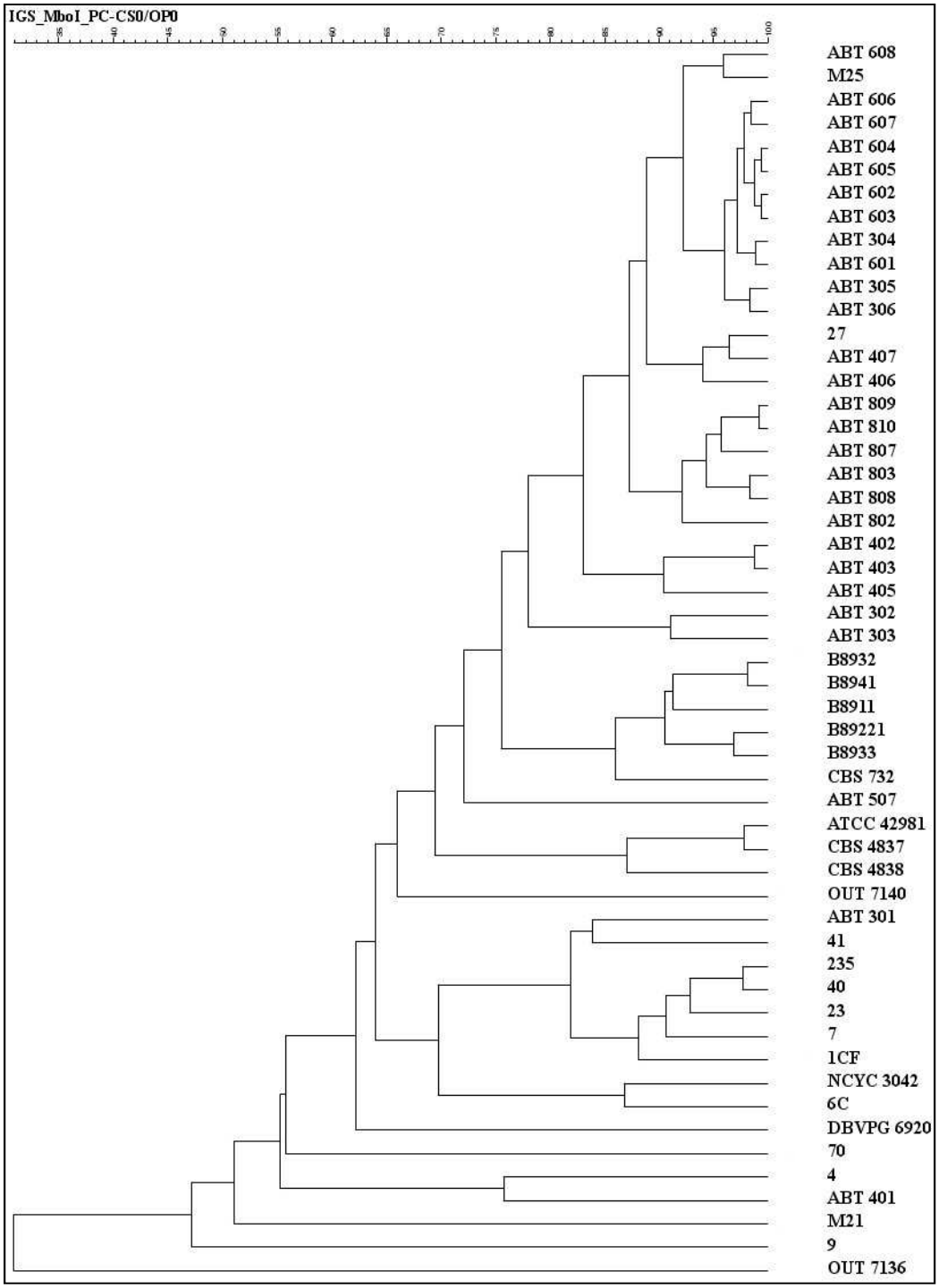
phylogenetic and taxonomic position, we conducted Southern blotting analysis and RNA structure analysis in these strains (**refer Chapter 7**).



Supplementary Figure S6.1. Pearson's coefficient based UPGMA dendrogram for *Zygosaccharomyces* strains (*Hap*II IGS-PCR). Robustness of the tree was estimated with bootstrap values on 1000 replicates indicated as a percentage.



Supplementary Figure S6.2. Pearson's coefficient based UPGMA dendrogram for *Zygosaccharomyces* strains (*Hha*I IGS-PCR). Robustness of the tree was estimated with bootstrap values on 1000 replicates indicated as a percentage.



Supplementary Figure S6.3. Pearson's coefficient based UPGMA dendrogram for *Zygosaccharomyces* strains (*Mbo*I IGS-PCR). Robustness of the tree was estimated with bootstrap values on 1000 replicates indicated as a percentage.

Chapter 7:

Extent of rDNA heterogeneity within individual genomes of *Zygosaccharomyces rouxii* complex strains

7.1. Introduction

Ribosomal DNA (rDNA) restriction analyses have provided much insight into the evolutionary biology of various organisms (Kurtzman and Robnett, 1998; Kurtzman et al., 2003; Sipiczki et al., 2013). The rDNA of yeasts is organized as large tandem arrays, which often complicates interpretation of rDNA divergence patterns in them (Proux-Wéra et al., 2013). Tandem repeated gene families are thought to undergo extensive homogenization through concerted evolution (Dover 1982; Ganley and Kobayashi, 2007). Evidence suggests that concerted evolution acts through unequal crossing over between repeating units, gene conversion or gene amplification. The homogenization of rDNA repeats is poorly understood, and its impact on the use of rDNA for phylogenetic inference is controversial. In fungi and yeasts, tandem arrays may contain 45–200 copies of the ribosomal region (Ganley and Kobayashi, 2007) and these copies may be distributed across one or more chromosomal locations (Ganley and Kobayashi, 2011; Proux-Wéra et al., 2013), thus allowing for significant variation within the genome of one individual. Intragenomic variation both in the ITS and 26S rDNA regions present significant challenges for phylogenetic analyses and species identification. Intragenomic variations are believed to be due to a relaxation of concerted evolution among the ribosomal DNA repeats in tandem arrays.

In a course of study aimed to unravel the widespread genomic complexity and diversity in *Z. rouxii* complex strains (**refer Chapter 3**; Solieri et al., 2013b), six strains ATCC 42981, CBS 4837 (=NCYC 1682), CBS 4838, OUT 7136, NCYC 3042 (*Z. pseudorouxii* nom. invar.), and *Z. sapae* strain ABT 301 were identified, based on their ITS region or 26S D1/D2 domain sequences

or both, to be heterogeneous. When a wider pool of 78 strains, retrieved from sugary and salty niches, was analyzed, we found similar rDNA heterogeneity (based on 5.8S ITS PCR-RFLP patterns and 26S D1-D2 sequencing in a considerable number of strains (**refer Chapter 6**), which when standard chemotaxonomic methods are employed appears indistinguishable from *Z. rouxii* and its close phylogenetic relatives (**refer Table 2.2 and Supplementary Table S2.1, Chapter 2**). Altogether, it appears that rDNA heterogeneity in these strains limits species delimitation and complicates taxonomical assignment. Even in more recent studies there seems to be confusion about the identity of *Zygosaccharomyces* yeast strains. James et al. (2005) reported that, in a phylogenetic study on *Zygosaccharomyces* yeasts, four strains CBS 4837, NCYC 3042, NCYC 3060 and NCYC 3061 all possessed identical D1/D2 sequences which differed significantly (i.e., by 15 and 25 base substitutions, respectively) from the type strains of both *Z. rouxii* and *Z. mellis*, indicating a significant misidentification to have occurred. Indeed, the NJ supported tree derived from 26S D1/D2 domain sequences they obtained showed that these strains to form a distinct phylogenetic group (bootstrap value, 100%) closely related to, but nevertheless, separate from, *Z. mellis* and *Z. rouxii*. Collectively, the results from the 26S rDNA sequencing suggest that these four yeasts, belong to a distinct, and hitherto undescribed, species of *Zygosaccharomyces*.

In order to establish the extent of rDNA heterogeneity in *Z. rouxii* complex (**refer Chapter 3**) and uncharacterized strains (**refer Chapter 6**), isolated in our laboratory and/or retrieved form culture collection and associated with high sugary and salty food niches. By PCR-RFLP analysis of 5.8S-ITS and D1D2 domains, we identified putative strains heterogeneous for either one or two phylogenetic markers, and then we cloned fragments from the amplified DNA. Sequence based phylogenetic tree could be informative in explaining evolutionary scenarios in relation to rDNA heterogeneity. Additionally, the variable sites in D1 and D2 regions that correspond to stretches of hairpin loops in the predicted RNA secondary structure of the rDNA region may impact the structural stability of rDNA and could also be informative in explaining evolutionary events that

result in rDNA heterogeneity. The purpose of the present study was to explore the significance of rDNA variation in a wider pool of *Zygosaccharomyces* strains, to see whether the patterns are indicative of bio-geographic, ecological, or microevolutionary aspects of these yeasts.

7.2. Materials and methods

7.2.1. Strains, medium and culture conditions

In this study, we considered six strains that showed rDNA heterogeneity with respect to ITS rDNA region or 26S rDNA region (**Supplementary Table S7.1 and Table 7.3**). These strains include NBRC0495, NBRC0505, NBRC10652, NBRC10669, NBRC10670 and M21 (**refer Chapter 6**). Besides this, six strains belonging to *Z. rouxii* complex (**refer Chapter 3**) were also considered. The strains were maintained and routinely grown in medium containing 1% yeast extract, 2% peptone, 2% glucose (YPD) with or without 2% agar. All the experiments were carried out from single-cell clones, obtained by plating out cells of their overnight cultures grown in the YPD onto YPDA plates and incubating them at 28°C for 2-3 days. All the strains were long term cryo-preserved at -80°C at the Unimore Microbial Culture Collection (website: www.umcc.unimore.it).

7.2.2. DNA manipulation

Genomic DNA (gDNA) isolation procedure utilizes physical disruption of the yeast cell wall using small glass beads and Phenol-Chloroform based extraction protocol as described by Hoffman & Winston (1987) and Sambrook et al. (1989) and as detailed in previous chapter (**refer section 2.2.2, Chapter 2**). Genomic DNA pellets obtained after extraction procedure were washed with 70% Ethanol, air dried for 20 min, resuspended in 35µL sterile double distilled water containing 1.5µL of RNAse (10mg/mL) and incubated at 37°C for 1hr in water bath. Finally, the

extracted DNAs were stored at -20°C. Resultant genomic DNAs were diluted to 100ng/µL and used for performing for further experiments.

7.2.3. PCR amplification and restriction analysis

The entire internal transcribed spacer (ITS) region (ITS 1, ITS 2 and the intervening 5.8S rRNA gene) for cloning purpose was amplified using primers ITS1 and ITS4 (White et al., 1990) as described in previous chapter (**refer section 2.2.3, Chapter 2**). Domains 1 and 2 of the 26S rRNA gene (D1/D2), for restriction digestion with *Ava*I and cloning purpose, were amplified using primers NL1 and NL4 (O'Donnell, 1993) as described in previous chapter (**refer section 2.2.3, Chapter 2**).

The *Ava*I restrcition digestion of 26S D1/D2 domain PCR amplicons was performed according to manufacturer's instruction. The restriction fragments were resolved electrophoretically using 2.0% (w/v) agarose gels in 0.5×TBE (89 mM Tris-borate, 2 mM EDTA pH 8) buffer, stained with 0.05% (v/v) ethidium bromide and visualized under UV light. Fragment lengths were estimated by comparing them with a GenRuller 100bp Plus DNA ladder (MBI Fermentas) as size marker. The gel was photographed using BioDoc Analyze (Biometra GmbH, Germany).

7.2.4. Cloning of 5.8S ITS region and 26S D1-D2 domain

PCR amplicons of 5.8S ITS regions and 26S rDNA D1-D2 domains were ligated into a PGEM vector and used for the transformation of the *E. coli* competent cells JM 109, according to manufacturer's instructions and previously described (**section 3.2.6, Chapter 3**). Fifteen to thirty white colonies (transformed cells) were picked aseptically and inoculated in LB tubes which thereafter were incubated overnight at 37°C. Plasmids from different overnight grown clones were extracted using Qiagen Miniprep kit according to manufacturer's instructions, and PCR-amplified

with the same primers pairs ITS1/ITS4 and NL1/NL4, respectively. The inserts were screened using the same enzyme previously used for RFLP analysis. At least two clones per insert were sequenced in both directions using the same primers.

7.2.5. Sequencing and phylogenetic analysis

Direct PCR amplicons and/or the cloned PCR amplicons of 5.8S ITS region and D1/D2 domain of 26S rDNA region was purified by using the DNA Clean and Concentrate TM-500 Kit (Zymo Research) and sequenced with the same primer used in PCR reaction by a commercial sequencing service provider (MWG Biotech, Germany).

Sequences of 5.8S-ITS regions and D1-D2 domain of 26S rDNA were edited and assembled using SeqMan software (DNASTAR) and BLASTed against the GenBank database to retrieve sequences of the closest relatives as well as sequences belong to mosaic strains (**refer section 3.2.1, Chapter 3**). All the sequences were aligned using the Clustal-W algorithm (Thompson et al., 1997). Phylogenetic analysis of 5.8S-ITS regions and 26S rDNA D1-D2 domains were carried out by the neighbour-joining (NJ) method with 1000 bootstrap iterations (Saitou & Nei, 1987). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches in both trees (Felsenstein, 1985). In the case of 26S rDNA D1-D2 NJ-tree, the evolutionary distances were computed using the Kimura 2-parameter method (Kimura, 1980), while the evolutionary distances among 5.8S-ITS rDNA sequences were computed using the Tamura-Nei method (Tamura and Nei, 1993). All the analyses were carried out in MEGA V6.0 (Tamura et al., 2011).

7.2.6. Southern blotting analysis

Approximately 5-8µg genomic DNA was digested overnight with selected enzymes, which either do not cut any of the previously cloned copies or cut one out of two heterogenous copies of the 5.8S ITS region or D1-D2 domain of 26S rDNA (**Table 7.1**). *In silico* digestion of the previously cloned sequences were carried out using the online tool NEBcutter V2.0 (New England Biolabs, Inc). The resulting digested genomic DNA was separated on a 0.8% agarose gel using 0.5X TBE buffer at 60V for 2 hr 30 min with a DIG-labelled molecular weight marker (DNA Molecular Weight Marker III, Ref: 11 218 603 910, Roche Diagnostics GmbH, Germany) After steps of denaturation (2x15min in Denaturation solution-1.5M NaCl, 0.5M NaOH), neutralization (2x15 min in Neutralization solution-1.5M NaCl, 0.5M Tris-HCl pH 7.0, 1mM EDTA), and equilibration (at least 20 min in 2X SSC-0.3M NaCl, 0.03M Sodium Citrate) (Sambrook et al., 1989), the DNA was blotted from the gel onto a positively charged nylon membrane (Ref: 11 209 299 001, Roche Diagnostics GmbH, Germany), by overnight capillary transfer, using 2X SSC buffer (0.3M NaCl, 0.03M Sodium Citrate). The blotted DNA was UV fixed on nylon membrane with UV crosslinker (UVC500, Hoefer) at 1200x100µJ/cm^2. The nylon membrane was then air dried for 2 hr, subsequently pre-hybridized for at least 30 min with 5 ml DIG Easy Hyb (Ref: 11 603 558 001, Roche Diagnostics GmbH, Germany) at 42°C in a hybridization oven (Amersham Pharmacia Biotech) and, finally, hybridized overnight at 42°C with 10 ml DIG Easy Hyb containing 200ng of a digoxigenin-labelled DNA probe (1:400 diluted, dilution as estimated by visually comparing the band intensities of probe), which had been previously heat-denatured by boiling for 5 min at 99°C. The digoxigenin-labelled 5.8S-ITS and D1-D2 26S rDNA probes were PCR-synthesized with the same primer pair used for the corresponding PCR reactions through a DIG PCR Labeling Mix^{PLUS} kit (Ref: 11 835 289 910, Roche Diagnostics GmbH, Germany) according to manufacturer's instructions.

After two stringency washes one in low stringency buffer (2X SSC containing 0.1% SDS) for 10 min followed by another in high stringency buffer (0.5X SSC containing 0.1% SDS) twice for 5 min], the hybridized bands were immune-detected with Anti-digoxigenin-AP Fab fragments (Ref: 11 093 274 910, Roche Diagnostics GmbH, Germany) conjugated to alkaline phosphatase and were then visualized with a chemiluminescence substrate (CDP-Star, *ready-to*-use, Ref: 12 041 677 001, Roche Diagnostics GmbH, Germany), according to manufacturer's instructions. Enzymatic dephosphorylation of this substrate by alkaline phosphatase leads to a light emission that was recorded on X-ray films (Kodak X-OMAT Scientific Imaging Film) for 20 min. The X-ray films are developed by washing X-ray films sequentially in Kodak GBX Developer, double distilled water, Kodak GBX Replenisher (Cat. No.: P7167, Sigma). Probe removal for rehybridization of a membrane was achieved by washing in stripping buffer (containing 0.2M NaCl, 0.1% SDS and 2X SSC) and incubating 37°C for 30 min. Following probe removal the stripped membrane can be washed in 2XSSC and then stored in 2XSSC until rehybridization at 2-8°C.

rDNA Heterogeneous region	Strains	Non-cutter restriction enzymes Name; incubation/inactivation conditions	Single-cutter restriction enzymes Name; incubation/inactivation conditions
26S D1-D2 domain	NBRC0495	BclI; 55°C overnight/No inactivation	MseI; 65°C overnight/No inactivation
	NBRC0505	BclI; 55°C overnight/No inactivation	SacI; 37°C overnight/65°C for 20mins
	NBRC10652	BclI; 55°C overnight/No inactivation	HphI; 37°C overnight/65°C for 20mins
	NBRC10669	BclI; 55°C overnight/No inactivation	AhdI; 37°C overnight/65°C for 20mins
	NBRC10670	BclI; 55°C overnight/No inactivation	AvaI; 37°C overnight/65°C for 20mins
5.8S ITS gene	NBRC0495	BclI; 55°C overnight/No inactivation	DraI; 37°C overnight/65°C for 20mins
	NBRC10652	BclI; 55°C overnight/No inactivation	DraI; 37°C overnight/65°C for 20mins
	NBRC10669	BclI; 55°C overnight/No inactivation	MspI; 37°C overnight/80°C for 20mins
	NBRC10670	BclI; 55°C overnight/No inactivation	MspI; 37°C overnight/80°C for 20mins
	M21	BclI; 55°C overnight/No inactivation	MspI; 37°C overnight/80°C for 20mins

Table 7.1. List of restriction enzymes used for discriminating heterogeneous copies of ITS and 26S rDNA region (in strains with rDNA heterogeneity).

7.2.7. RNA secondary structure prediction

The cloned D1-D2 domain sequence from representative strains chosen from 26S D1-D2 phylogenetic tree were aligned using Clustal-W2 (**Figure 7.1**). First the entire D1-D2 sequences were analyzed to identify the structures formed by the variable regions. Both were parts of hairpin

stems. Then all segments not involved in the stems were removed from the nucleotide sequences and new secondary structures were generated. For this, the sequence immediate upstream and downstream of the D1 domain and the D2 domain, representing hairpin-stem structure, were removed to generate sequence corresponding to the D1 domain and the D2 domain. For D1 domain, the sequences were cut at position 56thbp position and 195th bp in *Z. rouxii* CBS 732^{T} (GenBank accession number-AJ046112) and the corresponding D1 flanked region was cut from the other aligned sequences. For D2 domain, the sequences were cut at 354th bp position to 570thbp position in *Z. rouxii* CBS 732^{T} (GenBank accession number-AJ046112). The secondary structure of the stem-loop structure present in D1 and D2 domain sequence were predicted using RNA Structure Ver 5.4. The variable sites were localized in the structures and the potential effect of the substitutions on the secondary structure was examined by comparing the structures generated from the individual clones. The secondary structure with the lowest free energy (maximum stability) has been used for comparison. The generated secondary structures for all cloned D1-D2 domains were then compared with those of the corresponding parts of the *Saccharomyces cerevisiae* large subunit (26S) rRNA molecules (also available at http://www.rna.icmb.utexas.edu/.) and *Z. rouxii* large subunit (26S) rDNA molecule. The nucleotides variable positions in the D1 and the D2 domain were ascertained.

7.3. Results and discussion

7.3.1. 5.8S ITS rDNA sequence analysis reveals heterogeneity

Our previous chapter (**refer Chapter 3**) highlighted the occurrence of relaxation of concerted evolution in some strains of the *Zygosaccharomyces rouxii* complex, resulting in intra-strain variation of rRNA sequences within individual genome. To assess the extent of rDNA heterogeneity in other strains belonging to the *Z. rouxii* complex, we considered a wide pool of putative *Z. rouxii* related strains retrieved from sugary and salty niches and/or from recovered

from other culture collections (**refer Chapter 6**). The 5.8S-ITS PCR RFLP analysis discriminated 64 out of 79 strains (including *S. cerevisiae* strain ABT507) that showed shared restriction pattern with one or more strains and assigned them into 9 clusters (from G-1 to G-9). The remaining 15 strains displayed unique restriction pattern.

Based on 5.8S ITS PCR-RFLP clustering (**refer Table 6.2, Chapter 6**), we found 6 representative uncharacterized strains that resulted in complex band profile and suggested that these strains could possess multiple divergent variants of 5.8S-ITS regions. The strains were NBRC0495, NBRC10652, NBRC10669, NBRC10670, NBRC10672 and M21 (**with bold case and underline, refer Table 6.2, Chapter 6**). To verify this hypothesis, we cloned the 5.8S ITS sequence from these heterogeneous strains in competent *E .coli* JM 109 cells. Based on the *Hae*III restriction screening, we choose two clones for 5.8S ITS sequencing (**Supplementary Table S7.1)**. The two clones from strains NBRC0495, NBRC10652, NBRC10669, NBRC10670, and M21, respectively harbour polymorphisms at 43 sites (9 indels, 19 transitions, 15 transversions), 37 (8 indels, 14 transitions, 15 transversions), 46 (14 indels, 18 transitions, 14 transversions), 34 (10 indels, 10 transitions, 14 transversions), and 72 (21 indels, 32 transitions, 19 transversions). In most cases, we found that transition mutations are generated at higher frequency than transversions. Polymorphism was not studied for clones from NBRC10672. The BLAST search in NCBI identified percent identities of the cloned 5.8S ITS regions from strains heterogeneous with those from the *Z. rouxii* CBS 732^{T}, *Z. sapae*, *Z. mellis* and *Z. siamensis* (**Table 7.2)**.

Strains	Clones	Clones designation	Identity with *Z. rouxii*	Identity with *Z. Sapae* ABT 301^T			Identity with *Z. Mellis*			Identity with *Z. siamensis*
			CBS 732	copy 1	copy 2	copy 3	NBRC 1615	CBS 711	CBS 7412	ATCC 12572
NBRC0495	2 3	copy 1 copy 2	94.0% 87.0%	93.0% 87.0%	97.0% 99.0%	98.0% 99.0%	91.0% 89.0%	93.0% 87.0%	93.0% 87.0%	91.0% 92.0%
NBRC0505	3	3	87.0%	87.0%	99.0%	99.0%	89.0%	87.0%	87.0%	92.0%
NBRC0845	1	1	99.0%	99.0%	88.0%	89.0%	NA	100%	99.0%	NA
NBRC10652	6 7	copy 2 copy 1	99.0% 100%	99.0% 99.0%	91.0% 88.0%	89.0% 89.0%	NA NA	99.0% 100%	99.0% 99.0%	NA NA
NBRC10668	2	2	99.0%	99.0%	91.0%	89.0%	NA	99.0%	99.0%	NA
NBRC10669	10 11	copy 2 copy 1	87.0% 99.0%	87.0% 99.0%	99.0% 92.0%	99.0% 89.0%	89.0% NA	87.0% 99.0%	87.0% 99.0%	92.0% NA
NBRC10670	6 7	copy 1 copy 2	100% 93.0%	100% 92.0%	88.0% 97.0%	89.0% 98.0%	NA 92.0%	100% 92.0%	99.0% 92.0%	NA 92.0%
NBRC10672	3	3	99.0%	99.0%	91.0%	89.0%	NA	99.0%	99.0%	NA
M21	1 2	copy 1 copy 2	98.0% 88.0%	98.0% 88.0%	88.0% 99.0%	89.0% 98.0%	NA NA	100% 88.0%)	99.0% 88.0%	NA 91.0%

Table 7.2. Percent identity between ITS gene from heterogeneous strains and other species (*Z. rouxii* CBS 732T, *Z. sapae*, *Z. mellis* and *Z. siamensis*) as obtained from BLAST result. GenBank accession number of the ITS genes are AM943655.1 (*Z. rouxii* CBS 732^T), AM279465.1 (ABT 301^T copy 1), AM279464.1 (ABT 301^T copy 2), AM279696.1 (ABT 301^T copy 3), AB302841.1 (*Z. mellis* NBRC 1615), FN431889.1 (*Z. mellis* CBS 711), FN431890.1 (*Z. mellis* CBS 7412), and KC881076.1 (*Z. siamensis* ATCC 12572).

7.3.2. Phylogenetic analysis of heterogenic 5.8S ITS rDNA haplotypes

To investigate the genealogical relations among 5.8S ITS rDNA haplotypes, the sequences obtained in this study were aligned with 15 5.8S ITS sequences (10 were previously submitted by us) retrieved from GenBank. A total of 32-haplotype dataset was used to construct the NJ phylogenetic tree, which include sequences of phylogenetically close relative species and sequences used an outgroup. Tree topology delineated four major branches, named *Z. rouxii*, *Z. rouxii*-like, R, and *Z. sapae*, and one minor branch with low statistical supports (**Figure 7.1**). The major branches were supported by 100% bootstrapping value, except the *Z. rouxii* group. The result suggested that reticulation could occur in the evolution of these haplotypes. Remarkably, the 5.8S ITS haplotype 1 from the strain M21 (in *Z. rouxii* group) showed high divergence from the other *Z. rouxii*-like haplotypes (bootstrapping value 71%), suggesting that cryptic species could be present in this clade.

