Microbiology Monographs

Series Editor: Alexander Steinbüchel

Gerold Barth *Editor*

Yarrowia lipolytica

Genetics, Genomics, and Physiology

Springer

Microbiology Monographs
Volume 24

Series Editor: Alexander Steinbüchel
Münster, Germany

For further volumes:
http://www.springer.com/series/7171

Gerold Barth
Editor

Yarrowia lipolytica

Genetics, Genomics, and Physiology

Editor
Gerold Barth
Inst. Mikrobiologie
TU Dresden
Dresden
Germany

Series Editor
Alexander Steinbüchel
Institut für Molekulare Mikrobiologie und Biotechnologie
Westfälische Wilhelms-Universität
Münster
Germany

ISSN 1862-5576 ISSN 1862-5584 (electronic)
ISBN 978-3-642-38319-9 ISBN 978-3-642-38320-5 (eBook)
DOI 10.1007/978-3-642-38320-5
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013944661

© Springer-Verlag Berlin Heidelberg 2013
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Yarrowia lipolytica is an outstanding yeast, phylogenetically distant from *Saccharomyces cerevisiae* or other well-studied yeast species and standing alone at the bottom of the tree of hemiascomycetous yeasts. It was formerly called *Candida lipolytica*, *Endomycopsis lipolytica*, and *Saccharomycopsis lipolytica* and finally reclassified by van der Walt and von Arx (1980) as the type species of the genus *Yarrowia*. This genus was newly created and dedicated to David Yarrow, who has identified this genus. To differentiate this and other yeasts from the two well-studied yeast species *S. cerevisiae* and *Schizosaccharomyces pombe*, the artificial group of "nonconventional yeasts" has been created and was first used in the title of the conference on "Genetics of non-conventional yeasts" held in Weimar (Germany) in 1987. In the meantime, *Y. lipolytica* is the best-studied organism of this group. The detection of sexual reproduction in this yeast by Wickerham and colleagues in 1970 has forced genetic investigations and development of genetic tools as well as construction of laboratory strain lineages in American, French, and German groups for comparative studies. The sequencing of the complete genome of this yeast to which groups from Canada, France, Germany, Japan, Korea, and Spain have contributed has further encouraged studies with this yeast since 2004.

Y. lipolytica is used now as model organism for several studies of academic interest like degradation of *n*-alkanes, utilization as well as accumulation of lipids, biosynthesis and degradation of peroxisomes, secretion of proteins as well as metabolites, stress response, dimorphism, alternative intron splicing, genome evolution, and analysis of mitochondrial respiratory chain complex I.

The special physiological features of *Y. lipolytica* and the acceptance of its generally recognized as safe (GRAS) status make this dimorphic yeast significant for biotechnological applications. Properties like intracellular accumulation of oil, production of hydroxyl or dicarboxylic acids, as well as secretion of large amounts of organic acids focus the interest on this yeast as a potential producer of basic commodities, fine chemicals, or building blocks for chemical industry in the post-alkane area. Furthermore, the high capacity for secretion of proteases and lipases favors this yeast as a producer of these enzymes as well as of heterologous proteins.

The aim of the two *Microbiology Monograph* volumes is to summarize the huge knowledge of the biology and applications of this fascinating yeast. The first volume covers the genetics, genomics, and physiology of *Y. lipolytica*. The second volume provides a broad survey of biotechnological applications of this yeast and discusses new developments.

The editor wants to thank all authors for their excellent contributions, their substantial efforts, and especially for their lots of patience. Many thanks are also given to the series editor, Alexander Steinbüchel, for his help and encouragement and to Jutta Lindenborn, Springer, for her help during the publishing process. Finally, the editor wants to thank the people behind the scenes, especially Falk Matthäus, who supported the project in putting into effect.