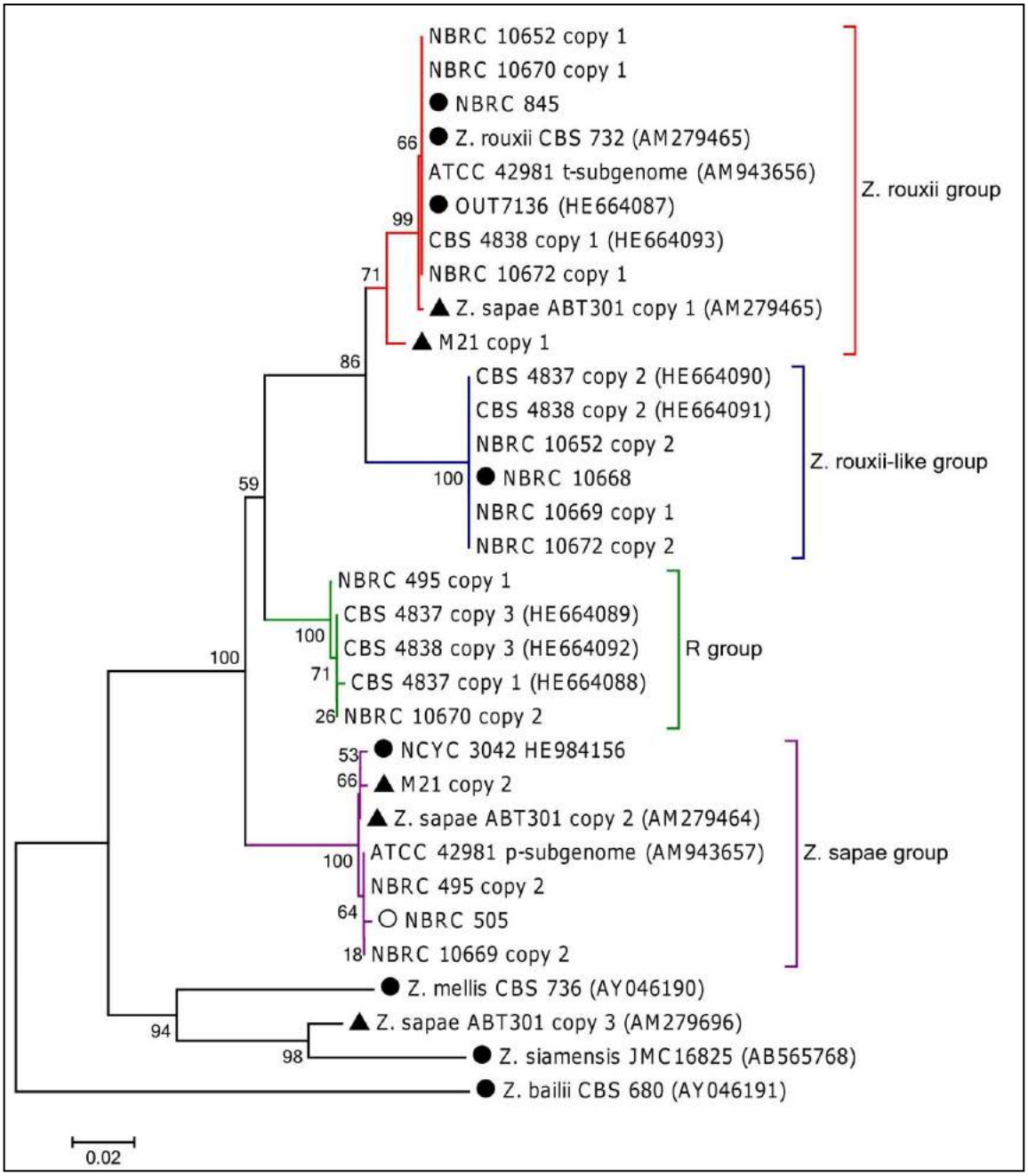


Figure 7.1. Phylogenetic tree based on 5.8S ITS sequence. The phylogeny was inferred using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Tamura-Nei method. All positions containing gaps and missing data were eliminated. *Z. rouxii*, *Z. rouxii* like, R, and *Z. sapae* clades are coloured in red, blue, green, and pink, respectively. Light blue branch underlines *Z. sapae* ABT301^{T} 5.8S-ITS copy 3. Black triangles represent strains with homogeneous D1-D2 26S rDNA genes and heterogeneous 5.8S-ITS rDNA regions. Black circles indicate strains with no heterogeneity both in D1-D2 26S rDNA genes and 5.8S-ITS rDNA regions. White circles indicate strains with heterogeneous D1-D2 26S rDNA genes and homogeneous 5.8S-ITS rDNA regions.

7.3.3. 26S D1-D2 rDNA sequence analysis reveals heterogeneity

All the uncharacterized representative strains along with *Z. rouxii* NBRC0845 and *Z. sapae* M21 (**refer Table 6.2, Chapter 6**) were submitted to sequencing of the 26S D1-D2 domains The direct sequencing of 26S D1-D2 PCR amplicons from the genomic DNA showed a high number of polymorphic nucleotide positions leading to the failure of sequencing reactions in 8 strains. We didn't performed D1-D2 sequencing for strain NBRC0525. Our sequencing effort was successful for only strain NBRC10668. The failure in sequencing reactions could arise from technical problems hampering amplification during PCR or the occurrence of double templates in the mix. To exclude the possibility of contaminated cultures, the sequencing was repeated after two times repeated subculturing of single colonies from strains with D1-D2 26S rRNA gene heterogeneity. Out of 8 strains, no polymorphic positions were found in two strains (NBRC0845 and M21), whereas 5 strains harboured polymorphisms in the 26S rRNA genes. We supposed that these 6 strains could possess multiple divergent variants of 26S D1-D2 regions. The strains were NBRC0495, NBRC0505, NBRC10652, NBRC10669, NBRC10670, and NBRC10672 (**showed in bold case and underlined in Table 6.2, Chapter 6**).

In order to verify this hypothesis, we cloned full-lengths PCR-amplified D1-D2 domain of the 26S rDNA region from the heterogeneous strains in competent *E .coli* JM 109 cells. Accordingly, the restriction digestion of 26S D1-D2 domain PCR amplicons with *Ava*I restriction endonucleases for these strains supported heterogeneity in their rDNA array (**Table 7.3)**. Significantly, those strains that are found heterogeneous with respect to their 5.8S ITS gene are heterogeneous in their 26S D1-D2 domain too. The only exception was that the strain M21, that showed 5.8S ITS heterogeneity, is not heterogeneous in its 26S D1-D2 domain; instead NBRC0505 was found to be heterogeneous in 26S D1-D2 but not in 5.8S ITS gene.

Strains	D1/D2 Sequencing	AvaI Pattern	AvaI Profile	D1-D2 Clones (AvaI Profile)	Clone designation	rDNA heterogeneity D1-D2	Reference
CBS 732	-	510-100	Zr	-		no	Chapter 6, 7
CBS 736	-	430-100	Zm	-		-	Chapter 6, 7
NCYC 3042	-	390-100	Zp	-		-	Chapter 6, 7
ABT 301	passed	390-100	Zp	-		-	Soleiri et al., 2013a, b
ATCC 42981	-	390-100	Zp (p-subgenome)	-		yes	Solieri et al., 2013b
	-	510-100	Zr (t-subgenome)	-			
CBS 4837	failed	510-390-100	Zr+Zp	**7**, 8, 9, 15 (Zr)	copy 1	yes	Soleiri et al., 2013b
				1, 2, 3, 4, 5, 6, 10, 11, 12, 13, 14, 16 (Zp)	copy 2		
CBS 4838	failed	510-390-100	Zr+Zp	3, 6, **7**, 8, 9, 11 (Zr)	copy 1	yes	Soleiri et al., 2013b
				1, 2, 4, 5, 10, 12 (Zp)	copy 2		
OUT 7136	passed	390-100	Zp	-	-	no	Soleiri et al., 2013b
NBRC0495	failed	510-390-100	Zr+Zp	**1**, 9, 10, 11, 13 (Zr)	copy 1	yes	Chapter 6, 7
				2, 3, 4, 5, 6, 7, 8, 12, 14, 15 (Zp)	copy 2		
NBRC0505	failed	510-430-100	Zr+Zm+Zp	**12** (Zm)	copy 1	yes	Chapter 6, 7
				1, 2, 3, 4, 5, 6, 7, 9, 10, 11, **13**, 14, 15 (Zp)	copy 2		
				8 (Zr)	-		
NBRC0525	not performed	510-390-100	Zr+Zp	not performed	-	-	Chapter 6, 7
NBRC0845	passed	510-100	Zr	-	-	no	Chapter 6, 7
NBRC10652*	failed	510-390-100	Zr+Zp	**12**, 13, 14 (Zr)	copy 1	yes	Chapter 6, 7
				1, 2, 3, 4, 5, 6, 7, 9, 10, **11**, 15 (Zp)	copy 2		
NBRC10668	passed	390-100	Zp	-		no	Chapter 6, 7
NBRC10669	failed	510-430-390-100	Zr+Zm+Zp	**2**, 3, 6, 7, 9, 10, 13, 14, 15 (Zr)	copy 1	yes	Chapter 6, 7
				1, 4, 5, 8, 11 (Zp)	copy 2		
				12 (Zm)	-		
NBRC10670	failed	510-430-390-100	Zr+Zm+Zp	**2**, 9 (Zm)	copy 1	yes	Chapter 6, 7
				1, **3**, 4, 5, 6, 7, 10, 11, 13, 14, 15 (Zr)	copy 2		
				12 (Zp)	-		
NBRC10672	failed	510-390-100	Zr+Zp	-	-	no	Chapter 6, 7
M21	passed	390-100	Zp	-	-	no	Chapter 6, 7

* Strain NBRC10652 needs more clones to be analyzed by RFLP and/or sequencing to confirm whether the RFLP pattern (510-390-100) is correct or it is different (probably 390-100) because our sequencing attempted yielded sequences having polymorphism at a single position, which could be due to sequencing error.

Table 7.3. D1-D2 sequencing of all representative and singlet strains (chosen based on 5.8S ITS PCR-RFLP based screening; refer Chapter 6**), *Ava*I pattern, profile, and clones designation (**for strains whose sequencing failed, i.e., heterogeneous strains**).**

.

Strains	Clones designation	Identity with *Z. rouxii* CBS 732^T	Identity with *Z. sapae* ABT 301 Copy 2	Identity with *Z. Mellis* NBRC1615	Identity with *Z. Siamensis* ATCC 12572
2	-	94.0%	95.0%	99.0%	99.0%
4		94.0%	96.0%	99.0%	100%
9	-	94.0%	96.0%	99.0%	99.0%
41	-	94.0%	96.0%	99.0%	100%
70	-	94.0%	96.0%	99.0%	99.0%
5CF	-				100%‡
NBRC0495	copy 1 copy 2	91.0% 97.0%	92.0% 99.0%	98.0% 95.0%	93.0% 95.0%
NBRC0505	copy 1 copy 2	94.0% 97.0%	95.0% 99.0%	99.0% 96.0%	99.0% 96.0%
NBRC0523	-	97.0%	99.0%	95.0%	95.0%
NBRC0845	-	97.0%	99.0%	95.0%	95.0%
NBRC10652	copy 2 copy 1	96.0% 96.0%	99.0% 99.0%	96.0% 96.0%	96.0% 97.0%
NBRC10668	3	97.0%	100%	NA	96.0%
NBRC10669	copy 2 copy 1	97.0% 94.0%	99.0% 96.0%	96.0% 99.0%	96.0% 99.0%
NBRC10670	copy 2 copy 1	94.0% 100%	99.0% 98.0%	97.0% 94.0%	97.0% 94.0%
NBRC10672	copy 1 copy 2	100% 98.0%	98.0% 100%	94.0% 96.0%	94.0% 96.0%
M21	-	97.0%	100%	NA	96.0%

Table 7.4. Percent identity between 26S D1-D2 region from heterogeneous strains and other species (*Z. rouxii* CBS 732T, *Z. sapae*, *Z. mellis* and *Z. siamensis*) as obtained from BLAST result. GenBank accession number of the ITS genes are **JQ689016.1** (*Z. rouxii* CBS 732^T), **AM947682.1** (*Z. sapae* ABT 301^T copy 2), **AB302837.1** (*Z. mellis* NBRC 1615), **KC881059.1** (*Z. siamensis* ATCC 12572), and **KF273863.1 (*D. hansenii*)**‡.

Based on the *Ava*I restriction screening, we choose two clones for 26S D1-D2 sequencing. Different copies of 26S D1-D2 rDNA region in NBRC0495, NBRC0505, NBRC10652, NBRC10669, and NBRC10670 respectively harbour polymorphism at 49 sites (5 indels, 23 transitions, 22 transversions), 23 (3 indels, 15 transitions, 5 transversions), 1 (1 transition), 23 (2 indels, 15 transitions, 6 transversions), and 21 (1 indel, 12 transitions, 8 transversions). Only the part of the sequence that aligned with the sequence related to other copy was considered for estimating polymorphic sites. In all cases, we found that transition mutations are generated at higher frequency than transversions. A single transition in strain NBRC10562 could be attributed to sequencing error. Polymorphism was not studied for clones from NBRC10672.

The Blastn search revealed two different 26S D1-D2 haplotypes are dominant in the genome of these strains (**Table 7.4**). Strains NBRS10670 and NBRC 10672 harboured one haplotype 99% identical to *Z. sapae* and the other one 100% identical to *Z. rouxii*. Strains NBRC10669, NBRC495 and 505 possessed one haplotype more related to *Z. mellis*/*Z. siamensis* and the other one 100% identical to *Z. sapae*. Finally the strain NBRC 10652 was found to possess two *Z. sapae*-like D1-D2 haplotypes highly similar to each other. This result disagreed with that previously obtained by the direct *Ava*I-restriction pattern of D1-D2 26S region PCR amplicon, which discriminated between one *Z. sapae* and one *Z. rouxii*-like sequences in the NBRC10652 genomic DNA. D1-D2 26S rDNA domain cloning and sequencing of strain NBRC10652 are being currently repeated in our Laboratory to solve this incongruence. Overall, the Blastn result showed high congruency with the NJ tree based phylogenetic tree (**Figure 7.2**).

The multiple sequence alignment of all cloned sequences identified variability (mostly heterogeneous) at 103 sites (**Table 7.5 and 7.6**). The variable positions were not distributed evenly along the entire length of the D1-D2 domains (**Figure 7.3**). The majority of them grouped in two short regions, which we have referred as variable region 1 (VR1) (**shaded in green, Figure 7.3**) and variable region 2 (VR2) (**shaded in red, Figure 7.3**). Twenty seven variable sites of the

D1 domain formed VR1 and sixty-six sites of the D2 domain comprised VR2. Out of 27 variable sites in D1 region, in 25 sites usually two nucleotides alternated. There were only two sites (at position 168bp and 169bp) in D1 region where more than two different nucleotides occurred when all cloned sequences were compared. Among the 25 variable sites, at 15 sites (in blue shade), the nucleotide remained conserved in all heterogeneous *Zygosaccharomyces* strains and alternation has taken place just with respect to *S. cerevisiae* (**Table 7.5**). Out of 66 variable sites in D1 region, in 61 sites usually two nucleotides alternated. There were only five sites (at position 445, 456, 457, 473 and 535bp) in D1 region where more than two different nucleotides occurred when all cloned sequences were compared. Among the 61 variable sites, at 15 sites (in blue shade), the nucleotide remained conserved in all heterogeneous *Zygosaccharomyces* strains and alternation has taken place with respect to *S. cerevisiae* (**Table 7.6**). The presence of ambiguous nucleotides in the database of D1-D2 sequences of the type strain and the strain studied in this chapter possess diverse large subunit rRNA genes in their rDNA arrays. The BLAST search in NCBI with the cloned sequences did not identify identical sequences with the exception of few strains. Percent identity of 26S D1-D2 region from strains heterogeneous with that from the *Z. rouxii* CBS 732^{T}, *Z. sapae*, *Z. mellis* and *Z. siamensis* obtained from BLAST result has been depicted in **Table 7.4**.

Strains	60	88	89	91	92	94	96	97	101
Z. rouxii CBS 732	T	A	A	T	G	T	C	T	A
CBS 4837 copy 2	T	G	A	T	G	T	C	T	A
CBS 4838 copy 2	T	A	A	T	G	T	C	T	A
ABT 301	T	G	A	T	G	T	C	T	A
NBRC0495 copy 1	C	G	A	C	G	T	C	T	G
NBRC0495 copy 2	T	A	A	T	G	T	C	C	A
NBRC0505 copy 1	C	G	A	C	G	T	C	T	G
NBRC0505 copy 2	T	G	A	T	G	T	C	T	A
NBRC10652 copy 1	T	A	A	T	G	T	C	T	A
NBRC10669 copy 2	C	G	A	T	G	T	C	T	A
NBRC10669 copy 1	T	G	A	C	G	T	C	T	G
NBRC10670 copy 1	T	G	A	C	G	T	C	T	G
S. cerevisiae CBS 1171	del	G	G	C	A	C	T	T	G
Strains	**103**	**105**	**106**	**107**	**112**	**151**	**158**	**165**	**166**
Z. rouxii CBS 732	T	G	C	C	C	C	A	G	T
CBS 4837 copy 2	T	G	C	C	C	C	A	G	T
CBS 4838 copy 2	T	G	C	C	C	C	A	G	T
ABT 301	T	G	C	C	C	C	A	G	T
NBRC0495 copy 1	T	G	C	G	C	C	A	A	T
NBRC0495 copy 2	T	G	C	C	C	C	A	G	T
NBRC0505 copy 1	T	G	C	G	C	C	A	A	T
NBRC0505 copy 2	T	G	C	C	C	C	A	G	T
NBRC10652 copy 1	T	G	C	C	C	C	A	G	T
NBRC10669 copy 2	T	G	C	C	C	C	A	G	T
NBRC10669 copy 1	T	G	C	G	C	C	A	A	T
NBRC10670 copy 1	T	G	C	G	C	C	A	A	T
S. cerevisiae CBS 1171	C	T	T	C	T	T	T	G	A
Strains	**168**	**169**	**171**	**172**	**183**	**185**	**186**	**189**	**190**
Z. rouxii CBS 732	A	T	C	A	G	A	C	T	T
CBS 4837 copy 2	A	T	C	A	G	G	C	T	C
CBS 4838 copy 2	A	T	C	A	G	A	C	T	T
ABT 301	A	T	C	A	G	G	C	T	C
NBRC0495 copy 1	T	A	C	A	G	G	C	T	C
NBRC0495 copy 2	A	T	C	A	G	G	C	T	C
NBRC0505 copy 1	T	A	C	A	G	G	C	T	C
NBRC0505 copy 2	G	T	C	A	G	G	C	T	C
NBRC10652 copy 1	A	T	C	A	G	G	C	T	C
NBRC10669 copy 2	A	T	C	A	G	G	C	T	C
NBRC10669 copy 1	T	A	C	A	G	G	C	T	C
NBRC10670 copy 1	T	A	C	A	G	G	C	T	C
S. cerevisiae CBS 1171	T	G	del	G	A	G	T	C	T

Table 7.5. **Nucleotide variable positions in the D1 domain.**

Strains	365	385	386	387	394	401	415	416	433	438	440
Z. rouxii CBS 732	A	C	T	C	C	G	G	C	A	T	G
CBS 4837 copy 2	G	C	T	C	C	G	G	C	A	T	G
CBS 4838 copy 2	A	C	T	C	C	G	G	C	A	T	G
ABT 301	A	C	T	C	C	G	G	C	A	T	G
NBRC0495 copy 1	A	C	T	C	C	G	G	C	G	G	A
NBRC0495 copy 2	A	C	T	C	C	G	G	C	A	T	G
NBRC0505 copy 1	A	C	T	C	C	G	G	C	A	T	G
NBRC0505 copy 2	A	C	T	C	C	G	G	C	A	T	G
NBRC10652 copy 1	A	C	T	C	C	G	G	C	A	T	G
NBRC10669 copy 2	A	C	T	C	C	G	G	C	A	T	G
NBRC10669 copy 1	A	C	T	C	C	G	G	C	A	T	G
NBRC10670 copy 1	A	C	T	C	C	G	G	C	A	T	G
S. cerevisiae CBS 1171	A	T	C	T	T	A	T	T	A	T	G

Strains	441	445	449	452	453	454	455	456	457	458	461
Z. rouxii CBS 732	C	A	T	A	T	C	T	C	T	G	A
CBS 4837 copy 2	C	A	G	A	T	C	C	T	C	G	A
CBS 4838 copy 2	C	A	G	A	T	C	C	T	C	G	A
ABT 301	C	A	G	A	T	C	C	T	C	G	A
NBRC0495 copy 1	T	T	T	T	G	A	C	T	A	A	A
NBRC0495 copy 2	C	A	G	A	T	C	C	T	C	G	A
NBRC0505 copy 1	T	A	G	A	G	C	C	T	T	G	G
NBRC0505 copy 2	C	A	G	A	T	C	C	T	C	G	A
NBRC10652 copy 1	C	A	G	A	T	C	C	T	C	G	A
NBRC10669 copy 2	T	A	G	A	G	C	C	T	T	G	A
NBRC10669 copy 1	C	A	G	A	T	C	C	T	C	G	A
NBRC10670 copy 1	C	A	G	A	T	C	C	T	C	G	A
S. cerevisiae CBS 1171	T	C	T	A	T	C	C	A	T	A	A

Strains	466	467	469	470	471	473	474	475	476	478	479
Z. rouxii CBS 732	G	G	T	T	C	T	T	C	T	C	G
CBS 4837 copy 2	G	G	T	del	C	T	G	C	C	C	T
CBS 4838 copy 2	G	G	T	del	C	T	G	C	C	C	T
ABT 301	G	G	T	del	C	T	G	C	C	C	T
NBRC0495 copy 1	G	G	T	del	C	A	del	C	T	C	G
NBRC0495 copy 2	G	G	T	del	C	T	G	C	C	C	T
NBRC0505 copy 1	G	A	T	del	C	T	T	T	T	T	T
NBRC0505 copy 2	G	G	T	del	C	T	G	C	C	C	T
NBRC10652 copy 1	G	G	T	del	C	T	G	C	C	C	T
NBRC10669 copy 2	G	A	T	del	C	T	T	T	T	T	T
NBRC10669 copy 1	G	G	C	del	C	T	G	C	C	C	T
NBRC10670 copy 1	G	G	T	del	C	T	G	C	C	C	T
S. cerevisiae CBS 1171	A	G	T	del	del	G	del	C	C	C	G

Strains	**481**	**482**	**483**	**484**	**485**	**486**	**489**	**498**	**499**	**502**	**503**
Z. rouxii CBS 732	G	A	G	G	G	A	G	C	A	G	G
CBS 4837 copy 2	G	T	A	G	G	A	G	C	G	G	G
CBS 4838 copy 2	G	T	A	G	G	A	G	C	A	G	G
ABT 301	G	T	A	G	G	A	G	C	G	G	G
NBRC0495 copy 1	T	del	del	G	G	A	G	T	T	T	T
NBRC0495 copy 2	G	T	A	G	G	A	G	C	G	G	G
NBRC0505 copy 1	G	C	G	G	G	G	G	C	A	G	G
NBRC0505 copy 2	G	T	A	G	G	A	G	C	G	G	G
NBRC10652 copy 1	G	T	A	G	G	A	G	C	G	G	G
NBRC10669 copy 2	G	C	G	G	G	G	G	C	A	G	G
NBRC10669 copy 1	G	T	A	G	G	A	G	C	G	G	G
NBRC10670 copy 1	G	T	A	G	G	A	G	C	G	G	G
S. cerevisiae CBS 1171	T	del	del	del	A	A	A	T	T	G	del

Strains	**504**	**513**	**515**	**516**	**517**	**518**	**522**	**527**	**528**	**533**	**535**
Z. rouxii CBS 732	A	A	C	T	G	G	T	T	A	A	A
CBS 4837 copy 2	A	A	C	T	G	G	T	T	A	A	A
CBS 4838 copy 2	A	A	C	T	G	G	T	T	A	A	A
ABT 301	A	A	C	T	G	G	T	T	A	A	A
NBRC0495 copy 1	G	T	T	C	T	A	C	A	C	T	T
NBRC0495 copy 2	A	A	C	T	G	G	T	T	A	A	A
NBRC0505 copy 1	A	A	C	C	G	G	T	T	A	A	T
NBRC0505 copy 2	A	A	C	T	G	G	T	T	A	A	A
NBRC10652 copy 1	A	A	C	T	G	G	T	T	A	A	A
NBRC10669 copy 2	A	A	C	C	G	G	T	T	A	A	T
NBRC10669 copy 1	A	A	C	T	G	G	T	T	A	A	A
NBRC10670 copy 1	A	A	C	T	G	G	T	T	A	A	A
S. cerevisiae CBS 1171	A	A	C	T	G	G	T	A	C	A	G

Strains	**536**	**537**	**538**	**539**	**541**	**543**	**550**	**561**	**563**	**564**	**567**
Z. rouxii CBS 732	T	T	T	T	T	A	T	G	T	A	T
CBS 4837 copy 2	T	T	T	T	T	A	T	G	T	A	T
CBS 4838 copy 2	T	T	T	T	T	A	T	G	T	A	T
ABT 301	T	T	T	T	T	A	T	G	T	A	T
NBRC0495 copy 1	T	T	del	T	A	T	T	A	C	T	A
NBRC0495 copy 2	T	T	T	T	T	A	T	G	T	A	T
NBRC0505 copy 1	C	T	T	A	T	A	T	G	T	-	-
NBRC0505 copy 2	T	T	T	T	T	A	T	G	T	A	T
NBRC10652 copy 1	T	T	T	T	T	A	T	G	T	A	T
NBRC10669 copy 2	C	T	T	A	T	A	T	G	T	A	T
NBRC10669 copy 1	T	T	T	T	T	A	T	G	T	A	T
NBRC10670 copy 1	T	T	T	T	T	A	T	G	T	A	T
S. cerevisiae CBS 1171	T	A	del	A	T	A	C	G	T	A	T

Table 7.6. **Nucleotide variable positions in the D2 domain.**

7.3.4. Phylogenetic analysis of heterogenic 26S D1-D2 rDNA haplotypes

One model to explain divergent intra-genomic rDNA copies is gradual spread of a "master repeat" by serial duplications during which the new copies acquire novel mutations in their inherited sequences. If this is the case, then the evolutionary history can be reconstructed by a phylogenetic analysis producing a bifurcating phylogenetic tree. Therefore, to investigate the genealogical relations among D1-D2 26S rDNA haplotypes, the sequences obtained in this study were aligned with 14 D1-D2 26S sequences (6 were previously submitted by us) retrieved from GenBank. This 29-haplotype dataset was used to construct the NJ phylogenetic tree shown in **Figure 7.2**. Tree topology delineated three major branches, named *Z. sapae*, *Z. rouxii*, and *Z. mellis*/*Z. siamensis*, and several other minor branches with low statistical supports. These results suggested that reticulation could occur in the evolution of these haplotypes. Remarkably, NBRC0495 genome hosted a *Z. sapae* like D1-D2 26S haplotype and a D1-D2 26S haplotype highly divergent from the recognized species *Z. mellis* and *Z. siamensis* (boostrapping value 94%), suggesting that cryptic species could be present in this clade. Moreover, the same strain showed a *Z. sapae*-like 5.8S ITS haplotype and another divergent haplotype which clustered in R group.