Dresden, Germany Gerold Barth
March 2013

Contents

Comparative Genomics of *Yarrowia lipolytica* 1
Claude Gaillardin, Meriem Mekouar, and Cécile Neuvéglise

Mitochondrial Genomics and Proteomics of *Yarrowia lipolytica* 31
Stefan Kerscher and Ulrich Brandt

Transposable Elements and Their Activities in *Y. lipolytica* 57
Serge Casaregola and Gerold Barth

Noncoding RNA Genes Transcribed by RNA Polymerase III in *Yarrowia lipolytica* 79
Cécile Neuvéglise, Claude Gaillardin, and Christian Marck

Utilization of Hydrophobic Substrate by *Yarrowia lipolytica* 111
Ryouichi Fukuda and Akinori Ohta

Ambient pH Signalling in *Yarrowia lipolytica* 121
Sylvie Blanchin-Roland

The Gpr1/Fun34/YaaH Protein Family in the Nonconventional Yeast *Yarrowia lipolytica* and the Conventional Yeast *Saccharomyces cerevisiae* 145
Falk Matthäus and Gerold Barth

Sulphur Metabolism of the Cheese-Ripening Yeast *Yarrowia lipolytica* 165
Agnès Hébert, Jean Marie Beckerich, Sophie Landaud, and Pascal Bonnarme

Index 185

Comparative Genomics of *Yarrowia lipolytica*

Claude Gaillardin, Meriem Mekouar, and Cécile Neuvéglise

Abstract Well known for its potential for industrial applications detailed in other chapters of this book, *Yarrowia lipolytica* occupies an interesting position among hemiascomycetous yeasts due to its basal position in the phylum associated with profound physiological differences setting it apart from the model species, *Saccharomyces cerevisiae*. Its study may thus shed light both on the history of yeast genomes and on the emergence of new solutions invented by yeasts to carry out basic functions like chromosome maintenance or mating in this highly diversified but phylogenetically coherent group. This chapter will give an overview of the data accumulated by the Génolevures consortium on the genome of *Y. lipolytica*. Since its first release in 2004, this genome appeared deeply different in terms of structure, gene content and organisation from that of other known hemiascomycetes. The message told by this unusual yeast genome was however difficult to understand fully for lack of intermediates providing a link to other genomes and of closely related genomes to sort out idiosyncrasies from ancestral characteristics. The recent publication of the genome of *Pichia pastoris*, a distantly related species but also deep branching, as well as our ongoing work on species closely related to *Y. lipolytica* (the "*Y. lipolytica* clade") or on a third deep branching species, *Arxula adeninivorans*, begins to fill some holes and may help building of an integrated view of yeast genome evolution.

Contents

1 Introduction 2
2 Genomic Data Generation 4
3 Chromosome Maintenance Elements 7
  3.1 Centromeres 7
  3.2 Origins of Replication 9
  3.3 Telomeres 10

C. Gaillardin (✉) • M. Mekouar • C. Neuvéglise
AgroParisTech, UMR INRA 1319 Micalis, 78850 Thiverval-Grignon, France
e-mail: claude@grignon.inra.fr

G. Barth (ed.), *Yarrowia lipolytica*, Microbiology Monographs 24,
DOI 10.1007/978-3-642-38320-5_1, © Springer-Verlag Berlin Heidelberg 2013

4 Non-protein Coding Genes 11
4.1 Ribosomal DNA 11
4.2 tRNA and Other Non-protein Coding Transcripts 12
5 Protein Coding Genes 12
5.1 Gene Counts, Gene Density and Polymorphism 12
5.2 Gene Models and Introns 14
5.3 *Yarrowia*-Specific Genes and Dubious Genes 15
5.4 Pseudogenes, Mutated Genes and Missing Genes 16
5.5 Gene Families and Family Expansions 17
5.6 Horizontal Transfer 18
6 Miscellaneous Functions Encoded by the Genome 20
6.1 Replication, Repair and Meiotic Recombination 20
6.2 Sex and Mating 20
6.3 mRNA Metabolism 21
6.4 Transporters 22
7 Mobile Elements 23
8 Mitochondrial DNA 25
9 Conclusion 25
References 26