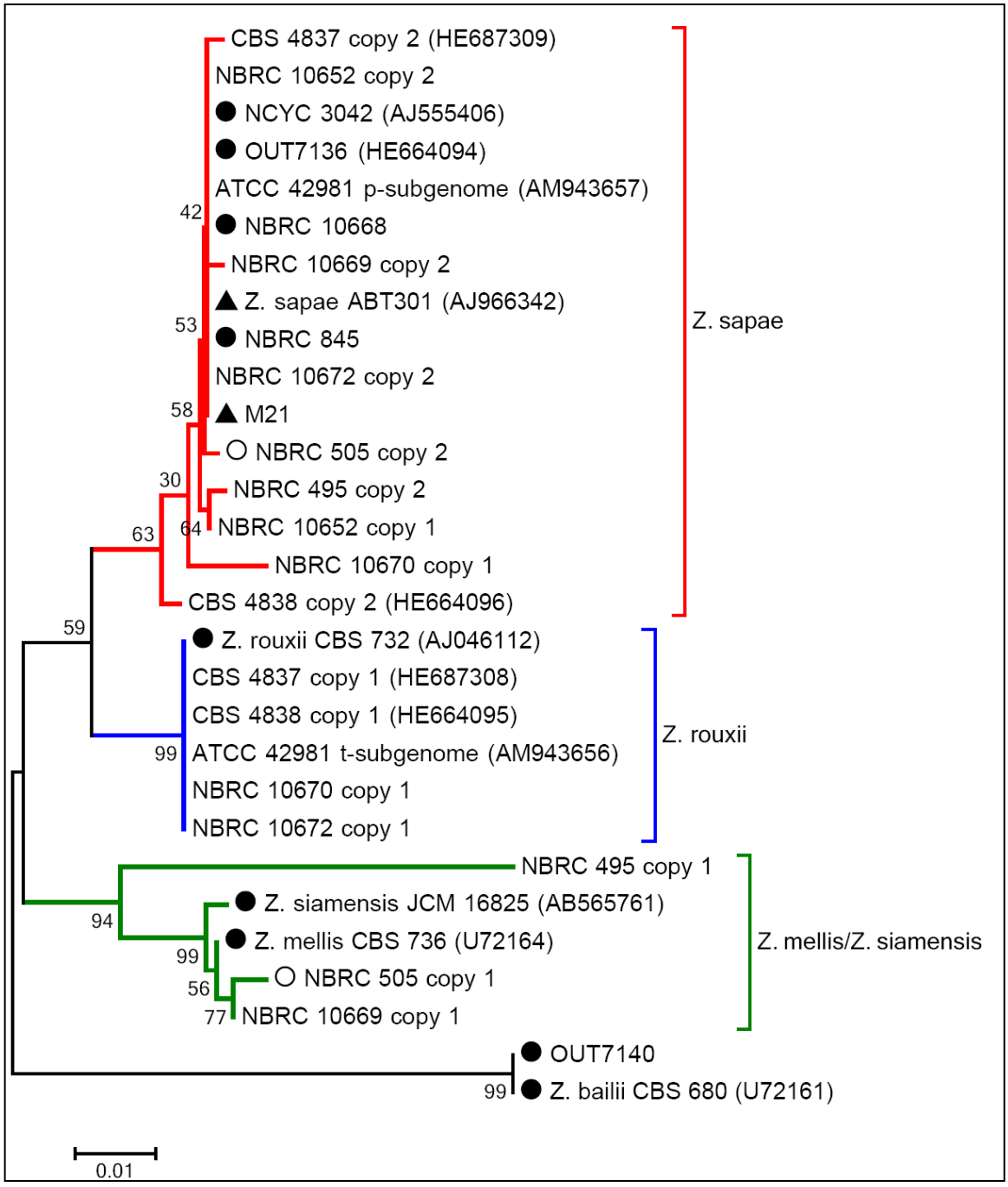


Figure 7.2. Phylogenetic relationships of strains based on D1-D2 26S rDNA gene. The evolutionary history was inferred using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Kimura 2-parameter method and are in the units of the number of base substitutions per site. Evolutionary analyses were conducted in MEGA 6.0. *Z. sapae*, *Z. rouxii*, and *Z. mellis*/*Z. siamensis* clades are coloured in red, blue, and green, respectively. Black triangles represent strains with homogeneous D1-D2 26S rDNA genes and heterogeneous 5.8S-ITS rDNA regions. Black circles indicate strains with no heterogeneity both in D1-D2 26S rDNA genes and 5.8S-ITS rDNA regions. White circles indicate strains with heterogeneous D1-D2 26S rDNA genes and homogeneous 5.8S-ITS rDNA regions.

7.3.5. Non-canonical and compensatory base-pair changes in hairpin stem in D1 and D2 region

To examine whether the sequence differences of the cloned D1-D2 domains entailed alterations in the structure of the encoded RNA molecules, we generated secondary structures for all cloned D1-D2 domains and compared them with those of the corresponding parts of the *Saccharomyces cerevisiae* large subunit (26S) rRNA molecules. The cloned sequences showed similar structures, in which both variable regions were involved in helical stems of hairpin loops (stem-loops) (**Figure 7.4** and **Figure 7.5**). The *S. cerevisiae* produced here were essentially identical to the corresponding stems in the secondary models of the complete rRNA sequences available in http://www.rna.icmb.utexas.edu/.

Structural variability was observed only in the stem parts, apparently due to the variable sites that grouped in the backfolding strands of the helices. Similar topological variability and grouping of variable sites in the back-folding strands were also observed in the hairpin loops that included the VR2 segments. As described above, both highly variable regions were located in the back-folding 3'- strands of the hairpins and almost all nucleotides of the variable sites paired with nucleotides of stable positions. This implies that each substitution at a variable site could alter the helical structure of its stem because it can disrupt the normal base pairing. Consistent with this, the hairpins of the clones showed variable patterns of paired and unpaired stretches (**Figure 7.4** and **Figure 7.5**). However, not all nucleotide substitutions caused changes in the secondary structures. The structural neutrality of certain nucleotide changes could be attributed to non-canonical base pairing referred to as wobble base paring.


Z. rouxii (AY046112)
CBS 4838 copy 2
NBRC0495 copy 2
NBRC10652 copy 1
CBS 4837 copy 2
NBRC10669 copy 2
ABT 301
NBRC0505 copy 2
NBRC10670 copy 2
NBRC0505 copy 1
NBRC10669 copy 1
NBRC0495 copy 1
S. cerevisiae (AY048154)

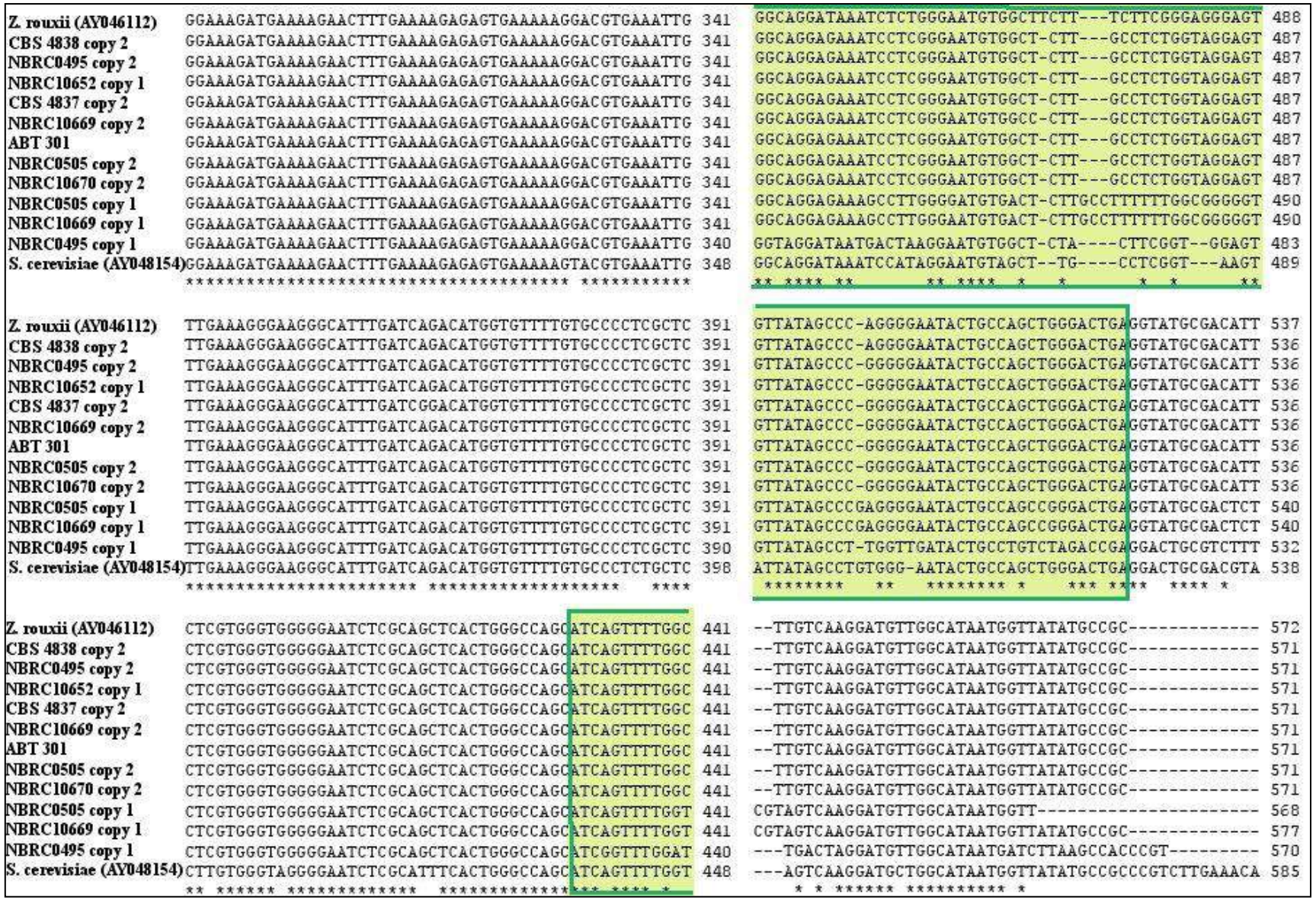
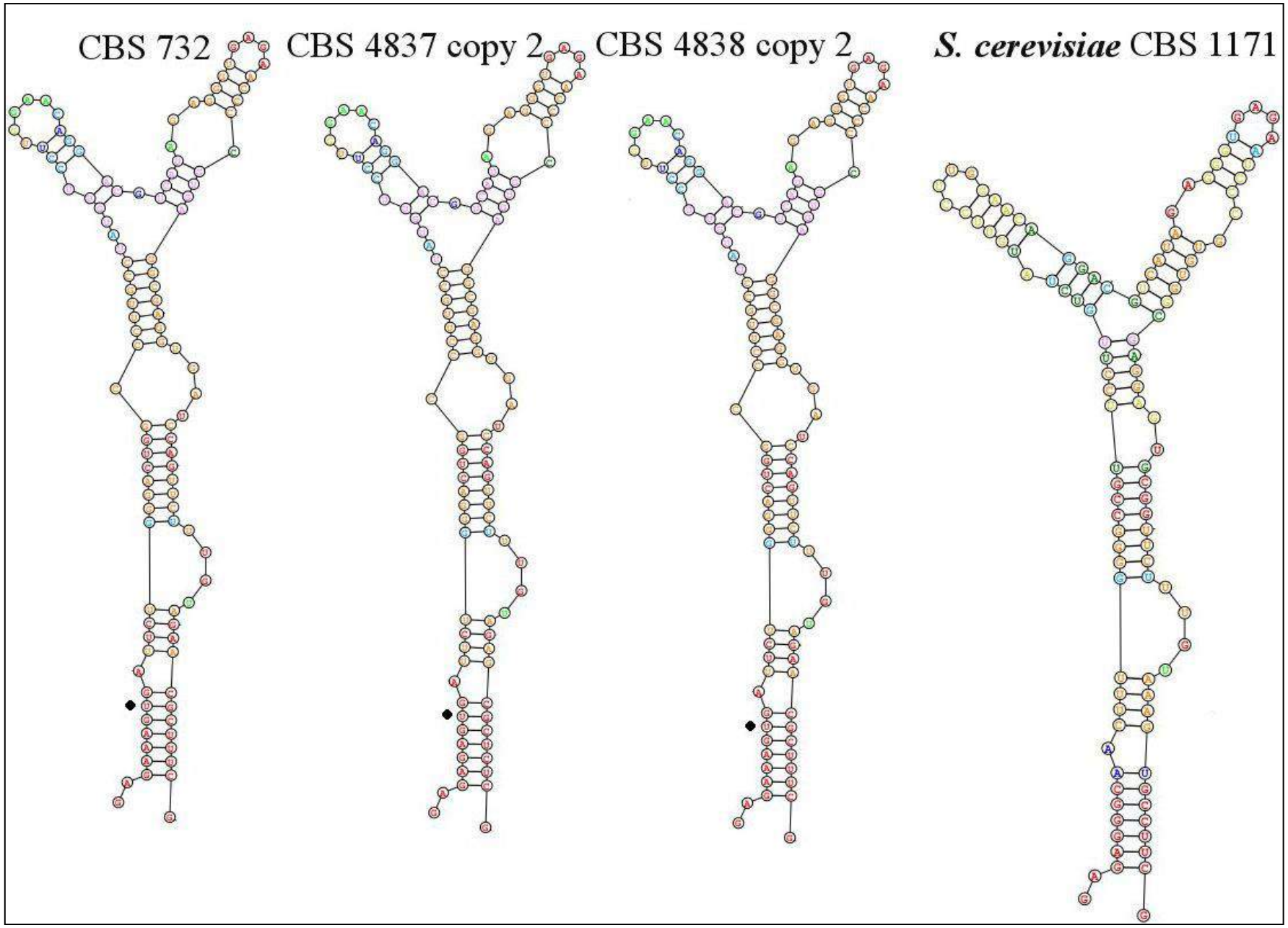
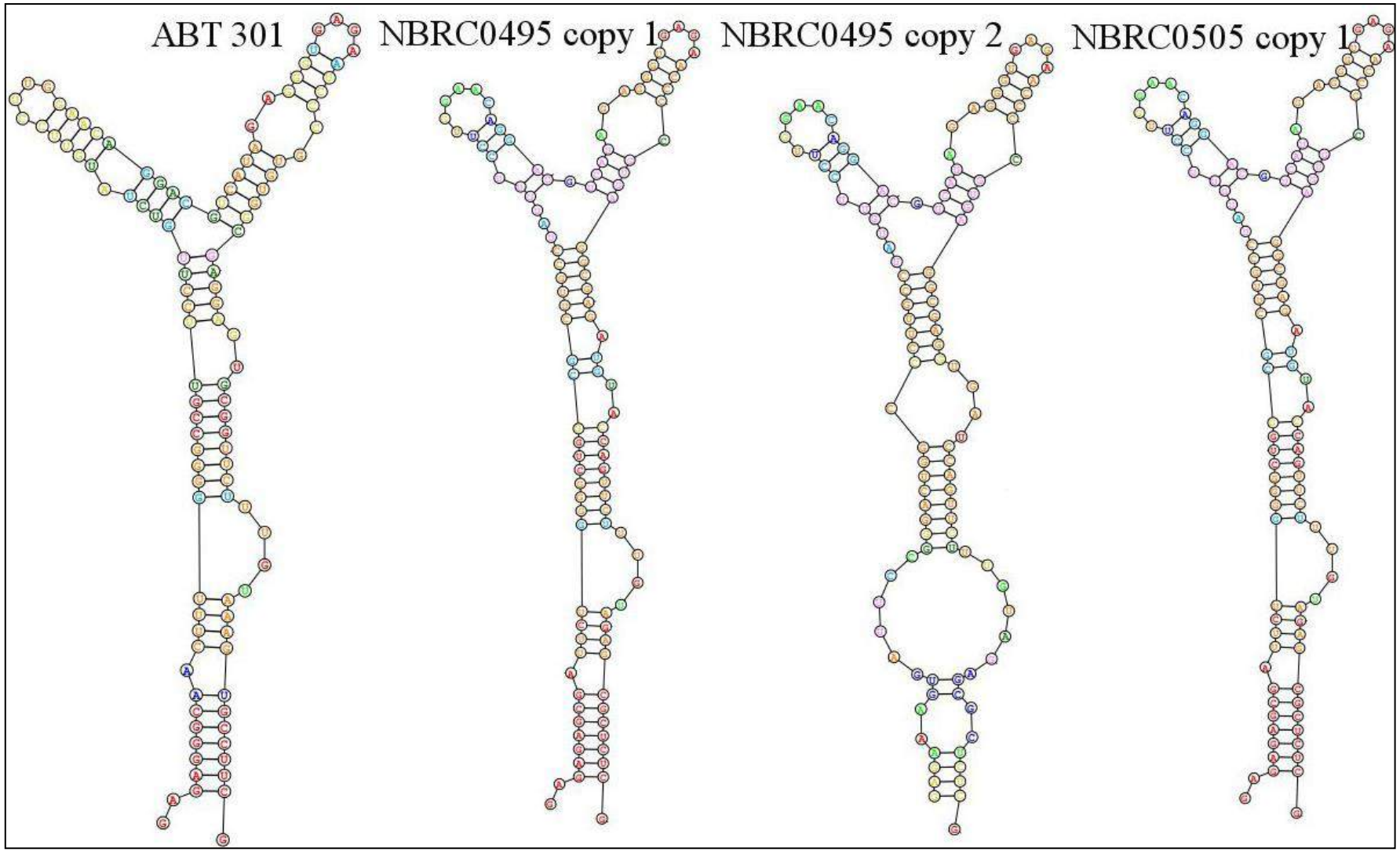


Figure 7.3. Multiple sequence alignment of representative 26S D1-D2 domain haplotypes. Red and green shades represent VR1 and VR2, respectively.

Due to this peculiarity of RNA, guanine can pair not only with cytosine but also with uracil in the RNA helix (Varani et al., 2000). Thus, the substitution of cytosine by thymine in the DNA sequence does not necessarily affect the structure of the RNA helix. One variable site of VR1 and five in VR2 of *Z. rouxii* CBS 732^T (and corresponding sites in heterogeneous strains as in aligned) at position 91, 455, 456, 469, 476, and 516 paired with stable guanines in the hybridizing segment of the hairpin stems (shown as a dot in *Z. rouxii* structure; **Figure 7.4** and **Figure 7.5**). Most of them had either T(U) or C in the cloned sequences. Their transitions did not alter the stem structure in many cloned sequences, confirming that wobble pairing of nucleotides did neutralize many substitutions indeed. However, in some cloned sequences the transition has structural alterations. So, the structural variability could be ascribed to substitutions in sites, where wobble pairing was not possible.


CBS 732
CBS 4837 copy 2
CBS 4838 copy 2
S. cerevisiae CBS 1171


ABT 301
NBRC0495 copy 1
NBRC0495 copy 2
NBRC0505 copy 1

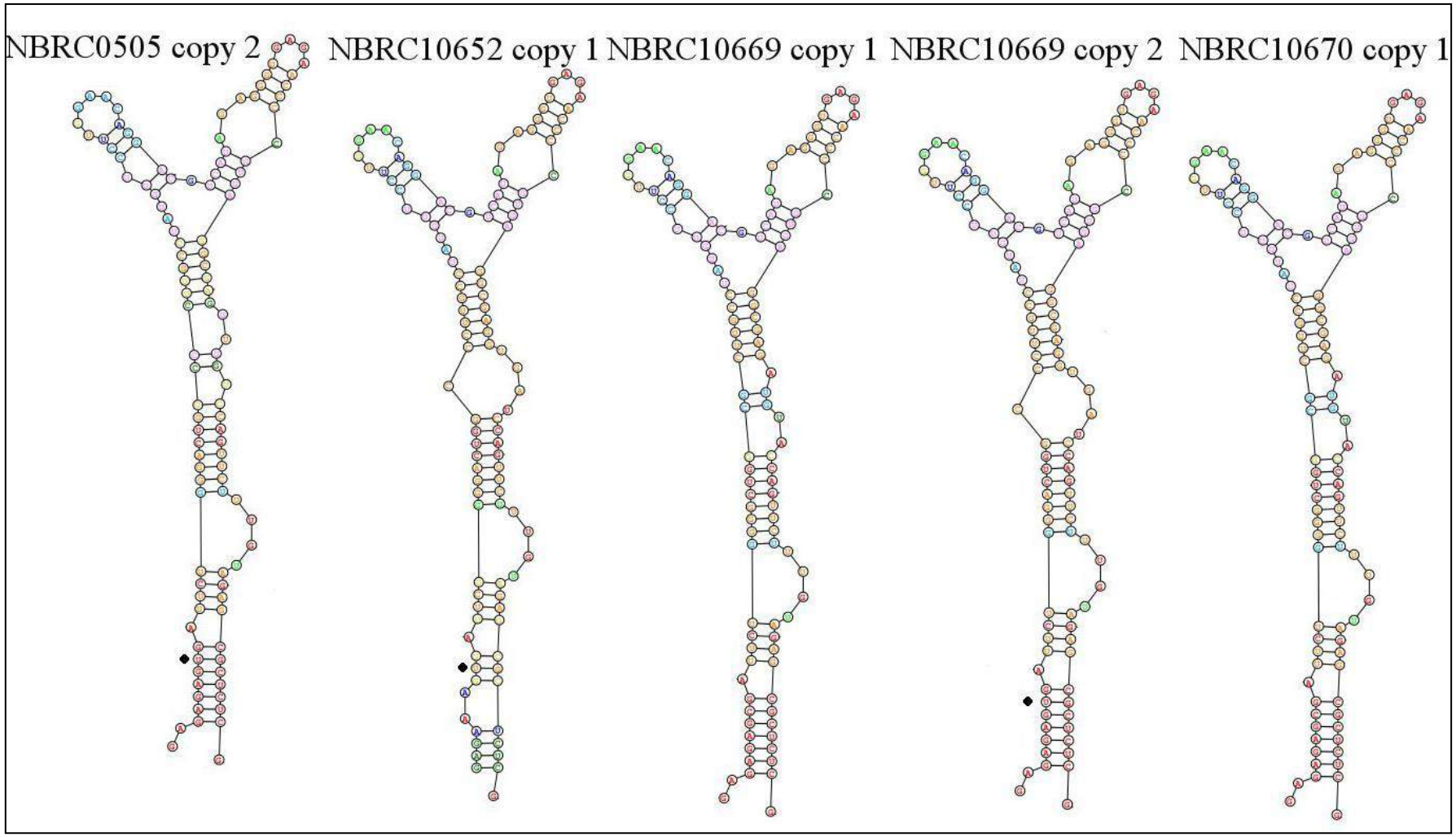
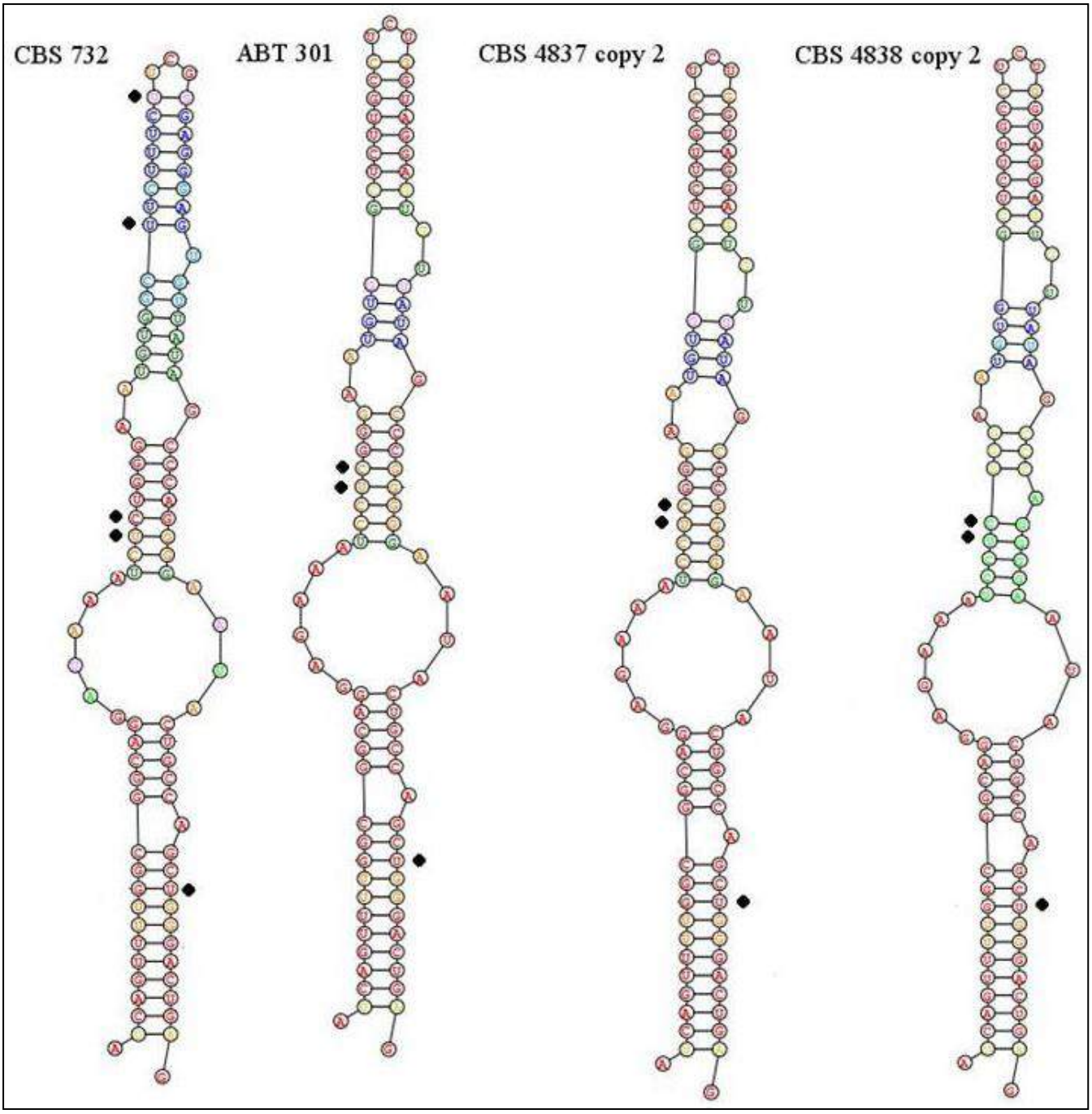
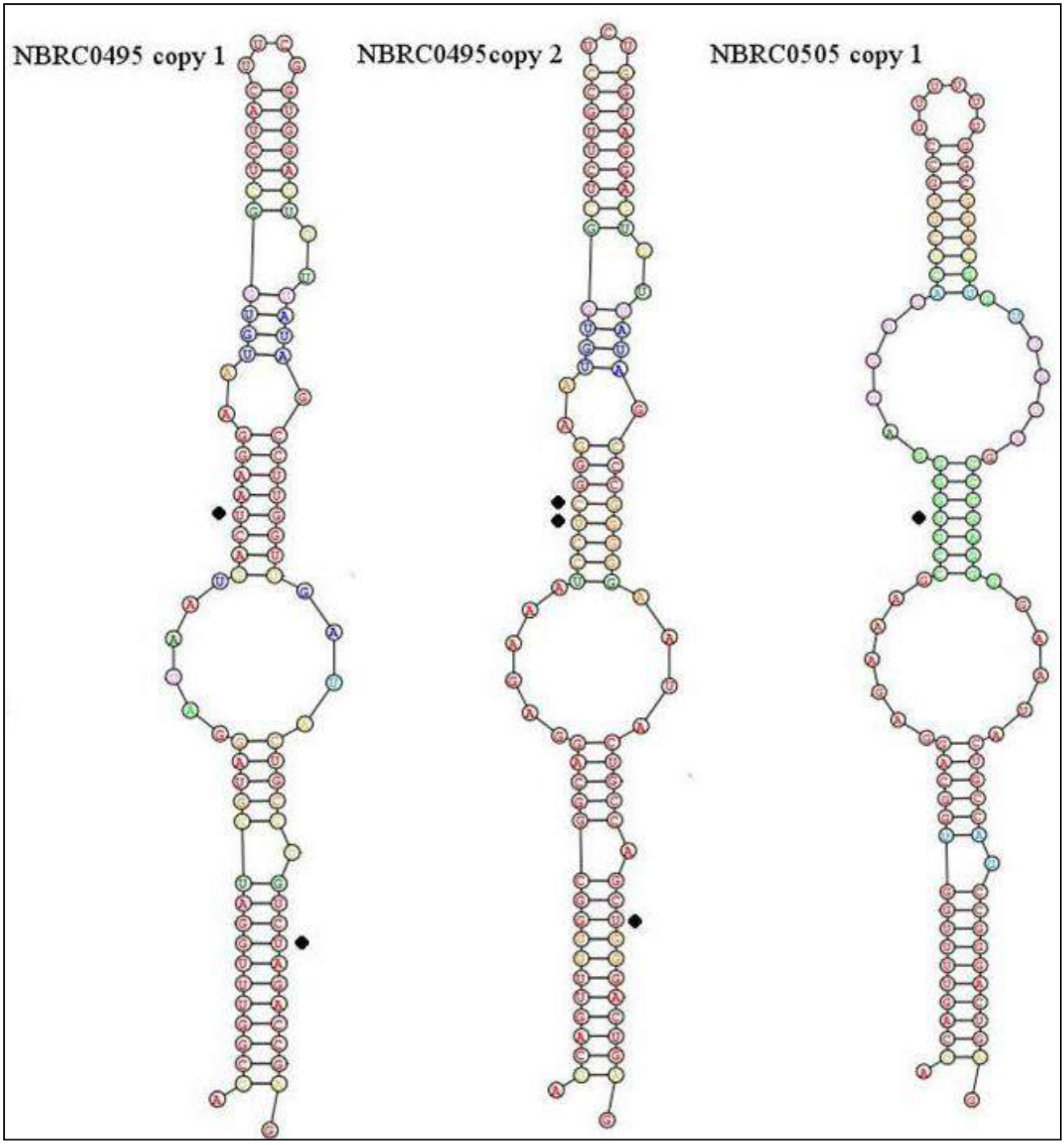
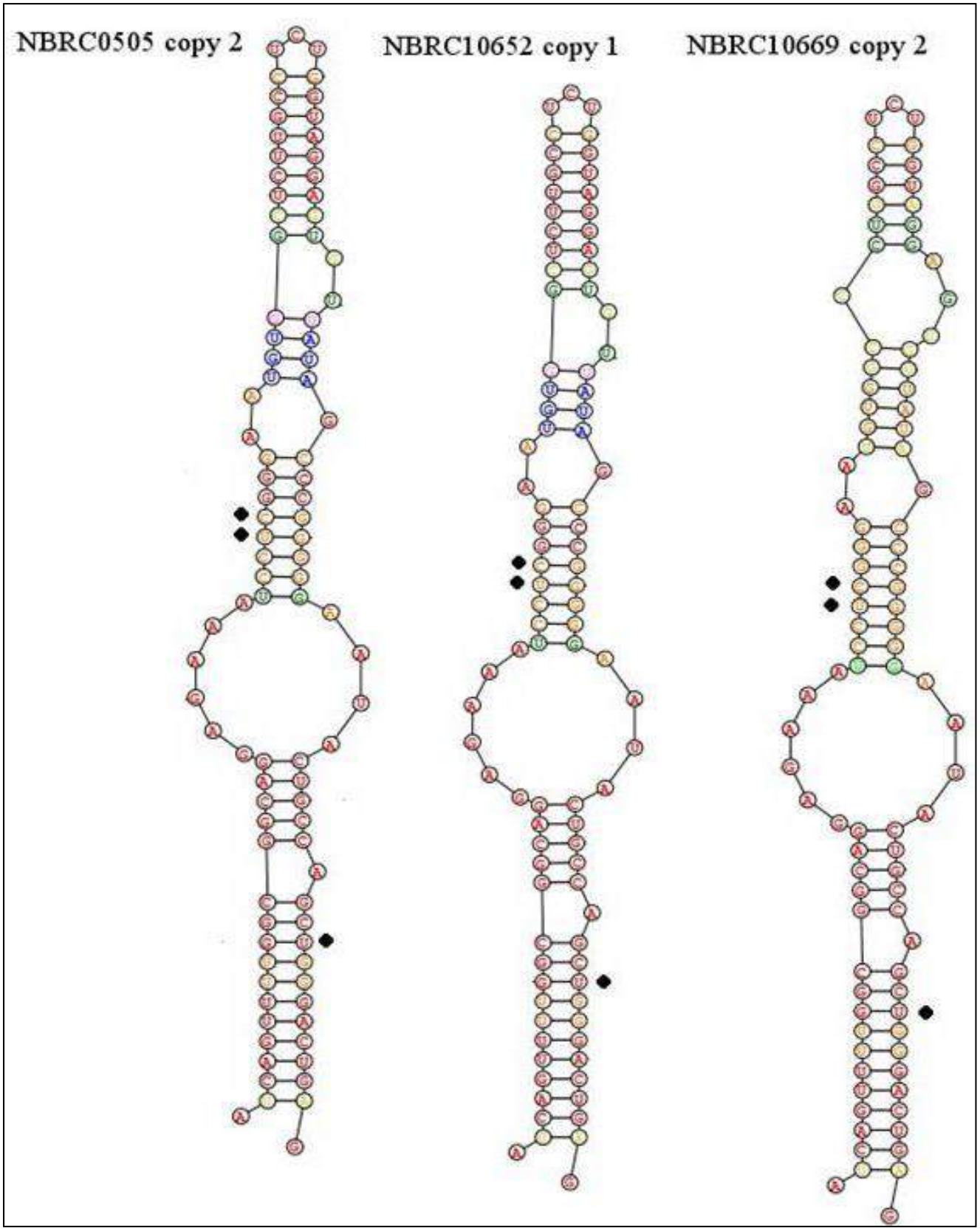