1 Introduction

First talks among the *Y. lipolytica* community about getting the whole genome sequence of their favourite species were exchanged during a meeting at Peñiscola (Spain) in 1999. The only completely sequenced yeast genome at that time was that of *S. cerevisiae*, although there were rumours about launching *Schizosaccharomyces pombe* and *C. albicans* genome projects. This was the time when the Génolevures I programme led by JL Souciet (Strasbourg) was started in association with 7 French yeast labs. Thanks to the sequencing centre of Génoscope in Evry (France), this was the first project aiming at exploratory sequencing at low coverage of a whole phylum—the hemiascomycetous yeasts. Analysis of the genomes of 13 hemiascomycetous species including *Y. lipolytica* strain W29 (CBS 7504, CLIB 89) evidenced an unsuspected diversity, larger than that of the chordates, and confirmed that *Y. lipolytica* was indeed the species more distantly related to *S. cerevisiae* and actually to all other species (Casaregola et al. 2000; Souciet et al. 2000). Further exchanges during the next *Yarrowia* meeting at Dresden, Germany, in 2002 resulted in the formation of a small consortium that engaged in financing on its own lab money an exploratory BAC sequencing project [G. Barth (U. Dresden), U. Brandt (U. Frankfurt), A. Dominguez (U. Salamanca), C. Gaillardin (INRA Grignon), JW Kim (Chungnam National U.), D. Ogrydziak (U. Davis), R. Rachubinski (U. Alberta)]. A BAC library was constructed and subcontracted to the Clemson University Genomic Institute for fingerprinting and BAC-end sequencing. In parallel, already sequenced genes were assigned to chromosomes separated by

pulse-field gel electrophoresis (PFGE), thus completing formerly obtained maps (Casaregola et al. 1997). These data were quite instrumental in promoting the choice of *Y. lipolytica* as one of the species to be fully sequenced in the second stage of the Génolevures programme. The Génolevures II programme associating now 14 partners (see http://www.genolevures.org/about.html#members) decided to sequence completely the genomes of four yeasts, scattered through the genomic tree of hemiascomycetes. The full genome of strain E150 (CLIB 122) was published in 2004 (Dujon et al. 2004), together with those of *Kluyveromyces lactis* and *Debaryomyces hansenii*, two other yeasts of biotechnological interest, and that of *Candida glabrata*, a significant human pathogen distant to *Candida albicans* and closer to *S. cerevisiae*.

Many more yeast genomes have since been published, and a few clades have been quite extensively studied including those close to *S. cerevisiae* (so-called post-whole genome duplication (WGD) species (Dujon 2006)) or to *Kluyveromyces lactis* (pre-WGD or so-called protoploid species (Souciet et al. 2009)) or to *Candida albicans* (yeasts who have reassigned the CTG codon or CTG clade (Butler et al. 2009; Rossignol et al. 2008)). *Y. lipolytica* remained however the only representative of early branching hemiascomycetous yeasts and was sitting quite isolated at the bottom of the tree. Understanding yeast evolution was therefore limited by the possible peculiarities of this stand-alone genome, and more data were clearly required. The recent publication of the genomes of *Pichia pastoris* (De Schutter et al. 2009) and our current work on *Arxula adeninivorans* (Génolevures, to be published), two species predating the emergence of the CTG clade, now permit to put the *Y. lipolytica* data in perspective.

More recently, we could gather two sets of complementary genomic data. The first set aimed at refining the gene models, particularly introns, and to highlight possible cases of alternative splicing: three cDNA libraries obtained under different growth conditions were end sequenced (Mekouar et al. 2010). The second set aimed at evaluating the origin and conservation of some genomic peculiarities observed in the *Y. lipolytica* genome. Closely related species can indeed be very useful in order to annotate and exploit more comprehensively the genome of a given species (see, e.g. Galagan et al. 2005; Kellis et al. 2003). The *Yarrowia* clade had been defined a few years ago as comprising six species (Kurtzman 2005), but additional species, very closely related to *Y. lipolytica* and sometimes even mating competent with it, were recently identified (Knutsen et al. 2007; Limtong et al. 2008) (Fig. 1). We obtained (C. Neuvéglise, to be published) genomic data at partial or full coverage of four of these species: *Candida alimentaria*, *Candida deformans*, *Candida galli* and *Candida phangngaensis*. Preliminary results from this comparative analysis will be discussed.

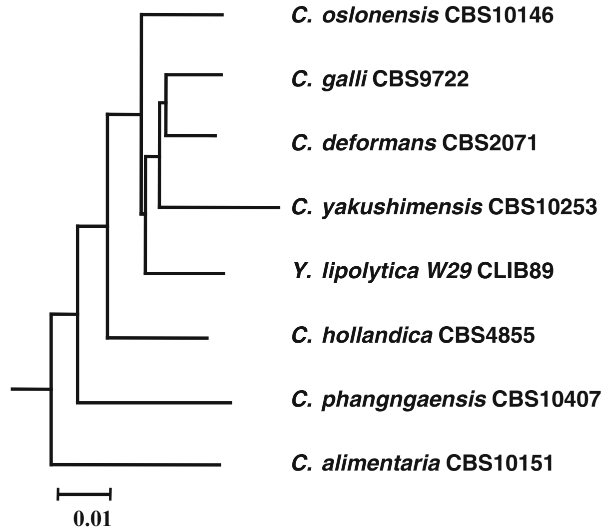


Fig. 1 Phylogeny of the species of the *Yarrowia* clade, based on 2,461 concatenated nucleotides of the genes encoding Act1, EF1-alpha and rDNA D1-D2 region