Figure 7.4. Predicted secondary structures of D1 hairpin-stem loops (for representative strains chosen from NJ based UPGMA phylogenetic tree). The wobble base pairing sites in 26S D1-D2 domain sequence of *Z. rouxii* and their corresponding (if present) in the sequence of other heterogeneous strains are marked with dots.


CBS 732
ABT 301
CBS 4837 copy 2
CBS 4838 copy 2


NBRC0495 copy 1
NBRC0495copy 2
NBRC0505 copy 1


NBRC0505 copy 2
NBRC10652 copy 1
NBRC10669 copy 2

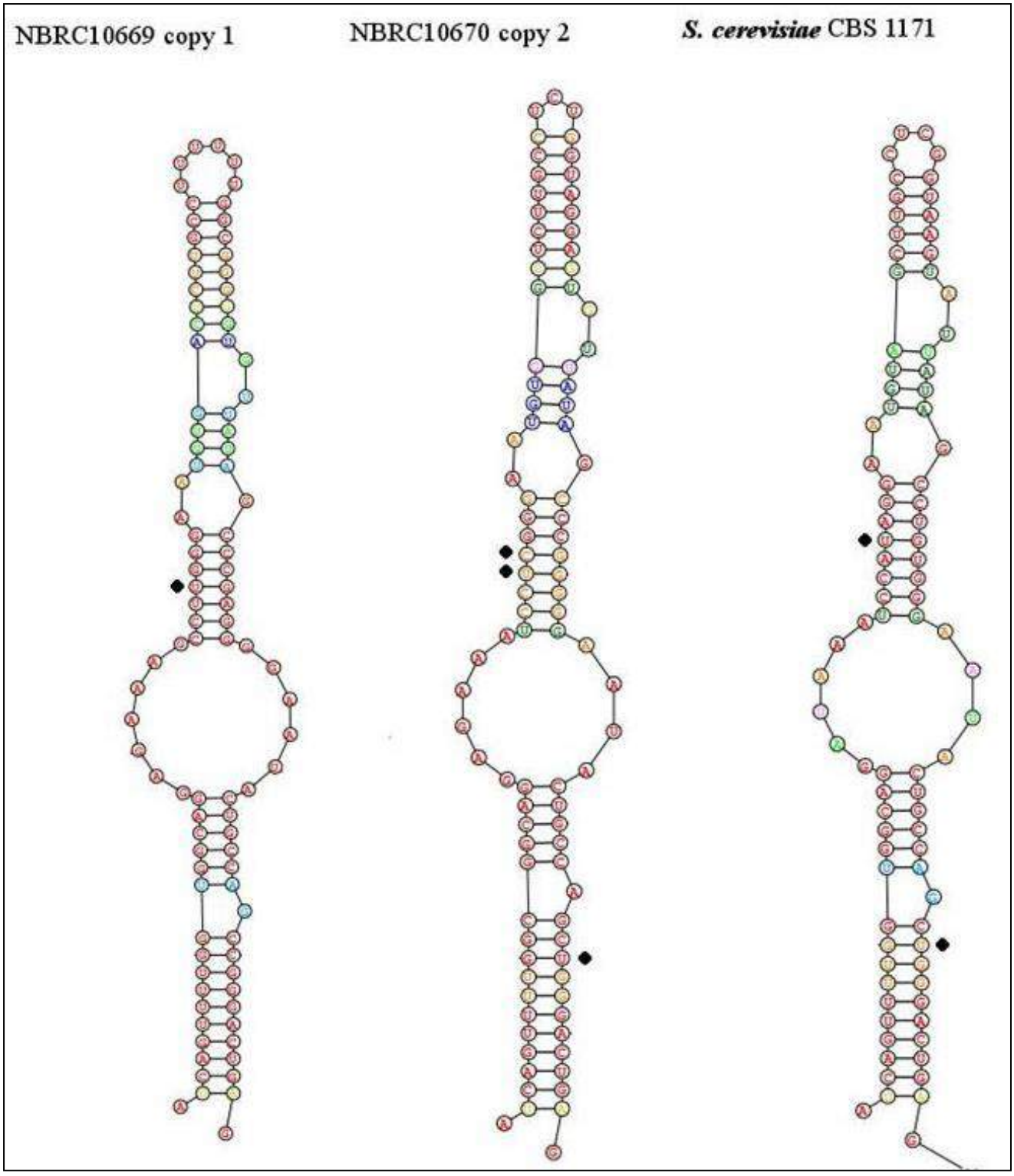


Figure 7.5. Predicted secondary structures of D2 hairpin-stem loops (of representative strains chosen from NJ based UPGMA phylogenetic tree). The wobble base pairing sites in 26S D1-D2 domain sequence of *Z. rouxii* and their corresponding (if present) in the sequence of other heterogeneous strains are marked with dots.

7.3.6. Southern blotting analysis confirmed heterogeneity

In order to confirm the presence of divergent intragenomic rDNA array in the genome of heterogeneous strains (5 strains each for 5.8S ITS and 26S D1-D2 Southern blotting), we performed Southern blot analysis, starting from the following assumptions:. First, the length of rDNA in strains with heterogeneous rDNA is same as the length of rDNA in *Z. rouxii* CBS 732^T, i.e. 9940bp (accession number: AM943655). Second, the ITS and 26S rDNA regions both start at same position in strains with rDNA heterogeneity and in *Z. rouxii* CBS 732^T. The ITS region and 26S rDNA region, in *Z. rouxii* CBS 732^T, respectively start at 4748bp and 5424bp. The strains that showed heterogeneity in their ITS region were NBRC0495, NBRC10652, NBRC10669, NBRC10670, and M21. The strains that showed heterogeneity in their 26S rDNA region were NBRC0495, NBRC0505, NBRC10652, NBRC10669, and NBRC10670. Cleavage sites on the forward strand are shown as a ^ and those on the complementary strand as _.

Detection of heterogeneity in ITS region

The genomic DNA of NBRC0495 was digested with six enzymes, *Bcl*I (recognize T^GATC_A), *Bam*HI, *Msp*I, *Dra*I (TTT^AAA), *Ssp*I, and *Eco*RI. *Bcl*I is a rare cutter and it cuts the *Z. rouxii* CBS 732^T rDNA array at two sites, one at 3922bp upstream of ITS region and other at 5795bp in its 26S rDNA region. *Bcl*I cuts at position 378bp in NBRC0495 copy 1 and at 382bp in NBRC0495 copy 2. It does not cut any of the two copies of ITS region in this strain. Immediate downstream it has restriction site at 3922bp in *Z. rouxii* CBS 732^T rDNA. If we assume this site to be conserved, the bands of 1880bp or 1884bp size are expected. Figure show a band of approximated size 1900bp suggesting that this site is conserved in NBRC0495. For detecting two heterogeneous copies of ITS region in strains NBRC0495 we choose *Dra*I restriction enzyme. *Dra*I cuts the *Z. rouxii* CBS 732^T ITS region at two sites, one at 4936bp upstream of ITS region and other at 5321bp. *Dra*I has two sites in NBRC0495 copy 1, one at 137bp other at 591bp. The

latter site was absent in NBRC0495 copy 2. Both copies of the 26S rDNA region of this strain is devoid of *Dra*I site. Immediate upstream in *Z. rouxii* CBS 732^{T}, *Dra*I site is present at 6430bp. The restriction digestion of genomic DNA, as expected, gave two bands, one of 454bp (591bp-137bp=454bp) related to copy 1 and other of 1545bp related to copy 1 (that recognize sequence between restriction site at 137 and upstream conserved *Dra*I site at 6430bp as in *Z. rouxii* CBS 732^{T}).

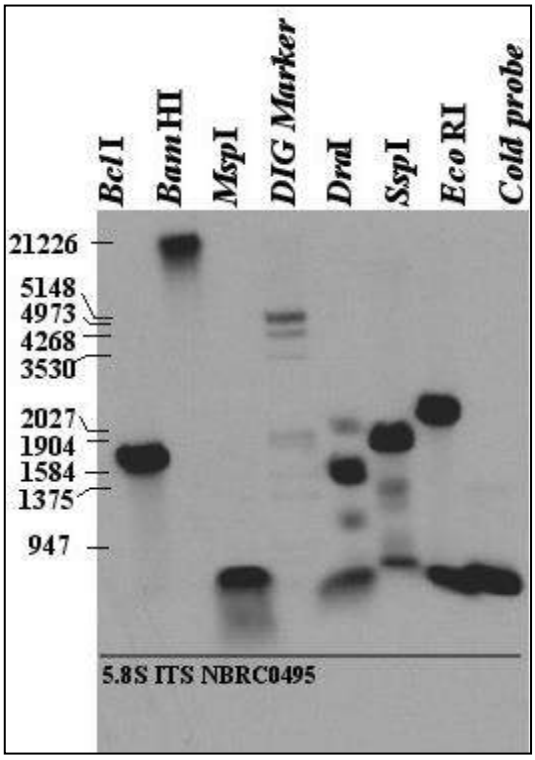
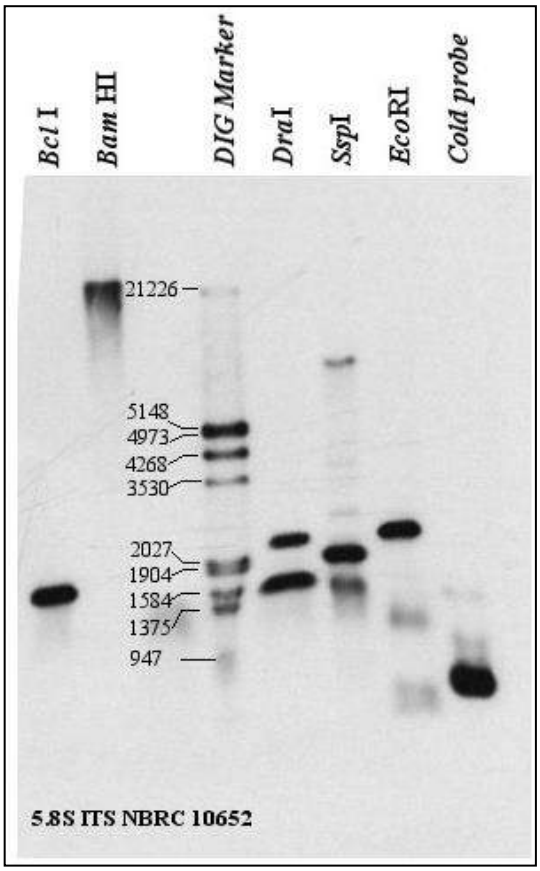


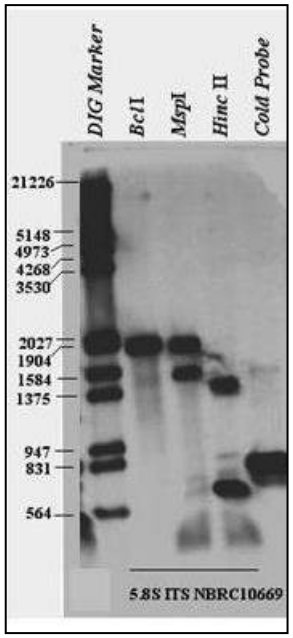
Figure 7.6. Detection of heterogeneous copies of 5.8S ITS region by Southern blot in NBRC0495 (upper) and NBRC10652 (lower).

The genomic DNA of NBRC10652 was digested with *Bcl*I, *Bam*HI, *Ban*I, *Dra*I, *Ssp*I, and *Eco*RI. Restriction digestion with *Bcl*I gave same band size as obtained from NBRC0495. We choose *Dra*I for detecting two heterogeneous copies of ITS region in strains NBRC0495. *Dra*I has two sites in ITS region of NBRC10652 copy 1, one at 221bp other at 606bp. The ITS region of NBRC10652 copy 2 has only one site at 224bp. Both copies of the 26S rDNA region of this strain is devoid of *Dra*I site. Immediate upstream in *Z. rouxii* CBS 732^{T}, *Dra*I site is present at 6430bp. The restriction digestion of genomic DNA, as expected, should give two bands, one of 454bp (606bp-221bp=385bp) related to copy 1 and other of 1458bp related to copy 2 (that recognize sequence between restriction site at 224 and upstream conserved *Dra*I site at 6430bp as in *Z. rouxii* CBS 732^{T}). However, the copy 1 specific band, due to its small length, could not be seen. Instead an additional band of approximately 2300bp was



seen. The slightly greater size of copy 2 specific band could be attributed to problem related to electrophoretic run.

For detection heterogeneous copies of ITS region in strain NBRC10669, we used three restriction enzymes, *Bcl*I, *Msp*I (C^CG_G), *Hinc*II. Restriction digestion with *Bcl*I gave same band size as obtained from NBRC0495 and NBRC10652. With *Msp*I restriction enzyme, we were able to discriminate between two copies of ITS region in NBRC10669. NBRC10669 copy 1 region contains one *Msp*I restriction site at 367bp. Copy 2 of ITS of this strain lack this site. Copy 1 of 26S rDNA region in this strain has two *Msp*I restriction sites, one at 19bp and other at 538bp. Similarly, copy 2 of 26S rDNA of this strain also two *Msp*I sites, one at 21bp and other at 518bp.



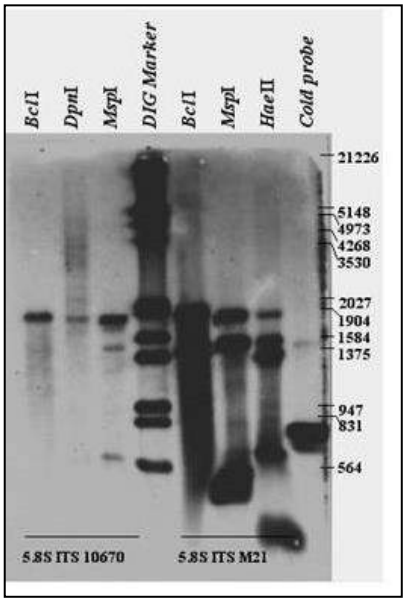
Immediate downstream to ITS region of *Z. rouxii* CBS 732^{T}, there is a *Msp*I site at 3612bp. If we assume this site to be conserved, the probe will detect sequence from 367bp in ITS copy 1 to downstream conserved site at 3612 and will give band of 1503bp (NBRC10669 ITS copy 1 specific). As *Msp*I site is absent in ITS copy 2, the probe will detect sequence from 19 or 21bp in 26S rDNA of NBRC10669 to downstream conserved site at 3612 and will give band of 1831 or 1833bp (NBRC10669 ITS copy 2 specific). We obtained two bands of expected size that suggest presence of two heterogeneous copies of ITS region in strain NBRC10669.

Figure 7.6. Detection of heterogeneous copies of 5.8S ITS region by Southern blot in NBRC10669

For detection heterogeneous copies of ITS region in strain NBRC10670, we used three restriction enzymes, *Bcl*I, *Dpn*I, *Msp*I. Restriction digestion with *Bcl*I gave same band size as obtained from NBRC0495, NBRC10652, and NBRC10669. With *Msp*I restriction enzyme, we were able to discriminate between two copies of ITS region in NBRC10670. NBRC10670 copy 1 region contains one *Msp*I restriction site at 367bp. Copy 2 of ITS of this strain lack this site. Copy 1 of 26S rDNA region in this strain has one *Msp*I restriction sites at 17bp. Instead copy 2 of 26S

rDNA of this strain has two *Msp*I sites, one at 19bp and other at 516bp. Immediate downstream to ITS region of *Z. rouxii* CBS 732^{T}, there is a *Msp*I site at 3612bp. If we assume this site to be conserved, the probe will detect sequence from 367bp in ITS copy 1 to downstream conserved site at 3612 and will give band of 1503bp (NBRC10670 ITS copy 1 specific). As *Msp*I site is absent in ITS copy 2, the probe will detect sequence from 17 or 19bp in 26S rDNA of NBRC10670 to downstream conserved site at 3612 and will give band of 1829 or 1829bp (NBRC10670 ITS copy 2 specific). We obtained two bands of expected size that suggest presence of two heterogeneous copies of ITS region in strain NBRC10670.

Figure 7.6. Detection of heterogeneous copies of 5.8S ITS region by Southern blot in NBRC10670 and M21



Heterogeneous copies of ITS region in strain M21 were detected using three restriction enzymes, *Bcl*I, *Msp*I and *Hae*II. Restriction digestion with *Bcl*I gave same band size as obtained from NBRC0495, NBRC10652, NBRC10669 and NBRC10670. With *Msp*I restriction enzyme, we were able to discriminate between two copies of ITS region in M21. M21 copy 1 region contains one *Msp*I restriction site at 371bp. Copy 2 of ITS of this strain lack this site. Strain M21 is not heterogeneous with respect to its 26S rDNA region and contain two *Msp*I restriction sites at 6bp and 503bp. Immediate downstream to ITS region of Z. rouxii CBS 732, there is a *Msp*I site at 3612bp. If we assume this site to be conserved, the probe will detect sequence from 371bp in ITS copy 1 to downstream conserved site at 3612 and will give band of 1507bp (M21 ITS copy 1 specific). As *Msp*I site is absent in ITS copy 2, the probe will detect sequence from 6bp in its 26S rDNA to downstream conserved site at 3612 and will give band of 1818bp (M21 ITS copy 2 specific). We obtained two bands of expected size that suggest presence of two heterogeneous copies of ITS region in strain M21.

Detection of heterogeneity in 26S rDNA region

Heterogeneous copies of 26S rDNA region in strain NBRC0495 were detected using three restriction enzymes, *Bcl*I, *Mse*I (T^TA_A) and *Hae*II. Restriction digestion with *Bcl*I gave same band size as obtained earlier. With *Mse*I restriction enzyme, we were able to discriminate between two copies of 26S rDNA region in NBRC0495. 26S rDNA of NBRC0495 copy 1 region contains one *Mse*I restriction site at 577bp. Copy 2 of 26S rDNA in this strain lacks this site. Strain NBRC0495 also contains four *Mse*I restriction sites at 53, 136, 206 and 590bp in its copy 1 of 26S rDNA region. In copy 2 of 26S rDNA region these sites were present, except the site at 590bp. Immediate upstream to 26S rDNA region of *Z. rouxii* CBS 732^{T}, there is a *Msp*I site at 6411bp. If we assume this site to be conserved, the probe will detect sequence from 206bp in ITS copy 2 to upstream *Mse*I site at 577bp in 26S rDNA copy 1 and will give band of 1047bp (NBRC0495 26S rDNA copy 1 specific). Similarly, *Msp*I site at 590bp in ITS copy 1 of this strain and assumed conserved *Mse*I site upstream to 26S rDNA at 6411bp in *Z. rouxii* CBS 732^{T} will give band of 1073bp (NBRC0495 26S rDNA copy 2 specific). As both bands (NBRC0495 26S rDNA copy 1 and copy 2 specific) have almost same size there is likelihood to obtain to intense band at this position. We obtained one intense band of expected size that suggests presence of two heterogeneous copies of 26S rDNA region in strain NBRC0495. Besides this, an additional band of approximate size 600bp was seen that could be obtain if ITS copy 1 from this strain pairs with 26S rDNA copy 1 sequence. However, we couldn't reason out the absence of similar band that correspond to ITS copy 2 sequence pairing with 26S rDNA copy 2 sequence.

For detection heterogeneous copies of 26S rDNA region in strain NBRC0505, we used four restriction enzymes, *Bcl*I, *Ban*I, *Hha*I and *Sac*I (G_AGTG^C). Restriction digestion with *Bcl*I gave same band size as obtained earlier at position 1900bp. With *Sac*I restriction enzyme, we were able to discriminate between two copies of 26S rDNA region in NBRC0505. 26S rDNA copy 1 region of NBRC0505 contains one *Sac*I restriction site at 56bp, which is absent in copy 2. *Sac*I is rare cutter and it cuts the *Z. rouxii* CBS 732^T array at only two positions, 4173bp and 7453bp. The Copy 2 of 26S rDNA of this strain could be detected by probe that will hybridize to sequence between *Sac*I conserved sites at 4173bp and 7453bp resulting in a band size of 3280bp. Probe could not detect the NBRC0505 copy 1 specific band (estimated size 1307) that result from *Sac*I conserved site at 4173bp and at 56bp in copy 1 of 26S rDNA of this strain. The possible reason could be that probe could not efficient hybridize with only 56bp sequence related to copy 1. We obtained another band of estimated size 7000bp that is not related with copy 2. This band could be related with copy 2. This band could be related with copy 1; however, it is possible only if the assumed conserved sites are not conserved in copy 1 harboring rDNA array in strain NBRC0505.

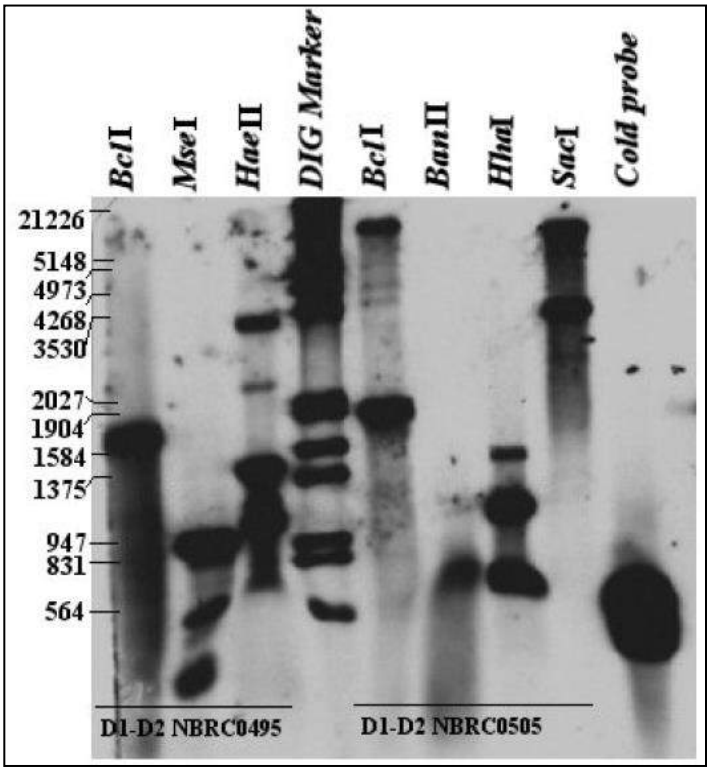


Figure 7.7. Detection of heterogeneous copies of 26S D1-D2 region by Southern blot in NBRC0495 and NBRC0505

For detection heterogeneous copies of 26S rDNA region in strain NBRC10652, we used three restriction enzymes, *Bcl*I, *Hph*I (5'-GGTGA$(N)_8$^-3' and 3'-CCACT$(N)_7$_-5') and *Ahd*I. Restriction digestion with *Bcl*I gave same band size as

obtained earlier at position 1900bp. With *Hph*I restriction enzyme, we were able to discriminate between two copies of 26S rDNA region in NBRC10670. With *Hph*I restriction we expected two bands. Probe will detect one band of 579/580bp size that corresponds to sequence between *Hph*I site at 199/198 (respectively in 26S rDNA copy 1 and copy 2) and upstream conserved *Hph*I site at position 6202bp in Z. rouxii rDNA. Another band of size 886/885bp would be expected that covers the sequence that range from 178/177bp (respectively in 26S rDNA copy 1 and copy 2) to downstream conserved *Hph*I site at 4716bp in *Z. rouxii* CBS 732^{T} rDNA. The latter band was not seen suggesting that the downstream *Hph*I site is not conserved in strain NBRC10652. Besides this, we found an additional band of approximated size 400bp suggesting that the strain is heterogeneous in its sequence of rDNA with respect to *Z. rouxii* CBS 732^{T} rDNA.

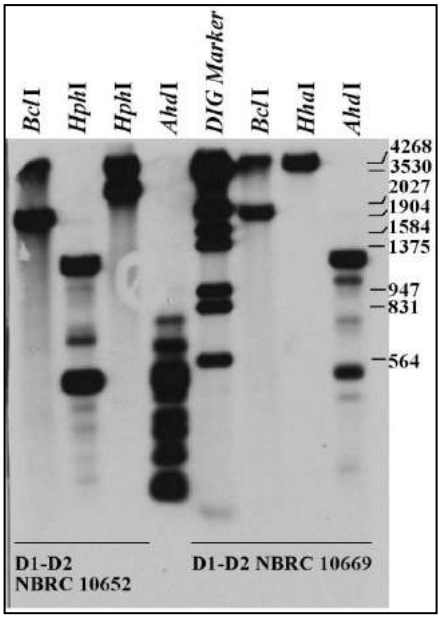
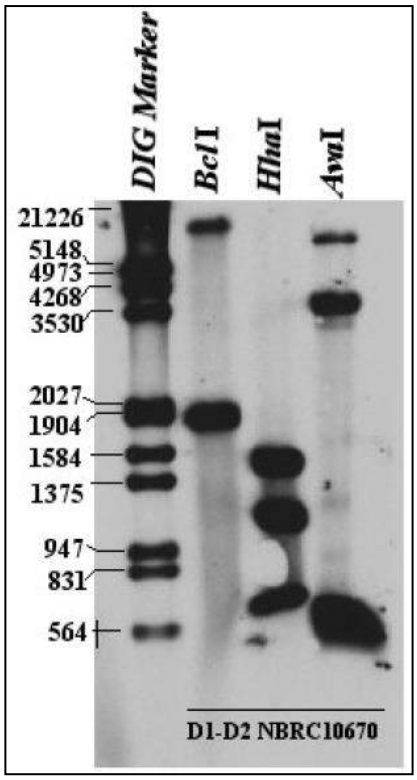


Figure 7.7. Detection of heterogeneous copies of 26S D1-D2 region by Southern blot in NBRC10652 and NBRC10669

For detection heterogeneous copies of 26S rDNA region in strain NBRC10669, we used three restriction enzymes, *Bcl*I, *Hha*I and *Ahd*I (GACNN_N^NNGTC). Restriction digestion with *Bcl*I gave same band size as obtained earlier at position 1900bp. With *Ahd*I we discriminated the two copies of 26S rDNA gene in strain NBRC10669. Strain NBRC10669 harbors one *Adh*I restriction site at position 556bp in its 26S rDNA copy 2, while the copy 1 is devoid of this restriction site. This enzyme is highly rare cutter and in the complete *Z. rouxii* CBS 732^{T} rDNA it cuts just a single site at position 5972bp that lies in its 26S rDNA region. For this reason, the restriction digestion of genomic DNA of strain NBRC10669 is expected to yield a band of 9940bp size corresponding to copy 2 of 26S rDNA of strain NBRC10669. However, we obtained band at 1100, 1000 and 500bp. As there is no restriction site for this enzyme in the ITS region copy 1 or

copy 2 of strain NBRC10669, we speculated that there are at least two *Ahd*I sites upstream to 26S rDNA in strain NBRC10669, one site lies in rDNA stretch that harbor copy 1 and other site at different position in rDNA stretch that harbors copy 2.

Figure 7.7. Detection of heterogeneous copies of 26S D1-D2 region by Southern blot in NBRC10670



For detection heterogeneous copies of 26S rDNA region in strain NBRC10670, we used three restriction enzymes, *Bcl*I, *Hha*I and *Ava*I (C^YCGR_G). Restriction digestion with *Bcl*I gave same band size as obtained earlier at position 1900bp. With *Ava*I restriction enzyme, we were able to discriminate between two copies of 26S rDNA region in NBRC10670. 26S rDNA copy 1 region of NBRC10670 contains three AvaI restriction site at 88, 475 and 515bp. The copy 2 has only one site at 86bp. *Ava*I cuts the *Z. rouxii* CBS 732^{T} rDNA array at 1441bp downstream to its ITS region and at 9225bp upstream to its 26S rDNA region. Probe detected copy 1 specific band that result from two *Ava*I restriction sites (one at 88bp and other at 475bp) in 26S rDNA copy 2 of this strain. The expected copy 1 specific band size was 387bp; however, band signal was too much high that hampers correct band size estimation. The probe did not detect the sequence that lies from *Ava*I site at 515bp (in 26S rDNA copy 1 of this strain) and upstream conserved *Ava*I site at 9225bp suggesting that this site is either absent or not conserved in NBRC10670 rDNA that harbor 26S rDNA copy 1. The copy 2 of 26S rDNA of strain NBRC10670 could be detected by probe that will hybridize to sequence between *Ava*I site at 86bp in its 26S rDNA region and conserved upstream AvaI sites at 9225bp resulting in a band size of 3713bp. Besides this, we obtained an additional band of approximated size around 6000bp. This band could result only if the *Ava*I site

at 1441bp downstream to ITS region is absent or not conserved. This band could be related to either one copy or the both copies of 26S rDNA region in strain NBRC10670.

7.4. Conclusion

So far, little attention has been devoted to intragenomic rDNA heterogeneity in yeasts because the ascomycetous yeasts are generally believed to have uniform rDNA repeat arrays due to homogenisation by concerted evolutions. In fact this uniformity has become a sort of basic tenet of molecular taxonomy and is routinely exploited in taxonomic classification of strains, species delimitation and mapping of phylogenetic relationships. Exceptions are certain postulated hybrid species (e.g. *Pichia sorbitophila*) and the alloploid and chimerical strains arisen from rare interspecies mating (e.g. in the genus *Saccharomyces*) which have different versions of LSU rDNA repeat units inherited from homozygous strains of euploid parental species (Dunn and Sherlock, 2008; Gordan and Wolfe, 2008; Sipiczki, 2008; Louis et al., 2012; Morales and Dujon, 2012). Postulated interspecific hybrids heterozygous for rDNA can also be produced under laboratory conditions but they are either sterile (Solieri et al., 2008b) or genetically unstable (Pfliegler et al., 2012) which can positively or negatively affect their chances to survive under natural and experimental conditions (**refer Chapter 5**). In principle, some strains examined here could also be postulated heterozygous hybrids as they are diploid or aneuploid; however, their ability to produce ascospores directly from cells with or without conjugation has to be ascertained (Soleiri et al., 2013a, b). However, the postulated hybrids usually have only two parental alleles (Gordan and Wolfe, 2008), whereas the *Zygosaccharomyces* strains examined here have at least two divergent 5.8S ITS and 26S D1-D2 sequences in their rRNA. We consider it more likely that they have mosaic arrays of rDNA repeats containing (most probably) paralogous D1-D2 segments differing by some substitutions.

Strains	ITS Clones (based on HaeIII Profile)	Clone designation	Heterogeneity ITS	Reference
CBS 732	-		no	Chapter 6, 7
CBS 736	-		-	Chapter 6, 7
NCYC 3042	-		-	Chapter 6, 7
ABT 301	2	copy 2	yes	Soleiri et al., 2013a, b
	10	copy 3		
	11	copy 1		
ATCC 42981	-		yes	Solieri et al., 2013b
	-			
CBS 4837	**13**	copy 2	yes	Soleiri et al., 2013b
	26	copy 1		
	29	copy 3		
CBS 4838	**1**	copy 2	yes	Soleiri et al., 2013b
	3	copy 3		
	4	copy 1		
OUT 7136	-	-	no	Soleiri et al., 2013b
NBRC0495	1, **2**, 6, 7, 8, 10, 13, 14, 15	copy 1	yes	Chapter 6, 7
	3, 4, 5, 9, 11, 12	copy 2		
NBRC0505	-		no	Chapter 6, 7
		-		
NBRC0525	not performed	-	no	Chapter 6, 7
NBRC0845	All similar	-	no	Chapter 6, 7
NBRC10652*	1-**6**, 8-15	copy 1	yes	Chapter 6, 7
	7	copy 2		
NBRC10668	-		no	Chapter 6, 7
NBRC10669	1-6, 8, 9, **10**, 12-15	copy 1	yes	Chapter 6, 7
	7, **11**	copy 2		
		-		
NBRC10670	1-**6**, 8-15	copy 1	yes	Chapter 6, 7
	7	copy 2		
		-		
NBRC10672	1, 2, 3, 5, 6, 7, 8, 9, 10, 12, 13, 15	not performed	yes	Chapter 6, 7
	4, 11	not performed		
M21*	1, 3, 4, 5, 6, 7, 8, 12, 14, 15	copy 1	yes	Chapter 6, 7
	2, 9, 10, 11, 13	copy 2	yes	Chapter 6, 7
		-	yes	Chapter 6, 7
* Strain M21 *Hae* III RFLP based screening of clones showed that clones representing copy 3 (similar to ABT 301 5.8S ITS copy 3) are missing.				

Supplementary Table S7.1. Clone designation for representative and singlet strains chosen based on 5.8S ITS-RFLP (**refer Chapter 6**)

Final conclusion

The present thesis contributes to better understand the relationships between phenotype and genome-level traits in the poorly studied, non-conventional yeasts belonging to the *Zygosaccharomyces rouxii* complex. *Z. rouxii* complex comprises of genetically and phenotypically divergent lineages that are characterized by gene copy number variations, and change in ploidy level and chromosome number (Solieri et al., 2013b). I hypothesized that the life cycle and breeding could be the evolutionary drivers of this wide spread genetic and functional variability in this complex; and demonstrated how genetic and functional variability are impacted by peculiar life cycle and mating system. Stress environment is known to induce genetic and phenotypic variability in yeasts (Berman & Hadany, 2012). I demonstrate that mating system in *Z. rouxii* complex is a hypermutagenic process that contributes to stress adaptation by generating progenies with different genetic and phenotypic outcomes (Solieri et al., 2014), of which some might be industrially important. The generation of divergent lineages could be a successful strategy under stress to increase the probability to achieve descendants improved in adaptation to hostile environments.

Bibliography

1) Aguiar C, Lucas C (2000) Yeasts killer/sensitivity phenotypes and halotolerance. Food Technol Biotechnol 38(1): 39–46.

2) Aguilera A, García-Muse T (2013) Causes of genome instability. Annu Rev Genet 47: 1-32.

3) Albertin W, Marullo P, Aigle M, Bourgais A, Bely M, et al. (2009) Evidence for autotetraploidy associated with reproductive isolation in *Saccharomyces cerevisiae*: towards a new domesticated species. J Evol Biol 22: 2157–2170.

4) Albertin W, Marullo P (2012) Polyploidy in fungi: evolution after whole genome duplication. Proc R Soc Lond B 279: 2497–2509.

5) Alby K, Schaefer D, Bennett RJ (2009) Homothallic and heterothallic mating in the opportunistic pathogen *Candida albicans*. Nature 460: 890–893.

6) Alcoba-Flórez J, Arévalo-Morales MDP, Pérez-Roth E, Laich F, Rivero-Pérez B (2007) Yeast molecular identification and typing. In: Méndez-Vilas AA, editor, Communicating Current Research and Educational Topics and Trends in Applied Microbiology, Spain: Formatex. Pp. 535-546.

7) Almeida AJ, Matute DR, Carmona JA, Martins M, Torres I (2007) Genome size and ploidy of Paracoccidioides brasiliensis reveals a haploid DNA content: Flow cytometry and GP43 sequence analysis. Fungal Gen Biol 44: 25-31.

8) Alonso-Vargas R, Garaizar J, Pontón J, Quindós G (2000) Utility of random amplified polymorphic DNA in the discrimination between *Candida albicans* and *Candida dubliniensis.* Rev Iberoam Micol 17: 10-13.

9) Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSIBLAST: a new generation of protein database search Programs. Nucleic Acids Res 25: 3389–3402.

10) Anderson JB, Sirjusingh C, Ricker N (2004) Haploidy, diploidy and evolution of antifungal drug resistance in *Saccharomyces cerevisiae*. Genetics 168: 1915–1923.

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

12) Andrade MJ, Rodríguez M, Casado EM, Bermúdez E, Córdoba JJ (2009) Differentiation of yeasts growing on dry-cured Iberian ham by mitochondrial DNA restriction analysis, RAPD-PCR and their volatile compounds production. Food Microbiol 26: 578–586.

13) Appel DJ, Gordon TR (1995) Intraspecific variation within populations of *Fusarium oxysporum* based on RFLP analysis of the intergenic spacer region of the rDNA. Exp Mycol 19: 120–128.

14) Araki H, Jearnpipatkul A, Tatsumi H, Sakurai T, Ushio K, et al. (1985) Molecular and functional organization of yeast plasmid pSR1. J Mol Biol 182: 191-203.

15) Ariño J, Ramos J, Sychrová H (2010) Alkali metal cation transport and homeostasis in yeasts. Microbiol Mol Biol Rev 74: 95-120.

16) Azumi M, Goto-Yamamoto N (2001) AFLP analysis of type strains and laboratory and industrial strains of *Saccharomyces* sensu stricto and its application to phenetic clustering. Yeast 18: 1145–1154.

17) Bakhrat A, Jurica MS, Stoddard BL, Raveh D (2004) Homology modeling and mutational analysis of HO endonuclease of yeast. Genetics 166: 721–728.

18) Baleiras Couto MM, van der Vossen JMBM, Hofstra H, Huis in't Veld JHJ (1994) RAPD analysis: a rapid technique for differentiation of spoilage yeasts. Int J Food Microbiol 24: 249–260.

19) Baleiras Couto MM, Eijsma B, Hofstra H, Huis in't Veld JHJ, van der Vossen JM (1996). Evaluation of molecular typing techniques to assign genetic diversity among *Saccharomyces cerevisiae* strains. Appl Environ Microbiol 62: 41–46.

20) Balloux F, Lehmann L, de Meeûs T (2003) The population genetics of clonal and partially clonal diploids. Genetics 164: 1635–1644.

21) Bañuelos MA, Sychrová H, Bleykasten-Grosshans C, Souciet JL, Potier S (1998) The Nha1 antiporter of *Saccharomyces cerevisiae* mediates sodium and potassium efflux. Microbiology 144: 2749-2758.

22) Baranyi J, Pin C (1999) Estimating bacterial growth parameters by means of detection times. Appl Environ Microbiol 65:732–736.

23) Barker BTP (1901). A conjugating 'yeast'. Philos Trans R Soc Lond B Biol Sci 194: 467–485.

24) Barnett JA, Payne RW, Yarrow D (1983) Yeasts: Characteristics and Identification. Cambridge: Cambridge University Press.

25) Barnett JA, Payne RW, Yarrow D (2000) Yeasts: Characteristics and Identification, 3rd edition, Cambridge UK: Cambridge University Press.

26) Baumgarth N, Roederer MA (2000) A practical approach to multicolor flow cytometry for immunophenotyping. J. Immunol. Methods 243: 77–97.

27) Belfort M, Roberts RJ (1997) Homing endonucleases: keeping the house in order. Nucleic Acids Res 25: 3379–3388.

28) Belloch C, Barrio E, Uruburu F, García DD, Querol A (1997) Characterization of four species of the genus *Kluyveromyces* by mitochondrial análisis. System Appl Microbiol 20: 397-408.

29) Belloch C, Barrio E, García MD, Querol A (2000) Phylogenetic reconstruction of the yeast genus *Kluyveromyces*: restriction map analysis of the 5.8S rRNA gene and the two ribosomal transcribed spacers. Syst Appl Microbiol 21: 266–273.

30) Belloch C, Pérez-Torrado R, González SS, Pérez-Ortín JE, García-Martínez J, et al. (2009) Chimeric genomes of natural hybrids between *Saccharomyces cerevisiae* and *Saccharomyces kudriavzevii*. Appl Environ Microbiol 75: 2534–2544.

31) Bendall SC, Nolan GP, Roederer M, Chattopadhyay PK (2012) A deep profiler's guide to cytometry. Trends Immunol 33: 323–332.

32) Benito B, Garciadeblás B, Rodríguez-Navarro A (2002) Potassium- or sodium-efflux ATPase, a key enzyme in the evolution of fungi. Microbiology 148: 933-941.

33) Berman J, Hadany L (2012) Does stress induce (para)sex? Implications for *Candida albicans* evolution. Trends Genet 28: 197-203.

34) Bhardway S, Sutar R, Bachhatawat AK, Singhi S, Chakrabarti A (2007) PCR based identification and strain typing of *Pichia anomala* using the ribosomal intergenic spacer region IGS1. J Med Microbiol 56: 185–189.

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

36) Billiard S, López-Villavicencio M, Hood ME, Giraud T (2012) Sex, outcrossing and mating types: unsolved questions in fungi and beyond. J Evol Biol 25: 1020–1038.

37) Birky CW (1996) Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. Genetics 144: 427–437.

38) Blackwell JR, Horgan R (1991) A novel strategy for the production of a highly expressed recombinant protein in an active form. FEBS Lett 295:10-12.

39) Bochner BR, Gadzinski P, Panomitros E (2001). Phenotype MicroArrays for high-throughput phenotypic testing and assay of gene function. Genome Res 11: 1246-1255.

40) Bon E, Casaregola S, Blandin G, Llorente B, Neuvéglise C, et al. (2003) Molecular evolution of eukaryotic genomes: hemiascomycetous yeast spliceosomal introns. Nucleic Acids Res 31: 1121–1135.

41) Borglin S, Joyner D, de Angelis KM, Khudyakov J, D'haeseleer P, Joachimiak MP, Hazen T (2012) Application of phenotypic microarrays to environmental microbiology. Curr Opin Biotechnol 23:41–48.

42) Botterel F, Desterke C, Costa C, Bretagne S (2001) Analysis of microsatellite markers of *Candida albicans* used for rapid typing. J Clin Microbiol 39(11): 4076-4081.

43) Bovers M, Hagen F, Kuramae EE, Diaz MR, Spanjaard L, et al. (2006) Unique hybrids between the fungal pathogens *Cryptococcus neoformans* and *Cryptococcus gattii.* FEMS Yeast Res 6: 599–607.

44) Boy-Marcotte E, Pierrot M, Bussereau F, Boucherie H, Jacquet M (1998) Msn2p and Msn4p control a large number of genes induced at the diauxic transition which are repressed by cyclic AMP in *Saccharomyces cerevisiae*. J Bacteriol 180: 1044-1052.

45) Branduardi P (2002) Molecular cloning and sequence analysis of the *Zygosaccharomyces bailii* HIS3gene encoding the imidazole glycerolphosphate dehydratase. Yeast 19: 1165-1170.

46) Branduardi P, Valli M, Brambilla L, Sauer M, Alberghina L, et al. (2004) The yeast *Zygosaccharomyces bailii*: a new host for heterologous protein production, secretion and for metabolic engineering applications. FEMS Yeast Res 4: 493-504.

47) Breuer U, Harms H (2006) *Debaryomyces hansenii*- An extremophilic yeast with biotechnological potential. Yeast 23: 415-437.

48) Brewster JL, de Valoir T, Dwyer ND, Winter E, Gustin MC (1993) An osmosensing signal transduction pathway in yeast. Science 259: 1760–1763.

49) Bubnick M, Smulian AG (2007) The MAT1 locus of *Histoplasma capsulatum* is responsive in a mating type-specific manner. Eukaryot Cell 6: 616-621.

50) Butinar L, Santos S, Spencer-Martins I, Oren A, Gunde-Cimerman N (2005) Yeast diversity in hypersaline habitats. FEMS Microbiol Lett 244: 229–234.

51) Butler G, Kenny C, Fagan A, Kurischko C, Gaillardin C et al. (2004) Evolution of the MAT locus and its HO endonuclease in yeast species. Proc Natl Acad Sci USA 101: 1632–1637.

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

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

54) Byrne KP, Wolfe KH (2007) Consistent patterns of rate asymmetry and gene loss indicate widespread neofunctionalization of yeast genes after whole-genome duplication. Genetics 175: 1341–1350.

55) Cai J, Robert IN, Collins MD (1996) Phylogenetic relationships among members of the ascomycetous yeast genera *Brettanomyces*, *Debaryomyces*, *Dekkera*, and *Kluyveromyces* deduced by small-subunit rRNA gene sequences. Int J Syst Bacteriol 46: 542-549.

56) Cai J, Zhao R, Jiang H, Wang W (2008) *De novo* origination of a new protein-coding gene in *Saccharomyces cerevisiae*. Genetics 179: 487–496.

57) Cao X, Hou L, Lu M, Wang C & Zeng B (2009) Genome shuffling of *Zygosaccharomyces rouxii* to accelerate and enhance the flavor formation of soy sauce. J Sci Food Agric 90: 281–285.

58) Carle GF, Olson MV (1985) An electrophoretic karyotype for yeast. Proc Natl Acad Sci USA 82: 3756–3760.

59) Casaregola S, Nguyen HV, Lepingle A, Brignon P, Gendre F, Gaillardin C (1998) A family of laboratory strains of *Saccharomyces cerevisiae* carry rearrangements involving chromosomes I and III. Yeast 14: 551–564.

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

61) Casas E, De Ancos B, Valderrama MJ, Cano P, Peinado JM (2004) Pentadiene production from potassium sorbate by osmotolerant yeasts. Int J Food Microbiol 94: 93–96.

62) Casey GP, Xiao W, Rank GH (1988) Application of pulsed field chromosome eletrophoresis in the study of chromosome XIII and the electrophoretic karyotype of industrial strains of *Saccharomyces* yeast. J Inst Brew 94: 239–243.

63) Causton HC, Ren B, Koh SS, Harbison CT, Kanin E, et al. (2001) Remodeling of yeast genome expression in response to environmental changes. Mol Biol Cell 12: 323-337.

64) Chen KW, Lin YH, Li S (2005) Comparison of four molecular typing methods to assess genetic relatedness of *Candida albicans* isolates in Taiwan. J Med Microbiol 54: 249-58.

65) Chester M, Leitch AR, Soltis PS, Soltis DE (2010) Review of the application of modern cytogenetic methods (FISH/GISH) to the study of reticulation (polyploidy/hybridisation). Genes 1: 166–192. (doi:10.3390/genes1020166)

66) Cliften PF, Fulton RS, Wilson RK, Johnston M (2006) After the duplication: gene loss and adaptation in *Saccharomyces* genomes. Genetics 172: 863–872.

67) Codón AC, Korhola M (1998) Chromosomal polymorphism and adaptation to specific industrial environments of *Saccharomyces* strains. Appl Microbiol Biotechnol 49: 154–163.

68) Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. Nature Rev Genet 9: 938–950.

69) Connell LB, Redman R, Rodriguez R, Barrett A, Iszard M, Fonseca A (2010) *Dioszegia antarctica* sp. nov. and *Dioszegia cryoxerica* sp. nov., psychrophilic basidiomycetous yeasts from polar desert soils in Antarctica. Int J Syst Evol Microbiol 60: 1466–1472.