2 Genomic Data Generation

All sequences from *Y. lipolytica* were obtained through the Sanger technology using either dye primers (Licor 4200 DNA sequencer) or dye terminators (ABI 3700 Genetic Analyser). A first set of sequences was obtained during the Génolevures I programme. DNA was from the wild-type strain W29 (CLIB89, CBS7504). A total of 4,940 random sequences with an average size of 995 bp were obtained, corresponding to about 4.9 Mb or 0.4× genome coverage (Casaregola et al. 2000).

A second set of sequences was later obtained from the genetic strain E150 (CLIB122) (Barth and Gaillardin 1996). A total of 147,279 reads were obtained from plasmidic inserts (3–5 kb) and BAC ends (insert size: 75 kb) and represented about 10× genome coverage. Six large scaffolds could be assembled, each corresponding to one chromosome. Assembled supercontigs were named Yali0A to Yali0F corresponding to chromosomes I to VI (Casaregola et al. 1997). Sequence finishing was done by a combination of primer walking and shotgun sequencing of repeated regions present on BACs or plasmids, so as to get individual sequences for most repeated regions like transposons (Dujon et al. 2004). The current assembly still contains 13 gaps (5 in Yali0A, 3 in Yali0D, 4 in Yali0E and 1 in Yali0F). Telomeric ends could not be assembled due to the presence of repeated sequences, including rDNA repeats. The assembled part of the nuclear genome corresponds to 20.5 Mb or 98 % of the full size estimated to be around 21 Mb (Casaregola et al. 1997).

Annotation of the *Y. lipolytica* genome was achieved through a combination of ab initio predictions of coding regions based on GeneMark, ORF identification, splicing site predictions and similarity searches in the UniProtKB database.

Predictions were manually curated to define potential translation starts, to refine intron definition and to propose a functional annotation. RNA coding genes (tRNA, rRNA, snRNA... see below) were individually searched. Annotation is an ongoing process and recent data obtained through cDNA data have been used to detect new genes or new introns (Mekouar et al. 2010). New gene definitions were also constantly provided by scientists of the *Y. lipolytica* community (Desfougeres et al. 2008; Hirakawa et al. 2009; Morgner et al. 2008). All elements were numbered according to (Durrens and Sherman 2005). Thus, a feature (e.g. YALI0A00110g) is defined by four letters defining the genus and the species, one digit referring to the assembled version, one letter for the chromosome, five digits and the suffix g for protein coding gene, s for cis-acting element and t for repeated features such as transposons, for example. The October 2008 version of the annotated sequence is accessible at http://www.genolevures.org/yali.html#, and more recent updates on the proteome can be downloaded from http://genome.jouy.inra.fr/genosplicing/. An overview of genome organisation is given in Fig. 2.

Data at low genomic coverage (3×) from species close to *Y. lipolytica* were obtained through Roche 454 GS-Flex shotgun sequencing of the following strains: *C. alimentaria* (CBS10151, isolated from cured ham), *C. deformans* (CBS2071, isolated from fingernails) and *C. phangngaensis* (CBS10407, isolated from seawater). Full coverage (22×) of *C. galli* (CBS9722, isolated from poultry meat) was obtained through the Roche 454 Titanium technology (shotgun and paired ends sequencing). Preliminary analysis of genomic data combined to FACS analysis indicated that *Y. lipolytica*, *C. deformans*, *C. galli* and *C. alimentaria* have similar genome sizes in the range of 20 Mb, whereas the *C. phangngaensis* genome is significantly smaller, making this species particularly relevant for understanding genome dynamic in this clade. Nearly complete chromosomes could be assembled for *C. galli* (one scaffold per chromosome) and comparison with *Y. lipolytica* evidenced strong similarities in gene content, intron distribution and overall synteny. A preliminary comparison of the different species of the *Y. lipolytica* clade is given in Table 1. One can immediately notice that, in addition to a smaller genome size, *C. phangngaensis* exhibits several peculiarities when compared to the other species of the clade: lower GC content, reduced number of homologues to the *Y. lipolytica* CDS and reduced frequency of conserved introns. Other features will be discussed below.