70) Cyert MS, Kunisawa R, Kaim D, Thorner J (1991) Yeast has homologs (*CNA1* and *CNA2* gene products) of mammalian calcineurin, a calmodulin-regulated phosphoprotein phosphatase. Proc Natl Acad Sci USA 88: 7376-7380.

71) Dakal TC, Solieri L, Giudici P (2014) Adaptive response and tolerance to sugar and salt stress in food yeast *Zygosaccharomyces rouxii*. Int J Food Microbiol. Manuscript accepted.

72) Dalle F, Dumont L, Franco N, Mesmacque D, Caillot D, et al. (2003) Genotyping of *Candida albicans* oral strains from healthy individuals by polymorphic microsatellite locus analysis. J Clin Microbiol 41(5): 2203-2205.

73) Dato L, Sauer M, Passolunghi S, Porro D, Branduardi P (2008) Investigating the multibudded and binucleate phenotype of the yeast *Zygosaccharomyces bailii* growing on minimal medium. FEMS Yeast Res 8: 906–915.

74) Deak T, Beuchat LR (1996) Handbook of Food Spoilage Yeasts. Boca Raton, FL: CRC Press.

75) Deak R, Bodai L, Aarts HJ, Maraz A (2004) Development of a novel, simple and rapid molecular identification system for clinical *Candida* species. Med Mycol 42(4): 311-318.

76) Deak T (2007) Handbook of food spoilage yeasts. Boca Raton, FL: CRC Press.

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

78) Dendis M, Horyath R, Michalek J, Ruzicka F, Grijalva M, et al. (2003) PCR-RFLP detection and species identification of fungal pathogens in patients with febrile neutropenia. Clin Microbiol Infect 9(12): 1191-202.

79) Despons L, Wirth B, Louis VL, Potier S, Souciet JL (2006) An evolutionary scenario for one of the largest yeast gene families. Trends Genet 22: 10–15.

80) Despons L, Baret PV, Frangeul L, Louis VL, Durrens P, et al. (2010) Genome-wide computational prediction of tandem gene arrays: application in yeasts. BMC Genomics 11: 56.

81) Dettman JR, Sirjusingh C, Kohn LM, Anderson JB (2007) Incipient speciation by divergent adaptation and antagonistic epistasis in yeast. Nature 447: 585-588.

82) de Barros Lopes M, Rainieri S, Henschke PA, Langridge P (1999) AFLP fingerprinting for analysis of yeast genetic variation. Int J Syst Bacteriol 49: 915–924.

83) de Clare M, Pir P, Oliver SG (2011) Haploinsufficiency and the sex chromosomes from yeasts to humans. BMC Biol 9:15.

84) de Jonge P, de Jongh FCM, Meijers R, Steensma HY, Scheffers WA (1986) Orthogonal-field-alternation gel electrophoresis banding patterns of DNA from yeast. Yeast 2: 193–204.

85) de Montigny J, Straub M, Potier S, Tekaia F, Dujon B, et al. (2000) Genomic exploration of the hemiascomycetous yeasts: 8. *Zygosaccharomyces rouxii*. FEBS Lett 487(1): 52–55.

86) de Nadal E, Alepuz PM, Posas F (2002) Dealing with osmostress through MAP kinase activation. EMBO Rep 3: 735-740.

87) Dhar R, Sagesser R, Weikert C, Yuan J, Wagner A (2011) Adaptation of *Saccharomyces cerevisiae* to saline stress through laboratory evolution. J Evol Biol 24: 1135–1153.

88) Diaz MR, Fell JW (2000) Molecular analysis of the IGS & ITS regions of the rDNA of the psychrophilic yeast in the genus Mrakia. Antonie van Leeuwenhoek 77: 7–12.

89) Diaz MR, Boekhout T, Kiesly T, Fell JW (2005) Comparative analysis of the intergenic spacer regions population structure of the species complex of the pathogenic yeasts *Cryptococcus neoformans*. FEMS Yeast Res 12: 1129–1140.

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

91) Diogo D, Bouchier C, d'Enfert C, Bougnoux ME (2009) Loss of heterozygosity in commensal isolates of the asexual diploid yeast *Candida albicans*. Fungal Genet Biol 46: 159–168.

92) Doniger SW, Kim HS, Swain D, Corcuera D, Williams M, et al. (2008) A catalog of neutral and deleterious polymorphism in yeast. PLoS Genet 4(8): e1000183. doi:10.1371/journal.pgen.1000183

93) Dover G (1982) Molecular drive: cohesive mode of species evolution. Nature 9: 111–116.

94) Dragosits M, Stadlmann J, Graf A, Gasser B, Maurer M, et al. (2010) The response to unfolded protein is involved in osmotolerance of *Pichia pastoris*. BMC Genomics 11: 207. doi: 10.1186/1471-2164-11-207.

95) Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, et al. (2004) Genome evolution in yeasts. Nature 430: 35–44.

96) Dujon, B (2010) Yeast evolutionary genomics. Nat Rev Genet 11: 512–524. doi: 10.1038/nrg2811

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

98) Durrens P, Nikolski M, Sherman D (2008) Fusion and fission of genes define a metric between fungal genomes. PLoS Comput Biol 4: e1000200.

99) Eilers PHC, Marx BD (1996) Flexible smoothing with b-splines and penalties. Statist Sci 11: 89–121.

100) Ekino K, Kwon I, Goto M, Yoshino S, Furukawa K (1999) Functional analysis of *HO* gene in delayed homothallism in *Saccharomyces cerevisiae* wy2. Yeast 15: 451–458.

101) Erdélyi B, Szabó A, Birincsik L, Hoschke Á (2004) Process development of methylenedioxyphenyl-acetone chiral bioreduction. J Mol Catal B Enzym 29: 195–199.

102) Erdélyi B, Szabó A, Seres G, Birincsik L, Ivanics J, Szatzker G, Poppe L (2006) Stereoselective production of (S)-1-aralkyl- and 1-arylethanols by freshly harvested and lyophilized yeast cells. Tetrahedron: Asymmetry 17: 268–274.

103) Ergon MC, Gulay Z (2005) Molecular epidemiology of *Candida* species isolated from urine at an intensive care unit. Mycoses 48(2): 126-231.

104) Esteve-Zarzoso B, Belloch C, Uruburu F, Querol A (1999) Identification of yeasts by RFLP analysis of the 5.8S rRNA gene and the two ribosomal internal transcribed spacers. Intern Jour Syst Bacteriol 49: 329-337.

105) Esteve-Zarzoso B, Zorman T, Belloch C, Querol A (2003) Molecular Characterisation of the species of the genus *Zygosaccharomyces*. Syst Appl Microbiol 26: 404–411.

106) Ezov TK, Chang SL, Frenkel Z, Segre AV, Bahalul N, et al. (2010) Heterothallism in *Saccharomyces cerevisiae* isolates from nature: effect of *HO* locus on the mode of reproduction. Mol Ecol. 19: 121-131.

107) Fabre E, Muller H, Therizols P, Lafontaine I, Dujon B, et al. (2005) Comparative genomics in hemiascomycete yeasts: evolution of sex, silencing, and subtelomeres. Mol Biol Evol 22(4): 856-873.

108) Fairhead C, Dujon B (2006) Structure of *Kluyveromyces lactis* subtelomeres: duplication and gene content. FEMS Yeast Res 6: 428–441.

109) Fan M, Chen LC, Ragan MA, Guttell JR, Warner R, et al. (1995) The 5SrRNA and the RNA intergenic spacer of the two varieties of *Cryptococcus neoformans*. J Med Vet Mycol 33: 215–221.

110) Fell JW, Blatt G (1999) Separation of strains of the yeasts *Xantophyllomyces dendrhrous* and *Phaffia rhodomyza* based on rDNa, IGS, and ITS sequence analysis. J Ind Microbiol Biot 1: 677–681.

111) Fell JW, Scorzetti G, Statzell-Tallman A, Boundy-Mills K (2007). Molecular diversity and intragenomic variability in the yeast genus *Xanthophyllomyces*: the origin of *Phaffia rhodozyma*? FEMS Yeast Res 7: 1399–1408.

112) Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783–791.

113) Fenton B, Malloch G, Germa F (1998) A study of variation in rDNA ITS regions shows that two haplotypes coexist within a single aphid genome. Genome 41: 337–345.

114) Fernandez-Espinar MT, Querol A, Ramón D (2000) Molecular characterization of yeasts strains by mitochondrial DNA restriction analysis. In: Spencer JFT, Spencer AL, editors, Methods in Biotechnology, New York: Humana Press Inc. Pp. 329-333.

115) Fitzpatrick DA, Logue ME, Butler G (2008) Evidence of recent interkingdom horizontal gene transfer between bacteria and *Candida parapsilosis.* BMC Evol Biol 8: 181.

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

117) Fogel S, Welch JW (1982) Tandem gene amplification mediates copper resistance in yeast. Proc Natl Acad Sci USA 79: 5342–5346.

118) Forche A, Alby K, Schaefer D, Johnson AD, Berman J, Bennett RJ (2008) The parasexual cycle in *Candida albicans* provides an alternative pathway to meiosis for the formation of recombinant strains. PLoS One 6(5): e110.

119) Forche A, Magee PT, Selmecki A, Berman J, May G (2009) Evolution in *Candida albicans* populations during a single passage through a mouse host. Genetics 182: 799–811.

120) Forche A, Abbey D, Pisithkul T, Weinzierl MA, Ringstrom T (2011) Stress alters rates and types of loss of heterozygosity in *Candida albicans*. mBio. 2: e00129-11.

121) Frank AC, Wolfe KH (2009) Evolutionary capture of viral and plasmid DNA by yeast nuclear chromosomes. Eukaryot Cell 8: 1521–1531.

122) Fraser JA, Heitman J (2003) Fungal mating-type loci. Curr Biol 13: R792-795.

123) Fraser JA, Giles SS, Wenink EC, Geunes-Boyer SG, Wright JR, et al. (2005) Same-sex mating and the origin of the Vancouver Island *Cryptococcus gattii* outbreak. Nature 437: 1360–1364.

124) Galeote V, Bigey F, Beyne E, Novo M, Legras J-L, et al. (2011) Amplification of a *Zygosaccharomyces bailii* DNA segment in wine yeast genomes by extrachromosomal circular DNA formation. PLoS ONE 6(3): e17872. doi:10.1371/journal.pone.0017872

125) Galeote V, Bigey F, Devillers H, Neuvéglise C, Dequina S (2013) Genome sequence of the food spoilage yeast *Zygosaccharomyces bailii* CLIB 213^{T}. Genome Announc 1(4): e00606-13. doi:10.1128/genomeA.00606-13

126) Gamberi T, Magherini F, Borro M, Gentile G, Cavalieri D, Marchi E, Modesti A (2009) Novel insights into phenotype and mitochondrial proteome of yeast mutants lacking proteins Sco1p or Sco2p. Mitochondrion 9: 103:114.

127) Ganley AR, Kobayashi T (2007) Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. Genome Res 17:184–191.

128) Ganley AR, Kobayashi T (2011) Monitoring the rate and dynamics of concerted evolution in the ribosomal DNA repeats of *Saccharomyces cerevisiae* using experimental evolution. Mol Biol Evol 28: 2883–2891.

129) Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, et al. (2000) Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell 11: 4241–4257.

130) Gerstein AC, Chun HJE, Grant A, Otto SP (2006) Genomic convergence toward diploidy in *Saccharomyces cerevisiae*. PLoS Genet 2: 1396–1401.

131) Ginovart M, Prats C, Portell X, Silbert M (2011) Exploring the lag phase and growth initiation of a yeast culture by means of an individual-based model. *Food Microbiol* 28: 810-817.

132) Giudici P, Solieri L, Pulvirenti A, Cassanelli S (2005) Strategies and perspectives for genetic improvement of wine yeasts. Appl Microbiol Biotechnol 66: 607–613.

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

134) Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, et al. (1996) Life with 6000 genes. Science 274: 546–567.

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

136) Gompert Z, Fordyce JA, Forister ML, Shapiro AM, Nice CC (2006) Homoploid hybrid speciation in an extreme habitat. Science 314: 1923–1925.

137) Gonzalez S, Barrio E, Querol A (2008) Molecular characterization of new natural hybrids of *Saccharomyces cerevisiae* and *S. kudriavzevii* in brewing. Appl Environ Microbiol 74: 2314–2320.

138) Gordon JL, Wolfe KH (2008) Recent allopolyploid origin of *Zygosaccharomyces rouxii* strain ATCC 42981. Yeast 25: 449–456.

139) Gordon JL, Armisén D, Proux-Wéra E, Óhéigeartaigh SS, Byrne KP, et al. (2011) Evolutionary erosion of yeast sex chromosomes by mating-type switching accidents. Proc Natl Acad Sci USA. 108: 20024–20029.

140) Gordon DJ, Resio B, Pellman D (2012) Causes and consequences of aneuploidy in cancer. Nat Rev Genet 13: 189-203.

141) Gorman R, Adley CC (2006) Pulsed-field gel electrophoresis as a molecular technique in *Salmonella* epidemiological studies. Methods Biotechnol 21: 81-90.

142) Gostinčar C, Gunde-Cimerman N, Turk M (2012) Genetic resources of extremotolerant fungi: A method for identification of genes conferring stress tolerance. Bioresour Technol 111: 360–367. doi: 10.1016/j.biortech.2012.02.039

143) Gottschalk PG, Dunn JR (2005) The five-parameter logistic: A characterization and comparison with the four-parameter logistic. Anal Biochem 343: 54–65.

144) Goutte C, Johnson AD, (1988) a1 protein alters the DNA binding specificity of alpha 2 repressor. Cell 52: 875–882.

145) Greig D, Leu JY (2009) Natural history of budding yeast. Curr Biol 19: R886–R890.

146) Greig D, Louis EJ, Borts RH, Travisano M (2002) Hybrid speciation in experimental populations of yeast. Science 298: 1773–1775.

147) Gresham, D. Desai MM, Tucker CM, Jenq HT, Pai DA, et al. (2008) The repertoire and dynamics of evolutionary adaptations to controlled nutrientlimited environments in yeast. PLoS Genet 4: e1000303. doi:10.1371/journal.pgen.1000303

148) Groth C, Hansen J, Piškur J (1999) A natural chimeric yeast containing genetic material from three species. Int J Syst Bacteriol 49: 1933–1938.

149) Gu X, Zhang Z, Huang W (2005) Rapid evolution of expression and regulatory divergences after yeast gene duplication. Proc Natl Acad Sci USA 102: 707–712.

150) Haber, JE, Rogers DT, McCusker J (1980) Homothallic conversion of yeast mating type genes occur by intra chromosomal recombination. Cell 22: 277-289.

151) Haber JE (1998) Mating-type gene switching in *Saccharomyces cerevisiae*. Annu Rev Genet 32: 561–599.

152) Haber JE (2006) Transpositions and translocations induced by site-specific double-strand breaks in budding yeast. DNA Repair 5: 998–1009.

153) Haber JE (2012) Mating-type genes and MAT switching in *Saccharomyces cerevisiae*. Genetics 191: 33-64.

154) Hadany L, Beker T (2003) On the evolutionary advantage of fitness-associated recombination. Genetics 165: 2167–2179.

155) Hadany L, Comeron JM (2008) Why are sex and recombination so common? Ann N Y Acad Sci 1133: 26-43.

156) Hafez M, Hausner G (2012) Homing endonucleases: DNA scissors on a mission. Genome 55: 553–569.

157) Hahnenberg K, Jia ZP, Young PC (1996) Functional expression of the *Schizosaccharomyces pombe* Na^+/H^+ gene, *sod2*, in *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA 93: 5031–5036.

158) Hall C, Brachat S, Dietrich FS (2005) Contribution of horizontal gene transfer to the evolution of *Saccharomyces cerevisiae*. Eukaryot Cell 4: 1102–1115.

159) Hall C, Dietrich FS (2007) The reacquisition of biotin prototrophy in *Saccharomyces cerevisiae* involved horizontal gene transfer, gene duplication and gene clustering. Genetics 177: 2293–2307.

160) Hamada T, Noda F, Hayashi K (1984) Structure of cell wall and extracellular mannans from *Saccharomyces rouxii* and their relationship to a high concentration of NaCl in the growth medium. Appl Environ Microbiol 48: 708–712.

161) Harrison E, Muir A, Stratford M, Wheals A (2011) Species specific PCR primers for the rapid identification of yeasts of the genus *Zygosaccharomyces*. FEMS Yeast Res 11: 356–365.

162) Hart B, Jonathan R, David O, McNaughton L (2002) Engineered improvements in DNA-binding function of the MATa1 homeodomain reveal structural changes involved in combinatorial control. J Mol Biol 316: 247-256.

163) Hartwell LH, Dutcher SK, Wood JS, Garvik B (1982) The fidelity of mitotic chromosome reproduction in *S. cerevisiae*. Rec Adv Yeast Mol Biol 1: 28–38.

164) Hassold T, Abruzzo M, Adkins K, Griffin D, Merrill M, et al. (1996) Human aneuploidy: Incidence, origin, and etiology. Environ Mol Mutagen 28: 167–175.

165) Hauck T, Brühlmann F, Schwab W (2003) Formation of 4-Hydroxy-2,5-Dimethyl-3[2H]-Furanone by *Zygosaccharomyces rouxii*: identification of an intermediate. Appl Environ Microbiol 69: 3911-3918.

166) Hawthrone DC (1963) A deletion in yeast and its bearing on the structure of the mating type locus. Genetics 48: 1727-1 729.

167) Hayford AE, Jespersen L (1999) Characterization of *Saccharomyces cerevisiae* strains from spontaneously fermented maize dough by profiles of assimilation, choromosome polymorphism, PCR and *MAL* genotyping. J Appl Microbiol 86: 284–294.

168) He Z, Crist M, Yen H, Duan X, Quiocho FA, et al. (1998) Amino acid residues in both the protein splicing and endonuclease domains of the PI-SceI intein mediate DNA binding. J Biol Chem 273: 4607-4615.

169) Hellborg L, Piškur J (2009) Complex nature of the genome in a wine spoilage yeast, *Dekkera bruxellensis*. Eukaryot Cell 8: 1739–1749.

170) Herskowitz I, Oshima Y. 1981. Control of cell type in Saccharomyces cerevisiae: mating and mating-type interconversion. In: Strathern JN, Jones EW, Broach JR, editors, The Molecular Biology of the Yeast *Saccharomyces*: Life Cycle and Inheritance, New York: Cold Spring Harbor Laboratory Press. Pp. 181–209.

171) Herskowitz I, (1988) Life cycle of the budding yeast *Saccharomyces cerevisiae*. Microbiol Rev 52: 536–553.

172) Herskowitz I, Rine J, Strathern NJ (1992) Mating-type determination and mating-type interconversion in *Saccharomyces cerevisiae*. In: Jones EW, Pringle JR, Broach JR, editors, The Molecular and Cellular Biology of the Yeast *Saccharomyces*, New York: Cold Spring Harbor Lab Press. pp. 583-656.

173) Hicks JB, Strathern NJ, Herskowitz I (1977) The cassette model of mating type interconversion. In: Bukhari A, Shapiro J, Adhya S, editors, DNA Insertion Elements, Plasmids and Episomes, New York: Cold Spring Harbor Lab Press. pp. 457-462.

174) Hicks WM, Kim M, Haber JE (2010) Increased mutagenesis and unique mutation signature associated with mitotic gene conversion. Science 329: 82–85.

175) Hiraoka M, Watanabe K, Umezu K, Maki H (2000) Spontaneous loss of heterozygosity in diploid *Saccharomyces cerevisiae* cells. Genetics 156(4): 1531-1548.

176) Ho CY, Adamson JG, Hodges RS, Smith M (1994) Heterodimerization of the yeast MATa1 and MAT alpha 2 proteins is mediated by two leucine zipper-like coiled-coil motifs. EMBO J. 13: 1403–1413.

177) Ho CY, Smith M, Houston ME, Adamson JG, Hodges RS (2002) A possible mechanism for partitioning between homo- and heterodimerization of the yeast homeodomain proteins MATa1 and MATalpha2. J Pept Res 59: 34-43.

178) Hodgkin J (2005) Karyotype, ploidy, and gene dosage. WormBook 25: 1–9.

179) Hoffman CS, Winston F (1987). A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. Gene 57: 267–272.

180) Hohmann S (2002) Osmotic stress signaling and osmoadaptation in yeasts. Microbiol Mol Biol Rev 66: 300–372.

181) Hou L, Wang M, Wang C, Wang C, Wang H (2013) Analysis of salt-tolerance genes in *Zygosaccharomyces rouxii*. Appl Biochem Biotechnol 170: 1417-1425.

182) Hsueh YP, Fraser JA, Heitman J (2008) Transitions in sexuality: recapitulation of an ancestral tri- and tetrapolar mating system in *Cryptococcus neoformans*. Eukaryot Cell 7: 1847- 1855.

183) Huber T, Faulkner G, Hugenholtz P (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics 20: 2317–2319.

184) Hull CM, Raisner RM, Johnson AD (2000) Evidence for mating of the 'asexual' yeast *Candida albicans* in a mammalian host. Science 14: 307–310.

185) Idnurm A (2011) Sex and speciation: the paradox that non-recombining DNA promotes recombination. Fungal Biol Rev 25: 121–127.

186) Innan H, Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. Nature Rev Genet 11: 97–108.

187) Ito M, Baba T, Mori H, Mori H (2005) Functional analysis of 1440 Escherichia coli genes using the combination of knock-out library and phenotype microarrays. Metab Eng 7(4): 318-327.

188) Iwaki T, Higashida Y, Tsuji H, Tamai Y, Watanabe Y (1998) Characterization of a second gene (*ZSOD22*) of Na^+/H^+ antiporter from salt-tolerant yeast *Zygosaccharomyces rouxii* and functional expression of *ZSOD2* and *ZSOD22* in *Saccharomyces cerevisiae*. Yeast 14: 1167–1174.

189) Iwaki T, Tamai Y, Watanabe Y (1999) Two putative MAP kinase genes, *ZrHOG1* and *ZrHOG2*, cloned from the salt-tolerant yeast *Zygosaccharomyces rouxii* are functionally homologous to the *Saccharomyces cerevisiae HOG1* gene. Microbiology 145: 241–248.

190) Iwaki T, Kurono S, Yokose Y, Kubota K, Tamai Y, et al. (2001) Cloning of glycerol-3-phosphate dehydrogenase genes (*ZrGPD1* and *ZrGPD2*) and glycerol dehydrogenase genes (*ZrGCY1* and *ZrGCY2*) from the salt- tolerant yeast *Zygosaccharomyces rouxii*. Yeast 18: 737-744.

191) Iyer PV, Singhal RS (2008) Production of glutaminase (E.C.3.2.1.5) from *Zygosaccharomyces rouxii*: statistical optimization using response surface methodology. Bioresour Technol 99: 4300–4307.

192) Iyer PV, Singhal RS (2010) Glutaminase production using *Zygosaccharomyces rouxii* NRRL-Y 2547: Effect of aeration, agitation regimes and feeding strategies. Chem Eng Technol 33: 52–62.

193) Jackson AP, Gamble JA, Yeomans T, Moran GP, Saunders D, et al. (2009) Comparative genomics of the fungal pathogens *Candida dubliniensis* and *Candida albicans*. *Genome Res.* 19: 2231–2244.

194) Jacques N, Sacerdot C, Derkaoui M, Dujon B, Ozier-Kalogeropoulos O, et al. (2010) Population polymorphism of nuclear mitochondrial DNA insertion reveals widespread diploidy associated with loss of heterozygosity in *Debaryomyces hansenii*. Eukaryot Cell 9: 449–459.

195) James SA, Collins MD, Roberts IN (1994) Genetic interrelationship among species of the genus *Zygosaccharomyces* as revealed by small subunit rRNA gene sequences. Yeast 10871-881.

196) James SA, Collins MD, Roberts IN (1996) Use of an rRNA internal transcribed spacer region to distinguish phylogenetically closely related species of the genera *Zygosaccharomyces* and *Torulaspora.* Int J Syst Bacteriol 46:189-194.

197) James SA, Stratford M (2003) Spoilage yeasts with emphasis on the genus *Zygosaccharomyces*. In: Boekhout T, Robert V, editors, Yeasts in Food Beneficial and Detrimental Aspects. Hamburg: Behr's Verlag. Pp. 171-187.

198) James SA, Bond CJ, Stratford M, Roberts IN (2005) Molecular evidence of natural hybrids in the genus *Zygosaccharomyces*. FEMS Yeast Res 5: 747–755.

199) James SA, O'Kelly MJ, Carter DM, Davey RP, van Oudenaarden A, et al. (2009) Repetitive sequence variation and dynamics in the ribosomal DNA array of *Saccharomyces cerevisiae* as revealed by whole genome resequencing. Genome Res 19: 626–635.

200) James SA, Stratford M (2011) *Zygosaccharomyces* Barker (1901). In: Kurtzman CP, Fell JW, Boekhout T, editors, The Yeasts, a Taxonomic Study, London: Elsevier. pp. 937–947.

201) Jensen R, Sprague GF, Herskowitz I (1983) Regulation of yeast mating-type interconversion: feedback control of HO gene expression by the mating-type locus. Proc Natl Acad Sci USA 80: 3035–3039

202) Jespersen L, van der Aa Kühle A, Petersen KM (2000) Phenotypic and genetic diversity of *Saccharomyces* contaminants isolated from lager breweries and their phylogenetic relationship with brewing yeasts. Int J Food Microbiol 60: 43–53.

203) Johnson PR, Swanson R, Rakhilina L, Hochstrasser M (1998) Degradation signal masking by heterodimerization of MATalpha2 and MATa1 blocks their mutual destruction by the ubiquitin-proteasome pathway. Cell 94: 217–227.

204) Johnston GC, Singer RA (1978) RNA synthesis and control of strain, ATCC 58716. We found that a protocol cell division in the yeast Saccharomyces cerevisiae. Cell 14: 951–958.

205) Johnston JR (1994) Pulsed field gel electrophoresis. In: Johnston JR, editor, Molecular Genetics of Yeast – A Practical Approach, New York: Oxford University Press. Pp. 83–96.

206) Johnston M, Andrews S, Brinkman R, Cooper J, Ding H, et al. (1994) Complete nucleotide sequence of *Saccharomyces cerevisiae* chromosome VIII. Science 265: 2077–2082.

207) Jin Y, Zhong H, Vershon AK (1999) The yeast a1 and alpha2 homeodomain proteins do not contribute equally to heterodimeric DNA binding. Mol Cell Biol 19: 585-593.

208) Johnston JR (1994) Pulsed field gel electrophoresis. In: Johnson JR, editor, Molecular Genetics of Yeast – A Practical Approach, New York: Oxford University Press. pp. 83–96.

209) Jones T, Federspiel NA, Chibana H, Dungan J, Kalman S, et al. (2004) The diploid genome sequence of *Candida albicans*. P Natl Acad Sci USA 101: 7329–7334.

210) Kageyama K, Senda M, Asano T, Suga H, Ishiguro K (2007). Intra-isolate heterogeneity of the ITS region of rDNA in *Pythium helicoides*. Mycol Res 111: 416–423.

211) Kahm M, Hasenbrink G, Lichtenberg-Frate H, Ludwig J, Kschischo M (2010) grofit: Fitting biological growth curves with R. J Statist Software 33: 1–21. (available at: http://cran.r-project.org/web/packages/grofit/).

212) Kashyap P, Sabu A, Pandey A, Szakacs G, Soccol CR (2002) Extra-cellular L-glutaminase production by *Zygosaccharomyces rouxii* under solid-state fermentation. Process Biochem 38: 307–312.

213) Kataoka S (2005) Functional effects of Japanese style fermented soy sauce (Shoyu) and its components. J Biosci Bioeng 100: 227–234.

214) Kaur R, Ma B, Cormack BP (2007) A family of glycosylphosphatidylinositol-linked aspartyl proteases is required for virulence of *Candida glabrata*. Proc Natl Acad Sci USA 104: 7628–7633.

215) Kavanaugh LA, Fraser JA, Dietrich FS (2006) Recent evolution of the human pathogen *Cryptococcus neoformans* by intervarietal transfer of a 14-gene fragment. Mol Biol Evol **23**: 1879–1890.

216) Ke R, Ingram PJ, Haynes K (2013) An integrative model of ion regulation in yeast. PLoS Comput Biol 9(1): e1002879.

217) Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423: 241–254.

218) Kellis M, Birren BW, Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature 428: 617–624.

219) Kim MD, Han KC, Kang HA, Rhee SK, Seo JH (2003) Coexpression of BiP increased antithrombotic hirudin production in recombinant *Saccharomyces cerevisiae*. J Biotechnol 101: 81–87.

220) Kim SH, Yi SV (2006) Correlated asymmetry of sequence and functional divergence between duplicate proteins of *Saccharomyces cerevisiae.* Mol Biol Evol 23: 1068–1075.

221) Kinclova O, Potier S & Sychrova H (2001) The Zygosaccharomyces rouxii strain CBS732 contains only one copy of the HOG1 and SOD2 genes. J Biotechnol 88: 151–158.

222) Kirchman PA, Botta G (2007) Copper supplementation increases yeast life span under conditions requiring respiratory metabolism. Mech Ageing Dev 128: 187-195.

223) Kiuchi K, Kakezawa M, Ebine H (1980) Mutants of *Saccharomyces rouxii* incapable of oxidizing linoleic acid. Nippon Shokuhin Kogyo Gakkaishi 27: 60–67.

224) Klis FM, Boorsma A, de Groot PWJ (2006) Cell wall construction in *S. cerevisiae*. Yeast 23: 185–202.

225) Knop M (2006) Evolution of the hemiascomycete yeasts: on life styles and the importance of inbreeding. BioEssays 28: 696–708.

226) Kondrashov AS (1994) The asexual ploidy cycle and the origin of sex. Nature 370: 213–216.

227) Kosekia S, Nonaka J (2012) Alternative approach to modeling bacterial lag time, using logistic regression as a function of time, temperature, pH, and sodium chloride concentration. Appl Environ Microbiol 78: 6103-6112.

228) Koszul R, Caburet S, Dujon B, Fischer G (2004) Eukaryotic genome evolution through the spontaneous duplication of large chromosomal segments. EMBO J 23: 234–243.

229) Koszul R, Fischer G (2009) A prominent role for segmental duplications in modeling eukaryotic genomes. C R Biol 332: 254–266.

230) Koutsoumanis K (2008) A study on the variability in the growth limits of individual cells and its effect on the behaviour of microbial populations. Int J Food Microbiol 128: 116–121.

231) Kumaran R, Shi-Yow Y, Jun-Yi L (2013) Characterization of chromosome stability in diploid, polyploid and hybrid yeast cells. PLoS ONE 8(7): e68094.

232) Kurtzman CP (1994) Molecular taxonomic of the yeasts. Yeast 10: 1727-1740.

233) Kurtzman CP, Fell JW (1998) The Yeasts, A Taxonomic Study, 4th edition, Amsterdam: Elsevier Science Publishers.

234) Kurtzman CP, Robnett CJ (1998) Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. Antonie Van Leeuwenhoek 73: 331–371.

235) Kurtzman CP (2003) Phylogenetic circumscription of *Saccharomyces*, *Kluyveromyces* and other members of the Saccharomycetaceae, and the proposal of the new genera *Lachancea*, *Nakaseomyces*, *Naumovia*, *Vanderwaltozyma* and *Zygotorulaspora*. FEMS Yeast Res 4: 233–245.

236) Kurtzman CP, Robnett CJ (2003) Phylogenetic relationships among yeasts of the '*Saccharomyces* complex' determined from multigene sequence analyses. FEMS Yeast Res 3: 417–432.

237) Kurtzman CP, Fell JW, Boekhout T (2011) The Yeasts, A Taxonomic Study, 5th edition, Amsterdam: Elsevier Science Publishers.

238) Kurtzman CP, Fell JW, Boekhout T, Robert V (2011) Methods for the isolation, phenotypic characterization and maintenance of yeasts. In: Kurtzman CP, Fell JW, Boekhout T, editors, The Yeasts, a Taxonomic Study, London: Elsevier. Pp. 87-110.

239) Lacefield S, Magendantz M, Solomon F (2006) Consequences of defective tubulin folding on heterodimer levels, mitosis and spindle morphology in *Saccharomyces cerevisiae*. Genetics 173: 635–646.

240) Lages F, Lucas C (1995) Characterization of a glycerol-H^+ symport in the halotolerant yeast *Pichia sorbitophila*. Yeast 11: 111-119.

241) Lages F, Silva-Graça M, Lucas C (1999) Active glycerol uptake is a mechanism underlying halotolerance in yeasts: a study of 42 species. Microbiology 145: 2577–2586.

242) Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947–2948.

243) Leandro MJ, Sychrová H, Prista C, Loureiro-Dias MC (2011) The osmotolerant fructophilic yeast *Zygosaccharomyces rouxii* employs two plasma-membrane fructose uptake systems belonging to a new family of yeast sugar transporters. Microbiology 157: 601-608.

244) Lee SC, Ni M, Li W, Shertz C, Heitman J (2010) The evolution of sex: a perspective from the fungal kingdom. Microbiol Mol Biol Rev 74: 298–340.

245) Lei X-H, Bochner BR (2013) Using Phenotype MicroArrays to determine culture conditions that induce or repress toxin production by *Clostridium difficile* and other microorganisms. PLoS ONE 8(2): e56545. doi:10.1371/journal.pone.0056545

246) Lenassi M, Zajc J, Gostinčar C, Gorjan A, Gunde-Cimerman N, et al. (2011) Adaptation of the glycerol-3-phosphate dehydrogenase Gpd1 to high salinities in the extremely halotolerant *Hortaea werneckii* and halophilic *Wallemia ichthyophaga*. Fungal Biol 115(10):959-70. doi: 10.1016/j.funbio.2011.04.001

247) Lenassi M, Gostinčar C, Jackman S, Turk M, Sadowski I, et al. (2013) Whole genome duplication and enrichment of metal cation transporters revealed by *de novo* genome sequencing of extremely halotolerant black yeast *Hortaea werneckii*. PLoS ONE 8(8): e71328.

248) Li T, Stark MR, Johnson AD, Wolberger C (1995) Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA. Science 270: 262–269.

249) Li D, Dong Y, Jiang Y, Jiang H, Cai J, et al. (2010) A *de novo* originated gene depresses budding yeast mating pathway and is repressed by the protein encoded by its antisense strand. Cell Res 20: 408–420.

250) Libuda DE, Winston F (2006) Amplification of histone genes by circular chromosome formation in *Saccharomyces cerevisiae.* Nature 443: 1003–1007.

251) Lieckfeldt E, Meyer W, Borner T (1993) Rapid identification and differentiation of yeasts by DNA and PCR fingerprinting. J Basic Microbiol 33: 413–425.

252) Lin X, Hull CM, Heitman J (2005) Sexual reproduction between partners of the same mating type in *Cryptococcus neoformans*. Nature 434: 1017-1021.

253) Lindsley DL, Sandler L, Baker BS, Carpenter AT, Denell RE, et al. (1972) Segmental aneuploidy and the genetic gross structure of the Drosophila genome. Genetics 71: 157–184.

254) Liti G, Barton DBH, Louis EJ (2006) Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. Genetics 174: 839–850.

255) Liu HJ, Li Q, Liu DH, Zhong JJ (2006) Impact of hyperosmotic condition on cell physiology and metabolic flux distribution of *Candida krusei*. Biochem Eng J 28: 92–98.

256) Llanos-Frutos R, Fernandez-Espinar MT, Querol A (2004) Identification of species of the genus *Candida* by analysis of the 5.8S rRNA gene and the two ribosomal internal transcribed spacers. Antonie Van Leeuwenhoek 85(3): 175-85.

257) López-Ribot JL, Kaufman-McAtee R, Kirkpatrick WR, Perea S, Frost-Patterson T (2000) Comparison of DNA-based typing methods to asses genetic diversity and relatedness among *Candida albicans* clinical isolates. Rev Iberoam Micol 17(2): 49-54.

258) López V, Querol A, Ramón D, Fernández-Espinar MT (2001) A simplified procedure to analyse mitochondrial DNA from industrial yeast. Int Food Microbiol 68(1-2): 75-81.

259) Louis VL, Despons L, Friedrich A, Matin T, Durrens P, et al. (2012) *Pichia sorbitophila*, an interspecies yeast hybrid, reveals early steps of genome resolution after polyploidization. G3 (Bethesda) 2: 299–311.

260) Loureiro V, Malfeito-Ferreira M (2003) Spoilage yeasts in the wine industry. Int J Food Microbiol 86: 23-50.

261) Mable BK, Otto SP (2001) Masking and purging mutations following EMS treatment in haploid, diploid and tetraploid yeast (*Saccharomyces cerevisiae*). Genet Res 77: 9–26.

262) Maeda T, Takekawa M, Saito H (1995) Activation of yeast *PBS2* MAPKK by MAPKKKs or by binding of an SH3-containing osmosensor. Science 269, 554–558.

263) Magee BB, Magee PT (2000) Induction of mating in *Candida albicans* by construction of MTLa and MTLα strains. Science 289: 310-313.

264) Magwene PM, Kayıkç Ö, Granek JA, Reininga JM, Scholl Z, et al. (2011) Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA 108: 1987-1992.

265) Mak A, Johnson AD (1993) The carboxy-terminal tail of the homeo domain protein alpha 2 is required for function with a second homeo domain protein. Genes Dev 7: 1862–1870.

266) Mallet J (2007) Hybrid speciation. Nature 446: 279–283.

267) Mallet S, Weiss S, Jacques N, Leh-Louis V, Sacerdot C, et al. (2012) Insights into the life cycle of yeasts from the CTG clade revealed by the analysis of the *Millerozyma* (*Pichia*) *farinosa* species complex. PLoS ONE 7: e35842.

268) Maqueda M, Zamora E, Rodríguez-Cousiño N, Ramírez M (2010) Wine yeast molecular typing using a simplified method for simultaneously extracting mtDNA, nuclear DNA and virus dsRNA. Food Microbiol 27: 205–209.

269) Marcet-Houben, M. & Gabaldón, T. (2010) Acquisition of prokaryotic genes by fungal genomes. Trends Genet 26: 5–8.

270) Marinoni G, Manuel M, Petersen RF, Hvidtfeldt J, Sulo P et al. (1999) Horizontal transfer of genetic material among *Saccharomyces* yeasts. J Bacteriol 181: 6488–6496.

271) Martin N, Ruedi EA, Leduc R, Sun FJ, Caetano-Anolles G (2007) Gene-interleaving patterns of synteny in the *Saccharomyces cerevisiae* genome: are they proof of an ancient genome duplication event? Biol. Direct 2: 23.

272) Martin T, Lu SW, van Tilbeurgh H, Ripoll DR, Dixelius C, et al. (2010) Tracing the origin of the fungal α1 domain places its ancestor in the HMG-box superfamily: implication for fungal mating-type evolution. PLoS ONE 5: e15199.

273) Martorell P, Fernández-Espinar MT, Querol A (2005) Molecular monitoring of spoilage yeasts during the production of candied fruit nougats to determine food contamination sources. International Journal of Food Microbiol 101: 293–302.

274) Martorell P, Stratford M, Steels H, Fernández-Espinar MT,Querol A (2007) Physiological characterization of spoilage strains of *Zygosaccharomyces bailii* and *Zygosaccharomyces rouxii* isolated from high sugar environments. Int J Food Microbiol 114: 234-242.

275) Marques AC, Vinckenbosch N, Brawand D, Kaessmann H (2008) Functional diversification of duplicate genes through subcellular adaptation of encoded proteins. Genome Biol 9: R54.

276) Marques-Bonet T, Girirajan S, Eichler EE (2009) The origins and impact of primate segmental duplications. Trends Genet 25: 443–454.

277) Masneuf I, Hansen J, Groth C, Piškur J, Dubourdieu D (1998) New hybrids between *Saccharomyces* sensu stricto yeast species found among wine and cider production strains. Appl Environ Microbiol 64: 3887–3892.

278) Matheos D, Kingsbury T, Ahsan U, Cunningham K (1997) Tcn1p/Crz1p, a calcineurin-dependent transcription factor that differentially regulates gene expression in *Saccharomyces cerevisiae*. Genes Dev 11: 3445–3458.

279) Matsumoto TK, Ellsmore AJ, Cessna SG, Low PS, Pardo JM, et al. (2002) An osmotically induced cytosolic Ca^{2+} transient activates calcineurin signaling to mediate ion homeostasis and salt tolerance of *Saccharomyces cerevisiae*. J Biol Chem 277: 33075–33080.

280) Matzke MA, Mette MF, Kanno T, Matzke AJ (2003) Does the intrinsic instability of aneuploid genomes have a causal role in cancer? Trends Genet 19: 253–256.

281) McKellar RC, Hawke A (2006) Assessment of distributions for fitting lag times of individual cells in bacterial populations. Int J Food Microbiol 106: 169–175.

282) McLysaght A, Hokamp K, Wolfe KH (2002) Extensive genomic duplication during early chordate evolution. Nat Genet 31: 200–204.

283) Meiron H, Nahon E, Raveh D (1995) Identification of the heterothallic mutation in *HO* endonuclease of *S. cerevisiae* using *HO*/ho chimeric genes. Curr Genet 28: 367-373.

284) Merico A, Sulo P, Piškur J, Compagno C (2007) Fermentative lifestyle in yeasts belonging to the *Saccharomyces* complex. FEBS J 274: 976–989.

285) Meyer W, Mitchell TG, Freedman EZ, Vilgalys R (1993) Hybridization probes for conventional DNA fingerprinting used as single primers in the polymerase chain reaction to distinguish strains of Cryptococcus neoformans. J Clin Microbiol 31: 2274-2280.

286) Meyer W, Mitchell TG (1995) Polymerase chain reaction fingerprinting in fungi using single primers specific to minisatellites and simple repetitive DNA sequences: strain variation in *Cryptococcus neoformans*. Electrophoresis 16: 1648-1656.

287) Morales L, Dujon B (2012) Evolutionary role of interspecies hybridization and genetic exchanges in yeasts. Microbiol Mol Biol Rev 76: 721–739.

288) Mori H (1973) Life cycle in a heterothallic haploid yeast, *Saccharomyces rouxii*. J Ferment Technol 51: 379–392.

289) Mori H, Onishi H (1967) Diploid hybridization in a heterothallic haploid yeast, *Saccharomyces rouxii*. Appl Microbiol 15: 928–934.

290) Mortimer RK, Romano P, Suzzi G, Polsinelli M (1994) Genome renewal: a new phenomenon revealed from a genetic study of 43 strains of *Saccharomyces cerevisiae* derived from natural fermentation of grape musts. Yeast 10: 1543–1552.

291) Mortimer RK, Hawthrone DC (1966) Genetic mapping in *Saccharomyces*. Genetics 53: 165-173.

292) Mortimer RK (2000) Evolution and variation of the yeast (*Saccharomyces*) genome. Genome Res 10: 403–409.

293) Moure CM, Gimble FS, Quiocho FA (2002) Crystal structure of the intein homing endonuclease PI-SceI bound to its recognition sequence. Nat Struct Biol 9: 764-770.

294) Muller H, Thierry A, Coppée JY, Gouyette C, Hennequin C, et al. (2009) Genomic polymorphism in the population of *Candida glabrata*: gene copy-number variation and chromosomal translocations. Fungal Genet Biol 46: 264–276.

295) Musacchio A, Salmon ED (2007) The spindle-assembly checkpoint in space and time. Nat Rev Mol Cell Biol 8: 379–393.

296) Nakao Y, Kanamori T, Itoh T, Kodama Y, Rainieri S, et al. (2009) Genome sequence of the lager brewing yeast, an interspecies hybrid. DNA Res 16: 115–129.

297) Naumova ES, Naumov GI, Masneuf-Pomarède I, Aigle M, Dubourdieu D (2005a) Molecular genetic study of introgression between *Saccharomyces bayanus* and *S. cerevisiae*. Yeast 22: 1099–1115.

298) Naumova S, Sukhotina NN, Naumov GI (2005b) Molecular markers for differentiation between the closely related dairy yeast *Kluyveromces lactis* var. *lactis* and wild *Kluyveromyces lactis* strains from the European "krassilnikovii" population. Microbiology 74: 329–335.

299) Nei M, Rooney AP (2005) Concerted and Birth-and-Death evolution of multigene families. Annu Rev Genet 39: 121–152.

300) Neuvéglise C, Feldman H, Bon E, Gaillardin C, Casarégola S (2002) Genomic evolution of the long terminal repeat retrotransposons in hemiascomycetous yeasts. Genome Res 12: 930–943.

301) Newell EW, Sigal N, Bendall SC, Nolan GP, Davis MM (2012) Cytometry by time-of-flight shows combinatorial cytokine expression and virus-specific cell niches within a continuum of $CD8^+$ T cell phenotypes. Immunity 36: 142–152.

302) Ni M, Feretzaki M, Li W, Floyd-Averette A, Mieczkowski P, et al. (2013) Unisexual and heterosexual meiotic reproduction generate aneuploidy and phenotypic diversity de novo in the yeast *Cryptococcus neoformans*. PLoS Biol 11(9): e1001653

303) Nickoloff JA, Chen EY, Heffron F (1986) A 24-base-pair DNA sequence from the MAT locus stimulates intergenic recombination in yeast. Proc Natl Acad Sci USA 83: 7831–7835.

304) Niwa O, Yanagida M (1985) Triploid meiosis and aneuploidy in *Schizosaccharomyces pombe*: an unstable aneuploid disomic for chromosome III. Curr Genet 9**:** 463–470.

305) Niwa O, Tange Y, Kurabayashi A (2006) Growth arrest and chromosome instability in aneuploid yeast. Yeast 23**:** 937–950.

306) Norbeck J, Blomberg A (2000) The level of cAMP-dependent protein kinase A activity strongly affects osmotolerance and osmo-instigated gene expression changes in *Saccharomyces cerevisiae*. Yeast 16(2): 121-137.

307) Novo M, Bigey F, Beyne E, Galeote V, Gavory F, et al. (2009) Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. Proc Natl Acad Sci USA 106: 16333–16338.

308) Orberá Ratón T (2004) Métodos moleculares de identificación de levaduras de interés biotecnológico. Rev Iberoam Micol 21: 15-19

309) Oda Y, Tonomura K (1995) Electrophoretic karyotyping of the yeast genus *Torulaspora*. Lett Appl Microbiol 21: 190–193.

310) O'Donnell K (1993) Fusarium and its near relatives. In: Reynolds DR, Taylor JW, editors, The Fungal Holomorph: Mitotic, Meiotic and Pleomorphic Speciation in Fungal Systematics, Wallingford: CAB International. pp. 225–233.

311) Ogawa Y, Tatsumi H, Murakami S, Ishida Y, Murakami K, et al. (1990) Secretion of *Aspergillus oryzae* alkaline protease in an osmophilic yeast, *Zygosaccharomyces rouxii*. Agric Biol Chem 54: 2521–2529.

312) Ohno S (1970) Evolution by Gene Duplication. New York: Springer Verlag.

313) Oliver SG, van dar Aart QJM, Agostoni-Carbone ML, Aigle M, Alberghina L, et al. (1992) The complete DNA sequence of yeast chromosome III. Nature: 357:38-46.

314) Oner MD, Erickson LE, Yang SS (1986) Utilization of spline functions for smoothing fermentation data and for estimation of specific rates. Biotechnol Bioeng 28: 902–918.

315) Onishi H (1963) Osmophilic yeasts. Adv Food Res 12: 53-94.

316) Onishi H, Suzuki T (1968) Production of D-mannitol and glycerol by yeasts. Appl Microbiol 16: 1847-1852.

317) Otto SP, Whitton J (2000) Polyploidy incidence and evolution. Annu Rev Genet 34: 401–437.

318) Papp B, Pal C, Hurst LD (2003) Evolution of *cis* regulatory elements in duplicated genes of yeast. Trends Genet 19: 417–422.

319) Parry EM, Cox BS (1970) The tolerance of aneuploidy in yeast. Genet Res 16: 333-340.

320) Paun O, Fay MF, Soltis DE, Chase MW (2007) Genetic and epigenetic alterations after hybridization and genome doubling. Taxon 56: 649–656.

321) Pavelka N, Rancati G, Zhu J, Bradford WD, Saraf A, et al. (2010) Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. Nature 468: 321-325.

322) Payen C, Koszul R, Dujon B, Fischer G (2008) Segmental duplications arise from pol32-dependent repair of broken forks through two alternative replication-based mechanisms. PLoS Genet. 4: e1000175.

323) Pena-Castillo L, Hughes TR (2007) Why are there still over 1000 uncharacterized yeast genes? Genetics 176: 7–14.

324) Pfau J, Amon A (2012) Chromosomal instability and aneuploidy in cancer: from yeast to man. EMBO Rep 13: 515-527.

325) Philley ML, Staben C (1994) Functional analyses of the *Neurospora crassa* MT a-1 mating type polypeptide. Genetics 137: 715-722.

326) Pinto PM, Resende MA, Koga-Ito CY, Ferreira JA, Tendler M (2004) rDNA-RFLP identification of *Candida* species in immunocompromised and seriously diseased patients. Can J Microbiol 50(7): 514-20.

327) Piškur J (2001) Origin of the duplicated regions in the yeast genomes. Trends Genet 17(6): 302-303.

328) Pitt JI, Hocking AD (2009) Fungi and Food Spoilage, 3rd edition, New York: Springer.

329) Plemenitas A, Vaupotic T, Lenassi M, Kogej T, Gunde-Cimerman N (2008) Adaptation of extremely halotolerant black yeast Hortaea werneckii to increased osmolarity: a molecular perspective at a glance. Stud Mycol 61: 67-75. doi: 10.3114/sim.2008.61.06

330) Poláková S, Blume C, Zárate J, Mentel M, Jorck-Ramberg D, et al. (2009) Formation of new chromosomes as a virulence mechanism in yeast *Candida glabrata*. Proc Natl Acad Sci USA 106: 2688–2693.

331) Porro D, Branduardi P (2009) Yeast cell factory: fishing for the best one or engineering it? Microb Cell Fact 8: 51.

332) Posas F, Wurgler-Murphy SM, Maed T, Witten EA, Thai TC, et al. (1996) Yeast Hog1 MAP kinase cascade is regulated by a multistep phosphorelay mechanism in the Sln1–Ypd1–Ssk1 'two component' osmosensor. Cell 86: 865–875.

333) Posas F, Saito H (1997) Osmotic activation of the HOG MAPK pathway via Ste11p MAPKKK: scaffold role of Pbs2p MAPKK. Science 276: 1702-1705.

334) Posas F, Chambers JR, Heyman JA, Hoeffler JP, de Nadal E, et al. (2000) The trancriptional response of yeast to saline stress. J Biol Chem 275: 17249–17255.

335) Pribylova L, de Montigny J, Sychrová H (2007a) Osmoresistant yeast *Zygosaccharomyces rouxii*: the two most studied wildtype strains (ATCC 2623 and ATCC 42981) differ in osmotolerance and glycerol metabolism. Yeast 24: 171–180.

336) Pribylova L, Farkaš V, Slaninová I, de Montigny J, Sychrová H (2007b) Differences in osmotolerant and cell-wall properties of two *Zygosaccharomyces rouxii* strains. Folia Microbiol 52: 241–245.

337) Pribylova L, Papouskova K, Sychrová H (2008) The salt tolerant yeast *Zygosaccharomyces rouxii* possesses two plasma-membrane Na^+/H^+-antiporters (ZrNha1p and ZrSod2-22p) playing different roles in cation homeostasis and cell physiology. Fungal Genet Biol 45: 1439-1447.

338) Prior C, Potier S, Souciet JL, Sychrová H (1996) Characterization of the *NHA1* gene encoding a Na^+/H^+ antiporter of the yeast *Saccharomyces cerevisiae*. FEBS Lett 387: 89-93.

339) Proft M, Struhl K (2002) Hog1 kinase converts the Sko1-Cyc8-Tup1 repressor complex into an activator that recruits SAGA and SWI/SNF in response to osmotic stress. Mol Cell 9: 1307–1317.

340) Proux-Wéra E, Byrne KP, Wolfe KH (2011) Evolutionary mobility of the ribosomal DNA array in yeasts. Genome Biol Evol 5(3): 525–531.

341) Pujol C, Daniels KJ, Lockhart SR, Srikantha T, Radke JB, et al. (2004) The closely related species *Candida albicans* and *Candida dubliniensis* can mate. Eukaryot Cell 3: 1015–1027.

342) Pulvirenti A, Caggia C, Restuccia C, Giudici P, Zambonelli C (2000) Inheritance of mitochondrial DNA in interspecific *Saccharomyces* hybrids. Ann Microbiol 50: 61-64.

343) Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, et al. (2012) The Pfam protein families database. Nucleic Acids Res 40: D290-D301.

344) Querol A, E Barrio (1990) A rapid and simple method for preparatuion of yeast mitochondrial DNA. Nucl Acids Res 18: 1657.

345) Querol A, Barrio E, Huerta T, Ramón D (1992) Molecular monitoring of wine fermentations conducted by active dry yeast strains. Appl Environ Microbiol 58: 2948–2953.

346) Quirós M, Martorell P, Valderrama MJ, Querol A, Peinado JM, et al. (2006) PCR-RFLP of the IGS region of rDNA, a useful tool for the practical discrimination between species of the genus *Debaryomyces*. Antonie van Leeuwenhoek 90: 211–219.