Oriented libraries of cDNAs were obtained from *Y. lipolytica* cells growing exponentially on YPD glucose or on YNB oleic acid or from cells in stationary phase on YPD glucose. Since most *Y. lipolytica* spliceosomal introns are in the 5′ end of the transcripts, 5′ ends of about 10,000 cDNAs obtained under each condition were sequenced by the Sanger dye terminator technology and aligned on the genomic data.

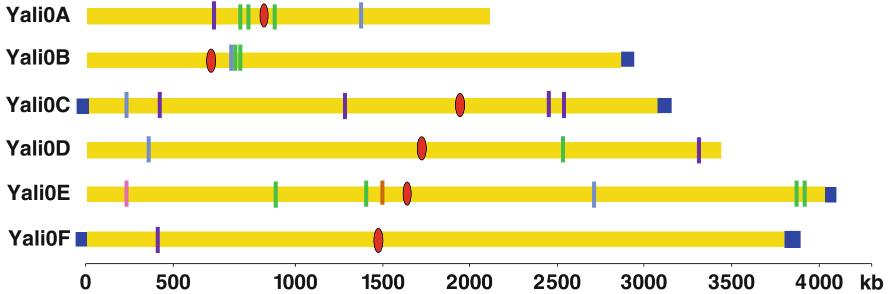


Fig. 2 *Yarrowia lipolytica* genome. Assembled contigs are depicted by *yellow bars* corresponding to 6 individual chromosomes. Sizes are given at the bottom in kb. Following elements are positioned: centromeres (*red ovals*), rDNA repeats (*blue squares*, not to scale). Full copies of mobile elements are depicted by *vertical bars*: *green* for Ylt1, *Brown* for Tyl6, *violet* for Ylli, *pink* for Fotyl and *blue* for Mutyl

Table 1 Genomic data from *Y. lipolytica* (YALI), *C. galli* (YAGA), *C. deformans* (YADE), *C. phangngaensis* (YAPH) and *C. alimentaria* (YAAL)

	YALI	YAGA	YADE	YAPH	YAAL
Ploïdy	Haploïd	Haploïd	Haploïd	Haploïd	Haploïd
Mating type	MATB	MATB	MATA	MATA	MATB
GC content	49 %	49 %	50 %	43 %	49 %
YALI homologues	(6,451)	6,194	6,066	5,407	5,843
Conservation/YALI					
• Identity		84 %	82 %	74 %	75 %
• Similarity		89 %	87 %	83 %	83 %
Introns					
• Conservation	(1,083 introns, 952 genes)		97 %	92 %	97 %
tRNA fusion					
• tRNA/5S	+	+	+	+	+
• tRNA/RUF70	+	+	+	+	+
Transposable elements					
• Gypsy	Ylt1, Tyl6	+	+	+	+
• Copia	–	+	Tyd5	+	+
• LINE	Ylli	+	+	+	+
• Mutator	Mutyl	–	–	–	+
• Pogo	Fotyl	+	+	–	+
Protein families					
Lipases	16	14	15	8	11
Esterases	4	3	4	2	4
Elongases	2	2	2	2	2
AcylCoA oxidases	6	6	6	5	6
Cytochrome P450	12	12	12	8	15
Aspartyl proteases	38	59	48	12	38
Alkaline proteases	10	12	9	2	9

3 Chromosome Maintenance Elements

Centromeres and chromosomal origins of replications have been functionally characterised through an autonomously replicating sequences (ARS) screening strategy (Fournier et al. 1993; Vernis et al. 1997). Chromosomal sequences able to confer extrachromosomal maintenance to plasmids in *Y. lipolytica* could be isolated at a very low frequency, and only two such sequences could be initially selected. Subsequent work showed that these *ARS* contained both an origin of replication (*ORI*) and a centromere (*CEN*), as indicated by the following arguments. *ARS* presence resulted in rather stable extrachromosomal maintenance of plasmids, but these were present at only one copy per cell and segregated 2:2 through meiosis. Single deletion from the chromosome of any one *ARS* was lethal, but *ARS* exchange at the locus was viable. Directing ectopic integration of an *ARS* resulted in low transformation frequencies and led to splitting of the target chromosome. This phenotype permitted subcloning of the *CEN* sequences. Minimal *CEN* sequences were in turn used to fish out chromosomal *ORI*s, and minimal *ORI*s could be used to fish out additional *CEN*s. Why replicating plasmids need both an *ORI* and a *CEN* to be maintained extrachromosomally is still mysterious, and various hypotheses have been put forward (Vernis et al. 2001).