347) Ramos J, Ariño J, Sychrová H (2011) Alkali-metal-cation influx and efflux systems in nonconventional yeast species. FEMS Microbiol Lett 317: 1-8.

348) Rep M, Krantz M, Thevelein JM, Hohmann S (2000) The transcriptional response of *saccahromyces cerevisiae* to osmotic shock. Hot1p and Msn2p/Msn4p are required for the induction of subsets of high osmolarity glycerol pathway-dependent genes. J Biol Chem 275: 8290-8300.

349) Replansky T, Koufopanou V, Greig D, Bell G (2008) *Saccharomyces sensu stricto* as a model system for evolution and ecology. Trends Ecol Evol 23: 494–501.

350) Ricchetti M, Fairhead C, Dujon B (1999) Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. Nature 402: 96–100.

351) Richards TA, Leonard G, Soanes DM, Talbot NJ (2011) Gene transfer into the fungi. Fungal Biol Rev 25: 98-110.

352) Rodicio R, Heinisch JJ (2010) Together we are strong: cell wall integrity sensors in yeast. Yeast 27: 531–540.

353) Rodrigues F, Ludovico P, Sousa MJ, Steensma HY, Corte-Real M, et al. (2003) The spoilage yeast *Zygosaccharomyces bailii* forms mitotic spores: a screening method for haploidization. Appl Environ Microbiol 69: 649–653.

354) Rolland T, Neuveglise C, Sacerdot C, Dujon B (2009) Insertion of horizontally transferred genes within conserved syntenic regions of yeast genomes. PLoS One 4(8): e6515.

355) Romero P, Patiño B, Quirós M, González-Jaén MT, Valderrama MJ, et al. (2005) Differential detection of Debaryomyces hansenii isolated from intermediate-moisture foods by PCR-RFLP of the IGS region of rDNA. FEMS Yeast Res 5: 455–461.

356) Rooney AP, Ward TJ (2005) Evolution of a large ribosomal RNA multigene family in filamentous fungi: birth and death of a concerted evolution paradigm. Proc Natl Acad Sci USA 102: 5084–5089.

357) Rosa CA, Lachance MA (2005). *Zygosaccharomyces machadoi* sp. nov., a yeast species isolated from a nest of the stingless bee *Tetragonisca angustula*. Lundiana 6: 27–29.

358) Rosenstraus MJ, Chasin LA (1978) Separation of linked markers in Chinese hamster cell hybrids: mitotic recombination is not involved. Genetics 90: 735–760.

359) Ruiz A, Ariño J (2007) Function and regulation of the *Saccharomyces cerevisiae ENA* sodium ATPase system. Eukaryot Cell 6: 2175–2183.

360) Rusnak F, Mertz P (2000) Calcineurin: form and function. Physiol Rev 4: 1483-1521.

361) Rustchenko E (2007) Chromosome instability in *Candida albicans*. FEMS Yeast Res 7: 2–11.

362) Ryall B, Eydallin G, Ferenci T (2012) Culture history and population heterogeneity as determinants of bacterial adaptation: the adaptomics of a single environmental transition. Microbiol Mol Biol Rev 76: 597-625

363) Rycoyska A, Valach M, Tomaska l, Bolotin-Fukuhara M, Nosek J (2004) Linear versus circular mitochondrial genomes: intraspecies veriability of micochondrial genome architecture in *Candida parasilopsis*. Microbiology 150: 1571-80.

364) Sacchetti M (1932) Ricerche sulla fermentazione di un mosto d'uva concentrato. Arch Microbiol 3: 473–491.

365) Sabate J, Cano J, Querol A, Guillamón JM (1998) Diversity of *Saccharomyces* strains in wine fermentations: analysis for two consecutive years. Lett Appl Microbiol 26: 452-455.

366) Sabate J, Cano J, Esteve-Zarzoso B, Guillamón JM (2002) Isolation and identification of yeasts associated with vineyard and winery by RFLP analysis of ribosomal genes and mitochondrial DNA. Microbiol Res 157: 1-8.

367) Sacerdot C, Casaregola S, Lafontaine I, Tekaia F, Dujon B, et al. (2008) Promiscuous DNA in the nuclear genomes of hemiascomycetous yeasts. FEMS Yeast Res 8: 846–857.

368) Saha BC, Sakakibara Y, Cotta MA (2007) Production of D-arabitol by a newly isolated *Zygosaccharomyces rouxii*. J Ind Microbiol Biotechnol 34: 519–523.

369) Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–425.

370) Saksinchai S, Suzuki M, Chantawannakul P, Ohkuma M, Lumyong S (2012). A novel ascosporogenous yeast species, *Zygosaccharomyces siamensis*, and the sugar tolerant yeasts associated with raw honey collected in Thailand. Fungal Divers 52: 123–139.

371) Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning: A Laboratory Manual, 2nd edition, New York: Cold Spring Harbor Laboratory Press.

372) Sancho T, Giménez-Jurado G, Malfeito-Ferrera M, Loureiro V (2000) Zymological indicators, a new concept applied to the detection of potential spoilage yeast species associated with fruit pulps and concentrates. Food Microbiol 17: 613–624.

373) Scannell DR, Frank AC, Conant GC, Byrne KP, Woolfit M, et al. (2007) Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. Proc Natl Acad Sci USA 104: 8397–8402.

374) Schacherer J, Tourrette Y, Potier S, Souciet JL, de Montigny J (2007) Spontaneous duplications in diploid *Saccharomyces cerevisiae* cells. DNA Repair 6: 1441–1452.

375) Schüller C, Brewster JL, Alexander MR, Gustin MC, Ruis H (1994) The HOG pathway controls osmotic regulation of transcription via the stress response element (STRE) of the *Saccharomyces cerevisiae CTT1* gene. EMBO J 13: 4382–4389.

376) Sekse C, Bohlin J, Skjerve E, Vegarud GE (2012) Growth comparison of several *Escherichia coli* strains exposed to various concentrations of lactoferrin using linear spline regression. Microb Inform Exp 2:5.

377) Sheltzer JM, Blank HM, Pfau SJ, Tange Y, George BM, et al. (2011) Aneuploidy drives genomic instability in yeast. Science 333: 1026–1030.

378) Shi, X., Karkut, T., Chamankhah, M., Alting-Mees, M., Hemmingsen, S.M., Hegedus, D., 2003. Optimal conditions for the expression of a single-chain antibody (scFv) gene in *Pichia pastoris*. Protein Expr Purif 28: 321–330.

379) Shin JH, MR Park, JW Song, Shin D, Jung SI, et al. (2004) Microevolution of *Candia albicans* strain during catheter-related candidemia. J Clin Microbiol 42(9): 4025-4031.

380) Silva-Graça M, Lucas C (2003) Physiological studies on long-term adaptation to salt stress in the extremely halotolerant yeast *Candida versatilis* CBS 4019 (syn. *C. halophila*). FEMS Yeast Res 3: 247–260.

381) Sipiczki M (2008) Interspecies hybridization and recombination in *Saccharomyces* wine yeasts. FEMS Yeast Res 8: 996–1007.

382) Sipiczki M, Pfliegler MP, Holb IJ (2013) *Metschnikowia* species share a pool of diverse rRNA genes differing in regions that determine hairpin-loop structures and evolve by reticulation. PLoS ONE 8(6): e67384. doi:10.1371/journal.pone.0067384

383) Simon UK, Weiß M (2008) Intragenomic variation of fungal ribosomal genes is higher than previously thought. Mol Biol Evol 25: 2251-2254.

384) Skelly DA, Merrihew GE, Riffle M, Connelly CF, Kerr EO (2013) Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. Genome Res 23. doi:10.1101/gr.155762.113

385) Sluis CVD, Tramper J, Wijffles RH (2001) Enhancing and accelerating flavor formation by salt tolerant yeasts in Japanese soy sauce processes. Trends Food Sci Technol 12: 322-327.

386) Smith A, Ward MP, Garrett S (1998) Yeast PKA represses Msn2p/Msn4p-dependent gene expression to regulate growth, stress response and glycogen accumulation. EMBO J: 17, 3556-3564.

387) Snowdon JA, Cliver DO (1996) Microorganisms in honey. Int J Food Microbiol 31: 1-26.

388) Solieri L, Landi S, de Vero L, Giudici P (2006) Molecular assessment of indigenous yeast population from traditional balsamic vinegar. J Appl Microbiol 101: 63–71.

389) Solieri L, Cassanelli S, Giudici P (2007) A new putative *Zygosaccharomyces* yeast species isolated from Traditional Balsamic Vinegar. Yeast 24: 403–417.

390) Solieri L, Cassanelli S, Croce MA, Giudici P (2008a) Genome size and ploidy level: new insights for elucidating relationships in *Zygosaccharomyces* species. Fungal Genet Biol 45: 1582–1590.

391) Solieri L, Antunez O, Perez-Ortin JE, Barrio E, Giudici P (2008b) Mitochondrial inheritance and fermentative : oxidative balance in hybrids between Saccharomyces cerevisiae and Saccharomyces uvarum. Yeast 25: 485–500.

392) Solieri L, Giudici P (2008) Yeasts associated to Traditional Balsamic Vinegar: ecological and technological features. Int J Food Microbiol 125(1): 36-45.

393) Solieri L, Dakal TC, Giudici P (2013a) *Zygosaccharomyces sapae* sp. nov., a novel yeast species isolated from Italian traditional balsamic vinegar. Int J Syst Evol Microbiol 63: 364–371.

394) Solieri L, Dakal TC, Croce MA, Giudici P (2013b) Unraveling genomic diversity of the *Zygosaccharomyces rouxii* complex with a link to its life cycle. FEMS Yeast Res 13: 245-258.

395) Solieri L, Dakal TC, Bicciato S (2014a) Quantitative phenotypic analysis of multi-stress response in *Zygosaccharomyces rouxii* complex. FEMS Yeast Res (manuscript accepted)

396) Solieri L, Dakal TC, Giudici P, Cassanelli S (2014b) Sex-determination system in the diploid yeast *Zygosaccharomyces sapae*. G3 (Bethesda) (manuscript accepted)

397) Souciet JL, Dujon B, Gaillardin C, Johnston M, Baret PV, et al. (2009) Comparative genomics of protoploid Saccharomycetaceae. Genome Res 19: 1696-709.

398) Srikantha TS, Lachke A, Soll DR (2003) Three mating type-like loci in *Candida glabrata*. Eukaryot Cell 2: 328-340.

399) Stajich JE, Dietrich FS, Roy SW (2007) Comparative genomic analysis of fungal genomes reveals intron-rich ancestors. Genome Biol 8: R223.

400) Stark MR, Escher D, Johnson AD (1999) A trans-acting peptide activates the yeast a1 repressor by raising its DNA-binding affinity. EMBO J 18: 621-1629.

401) Stathopoulos AM, Cyert MS (1997) Calcineurin acts through the *CRZ1*/*TCN1*-encoded transcription factor to regulate gene expression in yeast. Genes Dev 11: 3432–3444.

402) Stathopoulos-Gerontides A, Guo J, Cyert MS (1999) Yeast calcineurin regulates nuclear localization of the Crz1p transcription factor through dephosphorylation. Genes Dev 13: 798–803.

403) Steels S, James SA, Bond CJ, Roberts IN, Stratford M (2002) *Zygosaccharomyces kombuchaensis*: the physiology of a new species related to the spoilage yeasts *Zygosaccharomyces lentus* and *Zygosaccharomyces bailii*. FEMS Yeast Research 2: 113-121.

404) Stelling-Dekker NM (1931) Die Hefesammlung des central-bureau voor Schimmelcultures, II Teil 1. Die Sporogenen Hefen, Noord Hollandsche Uitgevers Maatschappij, Amsterdam.

405) Stephan F, Bah MS, Desterke D, Rezaiguia-Delclaux S, Foulet F, et al. (2002) Molecular diversity and routes of colonization of *Candida albicans* in a surgical intensive care unit, as studied using microsatellite markers. Clin Infect Dis 35(12): 1477-1483.

406) Stoddard BL (2005) Homing endonuclease structure and function. Q Rev Biophys 38: 49-95.

407) Stöver AG, Witek-Janusek L, Mathews HL (1998) A method for flow cytometric analysis of *Candida albicans* DNA. J Microbiol Methods 33: 191–196.

408) Stratford M (2006) Food and beverage spoilage yeast. In: Querol A, Fleet G, editors, The Yeast Handbook: Yeast in Food and Beverages, Germany: Springer-Verlag. pp. 335–379.

409) Strathern JN, Newlon CS, Herskowitz I, Hicks JB (1979) Isolation of a circular derivative of yeast chromosome III: implications for the mechanism of mating type interconversion. Cell 18: 309-319.

410) Strathern JN, Klar AJ, Hicks JB, Abraham JA, Ivy JM, et al. (1982) Homothallic switching of yeast mating type cassettes is initiated by a double-stranded cut in the *MAT* locus. Cell 31: 183–192.

411) Strathern JN (1988) Control and execution of mating type switching in *Saccharomyces cerevisiae*. In: Kucherlapati R., Smith G. R., editors, Genetic Recombination, Washington DC: ASM Press. Pp. 445–464.

412) Stříbný J, Kinclová-Zimmermannová O, Sychrová H (2012) Potassium supply and homeostasis in the osmotolerant non-conventional yeasts *Zygosaccharomyces rouxii* differ from *Saccharomyces cerevisiae*. Curr Genet 58: 255-264.

413) Suezawa Y, Suzuki M, Mori H (2008) Genotyping of a miso and soy sauce fermentation yeast, *Zygosaccharomyces rouxii*, based on sequence analysis of the partial 26S ribosomal RNA gene and two internal transcribed spacers. Biosci Biotechnol Biochem 72: 2452–2455.

414) Sugita T, Ikeda R, Shinoda T (2001) Diversity among strains of *Cryptococcus neoformans* isolates var. *gatti* as revealed by a sequence analysis of capsular polisaccaharide. Microbiol Immunol 69: 2080–2086.

415) Suh SO, Gujjari P, Beres C, Beck B, Zhou J (2013) Proposal of *Zygosaccharomyces parabailii* sp. nov. and *Zygosaccharomyces pseudobailii* sp. nov., two new species closely related to *Zygosaccharomyces bailii*. Int J Syst Evol Microbiol. doi:10.1099/ijs.0.048058-0

416) Swinnen IAM, Bernaerts K, Dens EJJ, Geeraerd AH, Van Impe JF (2004) Predictive modelling of the microbial lag phase: a review. Int J Food Microbiol 94: 137–159.

417) Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10(3): 512-526.

418) Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596-1599.

419) Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731-2739.

420) Tan S, Richmond TJ (1998) Crystal structure of the yeast MATalpha2/MCM1/DNA ternary complex. Nature 391: 660–666.

421) Tang XM, Kayingo G, Prior BA (2005) Functional analysis of the *Zygosaccharomyces rouxii* Fps1p homologue. Yeast 22: 571–581.

422) Tanner SD, Bandura DR, Ornatsky O, Baranov VI, Nitz M, et al. (2008) Flow cytometer with mass spectrometer detection for massively multiplexed single-cell biomarker assay. Pure Appl Chem 80: 2627–2641.

423) Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, et al. (2000). Phylogenetic species recognition and species concepts in fungi. Fungal Genet Biol 31: 21–32.

424) Taylor DJ, Bruenn J (2009) The evolution of novel fungal genes from non-retroviral RNA viruses. BMC Biol. **7**: 88.

425) Thevelein JM, de Winde JH (1999) Novel sensing mechanisms and targets for the cAMP protein kinase A pathway in the yeast *Saccharomyces cerevisiae*. Mol Microbiol 33: 904-908.

426) Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25: 4876–4882.

427) Thompson SL, Compton DA (2010) Proliferation of aneuploid human cells is limited by a p53-dependent mechanism. J Cell Biol 188: 369–381.

428) Toh A, Tada S, Oshima Y (1982) 2-p.m DNA-like plasmids in the osmophilic haploid yeast *Saccharomyces rouxii*. J Bacteriol 151: 1380-1390.

429) Toh A, Utatsu I (1985) Physical and functional structure of a yeast plasmid, pSB3, isolated from *Zygosaccharomyces bisporus*. Nucleic Acids Res 13: 4267-4283.

430) Tomaszewska L, Rywińska A, Gładkowski W (2012) Production of erythritol and mannitol by *Yarrowia lipolytica* yeast in media containing glycerol. J Ind Microbiol Biot 39: 1333-1343.

431) Tornai-Lehoczki J, Dlauchy D (2000) Delimination of brewing yeast strains using different molecular techniques. Int J Food Microbiol 62: 37–45.

432) Török T, Rockhold D, King AD (1993) Use of electrophoretic karyotyping and DNA–DNA hybridisation in yeast identification. Int J Food Microbiol 19: 63–80.

433) Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, et al. (2007) Effects of aneuploidy on cellular physiology and cell division in haploid yeast. Science 317**:** 916–924.

434) Torres EM, Williams BR, Amon A (2008) Aneuploidy: cells losing their balance. Genetics 179: 2737-746.

435) Torres EM, Dephoure N, Panneerselvam A, Tucker CM, Whittaker CA, et al. (2010) Identification of aneuploidy-tolerating mutations. Cell 143: 71–83.

436) Torriani S, Lorenzini M, Salvetti E, Felis GE (2011). *Zygosaccharomyces gambellarensis* sp. nov., an ascosporogenous yeast isolated from an Italian 'passito' style wine. Int J Syst Evol Microbiol 61: 3084–3088.

437) Tsong AE, Miller MG, Raisner RM, Johnson AD (2003) Evolution of a combinatorial transcriptional circuit: a case study in yeasts. Cell 115: 389–399.

438) Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, et. al. (2012) Primer3 – new capabilities and interfaces. Nucleic Acids Res 40: e115. doi: 10.1093/nar/gks596.

439) Usher J, Bond U (2009) Recombination between homoeologous chromosomes of lager yeasts leads to loss of function of the hybrid *GPH1* gene. Appl Environ Microbiol 75: 4573–4579.

440) Vaas LAI, Sikorski J, Michael V, Göker M, Klenk HP (2012) Visualization and curve-parameter estimation strategies for efficient exploration of phenotype microarray kinetics PLoS ONE 7(4): e34846. doi:10.1371/journal.pone.0034846

441) van Eck JH, Prior BA, Brandt EV (1993) The water relations of growth and polyhydroxy alcohol production by ascomycetous yeasts. J Gen Microbiol 139: 1047–1054.

442) van Zyl PJ, Prior BA (1990) Adaptation of *Zygosaccharomyces rouxii* to changes in water activity in transient continuous culture. Biotechnol Lett 12: 361-366.

443) Varani G, McClain WH (2000) The GU wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. EMBO Rep 1: 18–23.

444) Vershon AK, Johnson AD (1993) A short, disordered protein region mediates interactions between the homeodomain of the yeast alpha 2 protein and the MCM1 protein. Cell 72: 105–112.

445) Vicenzi JT, Zmijewski MJ, Reinhard MR, Landen BE, Muth WL, et al. (1997) Large-scale stereoselective enzymatic ketone reduction with in situ product removal via polymeric adsorbent resins. Enzyme Microb Technol 20: 494–499.

446) Vigentini I, Brambilla L, Branduardi P, Merico A, Porro D, et al. (2005) Heterologous protein production in *Zygosaccharomyces bailii*: physiological effects and fermentative strategies. FEMS Yeast Res 5: 647-652.

447) Vision TJ, Brown DG, Tanksley SD (2000) The origins of genomic duplications in Arabidopsis. Science 290: 2114–2117.

448) Vital-Lopez FG, Wallqvist A, Reifman J (2013) Bridging the gap between gene expression and metabolic phenotype via kinetic models. BMC Syst Biol 7: 63 doi:10.1186/1752-0509-7-63

449) Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, et al. (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23: 4407–4414.

450) Wang ZX, Kayingo G, Blomberg A, Prior BA (2002) Cloning, sequencing and characterization of a gene encoding dihydroxyacetone kinase from *Zygosaccharomyces rouxii* NRRL2547. Yeast 19: 1447–1458.

451) Wang DM, Yao YJ (2005). Intrastrain internal transcribed spacer heterogeneity in *Ganoderma* species. Can J Microbiol 51: 113–121.

452) Warringer J, Blomberg A (2003) Automated screening in environmental arrays allows analysis of quantitative phenotypic profiles in *Saccharomyces cerevisiae*. Yeast 20: 53–67.

453) Warringer J, Ericson E, Fernandez L, Nerman O, Blomberg A (2003) High-resolution yeast phenomics resolves different physiological features in the saline response. Proc Natl Acad Sci USA 20: 15724–15729.

454) Watanabe Y, Miwa S, Tamai Y (1995) Characterization of Na^+/H^+-antiporter gene closely related to the salt-tolerance of yeast *Zygosaccharomyces rouxii*. Yeast 11: 829–838.

455) Watanabe Y, Iwaki T, Shimono Y, Ichimiya A, Nagaoka Y, Tamai, Y (1999) Characterization of the Na1-ATPase gene (*ZENA1*) from the salt-tolerant yeast *Zygosaccharomyces rouxii*. J Biosci Bioeng 88: 136–142.

456) Watanabe Y, Hirasaki M, Tohnai N, Yagi K, Abe S, et al. (2003) Salt shock enhances the expression of *ZrATP2*, the gene for the mitochondrial ATPase β subunit of *Zygosaccharomyces rouxii*. J Biosci Bioeng 96: 193-195.

457) Watanabe Y, Tsuchimoto S, Tamai Y (2004) Heterologous expression of *Zygosaccharomyces rouxii* glycerol 3-phosphate dehydrogenase gene (*ZrGPD1*) and glycerol dehydrogenase gene (*ZrGCY1*) in *Saccharomyces cerevisiae*. FEMS Yeast Res 4: 505–510.

458) Watanabe J, Uehara K, Mogi Y (2013) Adaptation of the osmotolerant yeast *Zygosaccharomyces rouxii* to an osmotic environment through copy number amplification of *FLO11D*. Genetics 195: 393-405.

459) Watanabe J, Uehara K, Mogi Y (2013) Diversity of mating-type chromosome structures in the yeast *Zygosaccharomyces rouxii* caused by ectopic exchanges between MAT-Like Loci. PLoS ONE 8: e62121.

460) Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics 25: 1189-1191.

461) Weaver BA, Cleveland DW (2006) Does aneuploidy cause cancer? Curr Opin Cell Biol 18: 658–667.

462) Wei W, McCusker JH, Hyman RW, Jones T, Ning Y, et al. (2007) Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. Proc Natl Acad Sci USA 104: 12825–12830.

463) Wei Y, Wang C, Wang M, Cao X, Hou L (2013) Comparative analysis of salt-tolerant gene *HOG1* in a *Zygosaccharomyces rouxii* mutant strain and its parent strain. J Sci Food Agric 93: 2765–2770.

464) Weiffenbach B, Rogers DT, Haber JE, Zoller M, Russell DW, et al. (1983) Deletions and single base pair changes in the yeast mating type locus that prevent homothallic mating type conversions. Proc Natl Acad Sci USA 80: 3401–3405.

465) Weising K, Atkinson RG, Gardner RC (1995) Genomic fingerprinting by microsatellite-primed PCR: a critical evaluation. PCR Meth Applications 4: 249-255.

466) White TJ, Bruns T, Lee S, Taylor J (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, editors, PCR Protocols: a Guide to Methods and Applications, San Diego: Academic Press. Pp. 315–322.

467) Wickerham LJ, Burton KA (1960) Heterothallism in *Saccharomyces rouxii*. J Bacteriol 80: 492–495.

468) Williams BR, Prabhu VR, Hunter KE, Glazier CM, Whittaker CA, et al. (2008) Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells. Science 322: 703-709.

469) Wolberger C, Vershon AK, Liu B, Jhonson AD, Pabo CO (1991) Crystal structure of a MATa homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. Cell 67: 517-528.

470) Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387: 708–713.

471) Wong S, Butler G, Wolfe KH (2002) Gene order evolution and paleopolyploidy in hemiascomycete yeasts. Proc Natl Acad Sci USA 99: 9272–9277.

472) Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, et al. (2002) The genome sequence of *Schizosaccharomyces pombe*. Nature 415: 871–880.

473) Woolfit M, Rozpedowska E, Piskur J, Wolfe KH (2007) Genome survey sequencing of the wine spoilage yeast *Dekkera (Brettanomyces) bruxellensis*. Eukaryot Cell 6: 721–733.

474) Wrent P, Rivas E-M, Peinado JM, de Silóniz MI (2010) Strain typing of *Zygosaccharomyces* yeast species using a single molecular method based on polymorphism of the intergenic spacer region (IGS). Int J Food Microbiol 142: 89–96.

475) Wu W, Pujol C, Lockhart SR, Soll DR (2005) Chromosome loss followed by duplication is the major mechanism of spontaneous mating-type locus homozygosis in *Candida albicans*. Genetics 169: 1311–1327.

476) Wu Q, James SA, Roberts IN, Moulton V, Huber KT (2008) Exploring contradictory phylogenetic relationships in yeasts. FEMS Yeast Res 8: 641–650.

477) Yarrow D (1984). *Zygosaccharomyces* Barker. In Kreger-van Rij NJW, editor, The Yeasts - a Taxonomic Study, Amsterdam: Elsevier. Pp. 449-465.

478) Yarrow D (1998) Methods for the isolation and identification of yeasts. In Kurtzman CP, Fell JW, editors, The Yeasts, a Taxonomic Study, Amsterdam: Elsevier. Pp. 77-100.

479) Yona AH, Manor YS, Herbst RH, Romano GH, Mitchell A, et al. (2012) Chromosomal duplication is a transient evolutionary solution to stress. Proc Natl Acad Sci USA 109: 21010–21015.

480) Yu X, Gabriel A (2004) Reciprocal translocations in *Saccharomyces cerevisiae* formed by nonhomologous end joining. Genetics 166: 741–751.

481) Zeyl C, Vanderford T, Carter M (2003) An evolutionary advantage of haploidy in large yeast populations. Science 299: 555–557.

482) Zhang J, Biswas I (2009) A phenotypic microarray analysis of a *Streptococcus mutans liaS* mutant. Microbiology 155: 61–68

483) Zhao XQ, Bai FW (2009) Mechanisms of yeast stress tolerance and its manipulation for efficient fuel ethanol production. J Biotechnol 144: 23–30.

484) Zhaxybayeva O, Lapierre P, Gogarten JP (2004) Genome mosaicism and organismal lineages. Trends Genet 20: 254–260.

485) Zhu J, Pavelka N, Bradford WD, Rancati G, Li R (2013) Karyotypic determinants of chromosome instability in aneuploid budding yeast. PLoS Genet 8(5): e1002719. doi:10.1371/journal.pgen.1002719

486) Zwietering MH, Jongenburger I, Rombouts FM, van 'T Riet K (1990) Modeling of the bacterial growth curve. Appl Environ Microbiol 56: 1875-1881.