3.1 Centromeres

Only five out of the six expected centromeres could be isolated through the above screen (Matsuoka et al. 1993; Vernis et al. 2001). Six *CEN* sequences are however available since Yali0A *CEN* contains two adjacent sequences (*CEN1-1* and *CEN1-2*), each endowed with full *CEN* activity. No *CEN* could be isolated from Yali0F. Centromeres appear to be short sequences (in the range of 200–250 bp), rather AT rich (>70 % vs. ~50 % for the whole genome), with no obvious sequence conservation. Recent work on *Y. lipolytica CEN1-1* (Yamane et al. 2008a, b) identified an imperfect palindromic sequence, able to bind site-specific (uncharacterized) proteins, as necessary for plasmid transmission in *Y. lipolytica*. A 17- to 21-bp consensus sequence was retrieved in five out of the six *CEN* sequences so far characterised, but not in *CEN4*. When the genome was surveyed, we observed that each of the five functionally defined *CEN* regions was located in a large (2–3 kb) GC poor region (see Fig. 3a), including *CEN4*. There was only one such large GC anomaly per chromosome. The one localised on Yali0F did contain a version of the palindromic consensus (coordinates 1945874-1945894) (see Fig. 3b). We thus propose to assign *CEN6* to this region instead of the one proposed in (Dujon et al. 2004).

Similar GC anomalies do exist in the assembled chromosomes of *C. galli* and the same 21-bp consensus sequence can be identified in some of them at least. Although definite conclusions await direct experimental confirmation, it seems highly likely

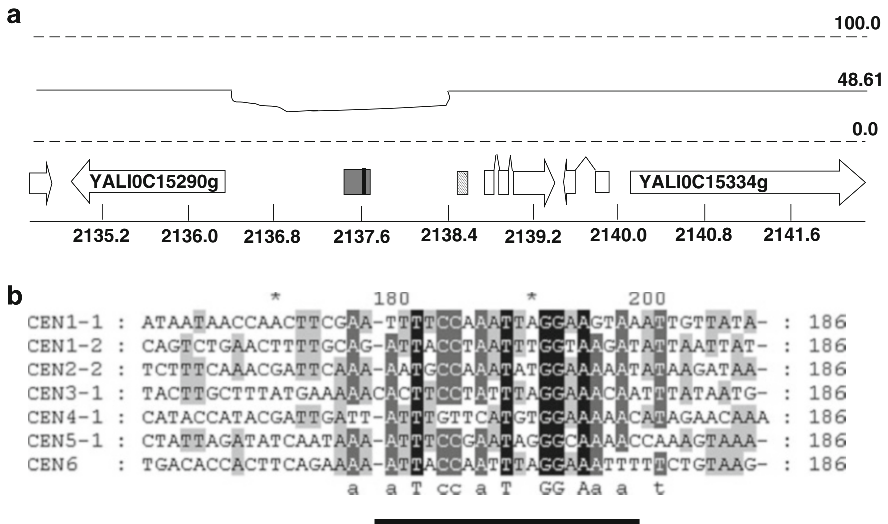


Fig. 3 Structure of centromeric regions. (**a**): *ARS18* region on Yali0C chromosome, showing positions of functionally characterised *CEN3-1* (*grey box*) and ORI3018 (*hatched box*), as well as adjacent CDS (*open arrows*). The 21 bp imperfect palindrome found in 5/6 *CEN*s is indicated by a *vertical black bar* (see text). The average GC % with a 2.5 SD cut-off and a 1,000 nt sliding window is shown on top (GC % are indicated on the *right*). (**b**) Partial alignment view of the 186 nt region used by (Yamane et al. 2008a, b) to define protein-binding sites in the centromeres of *Y. lipolytica*. Seven centromeric sequences are shown, including a putative match in *CEN4-1* and the newly proposed match in *CEN6* (see text). The 17–21 bp degenerated consensus sequence is *underlined*

that similar centromeric structures exist in all species of the *Yarrowia* clade, so that this structure predated emergence of the clade.

*Y. lipolytica CEN*s differ from the *S. cerevisiae* or *K. lactis* centromeres which contain the canonical CDI-CDEII-CDEIII structure (Hegemann and Fleig 1993; Heus et al. 1994). They also differ from the recently described *C. albicans* centromeres which have been reported to be larger, epigenetically defined regional structures with no obvious sequence conservation (Mishra et al. 2007; Sanyal et al. 2004). There is presently no evidence in *Y. lipolytica* for centromere neoformation upon *CEN* deletion as reported in *C. albican*s (Ketel et al. 2009). Analysis of the kinetochore-associated proteins (Meraldi et al. 2006) showed that *Y. lipolytica* was devoid of a set of 13 kinetochore proteins associated to point centromeres in *Saccharomyces* and *Kluyveromyces* yeasts but had retained Sim4, Fta1, Fta3 and Fta4 homologues which are absent from the former set of yeasts and characteristic of organisms with regional *CEN*s. This suggests that *Y. lipolytica* may represent an early stage of the evolution of regional centromeres toward punctual structures, different from the route taken by *C. albicans*.

3.2 Origins of Replication

Isolation of several chromosomal origins of replication was achieved by cloning in a *CEN* vector sequences able to promote autonomous replication (Vernis et al. 1999): a variety of origins were isolated, either unique or repeated in the genome, suggesting that, as in *S. cerevisiae*, there was on average one origin of replication every 20–50 kb. The unique origin *ori4021* as well as some at least of the repeated rDNA origins were shown to be active in the chromosomes by visualising replication intermediates on two-dimensional (2D) gels. Contrary to higher eukaryotic origins, where most initiation events occur within initiation zones of several kb (Mesner et al. 2006), initiation events on *Y. lipolytica* plasmids or chromosomes were localised at a discrete site within the *ORI*s. Repeated *ORI*s were found to correspond to the rDNA NTS region, a situation also reported in *S. cerevisiae* and *S. pombe* (Linskens and Huberman 1988; Sanchez et al. 1998) and to ends of chromosomes (see oriX009 in Sect. 3.3 below). The third repeated candidate oriX096 is likely to result from a cloning artefact, since it corresponds to three distinct regions of chromosomes F and C (unpublished observation).

Origins could be trimmed down to 115–159 bp and moved without losing activity to various chromosomal contexts (2D gel analysis) or to various *CEN* plasmid locations, suggesting that, as in *S. cerevisiae*, replication origins are short and context independent. They do not display the typical *S. cerevisiae* 11-bp ARS consensus (WTTTAYRTTTW). All *ORI*s except oriX009 contain 1–5 repeats of a TDCAAGTH motif, but nucleotides within this sequence can however be shuffled without impairing plasmid replication. So far, no consensus sequences could be identified in *Yarrowia* origins; neither was any significant compositional bias observed nor any specific structural constraint (bent DNA, regions of low ΔG facilitating DNA unwinding, etc.) (Vernis et al. 1999). *Yarrowia* origins thus share their small size with *S. cerevisiae* and their absence of consensus with *S. pombe* and higher eukaryotes (Cvetic and Walter 2005).

The six subunits of the origin recognition complex (ORC) are encoded by YALI0D10104g (*ORC1*), YALI0D22330g (*ORC2*), YALI0F14773g (*ORC3*), YALI0E15928g (*ORC4*), YALI0B01452g (*ORC5*) and YALI0F31647g (*ORC6*), although the two latter have diverged considerably from their homologues in *S. cerevisiae*. It should be stressed that while the ORC complex appears to bind to specific DNA sequences in the budding yeast, there is increasing evidence that metazoan ORC complexes do not rely at all on sequence-specific sites to be directed to origins of replication. In *S. cerevisiae*, a global survey of origins active in vivo (Xu et al. 2006) showed that the 11-bp *ARS* consensus (ACS) was not sufficient for origin function: the majority of ACS matches did not function as ORC binding sites. Identifying the mechanism of ORC recognition at *Y. lipolytica* origins thus remains an open question which may shed light on the evolution of origins of replication in hemiascomycetes.

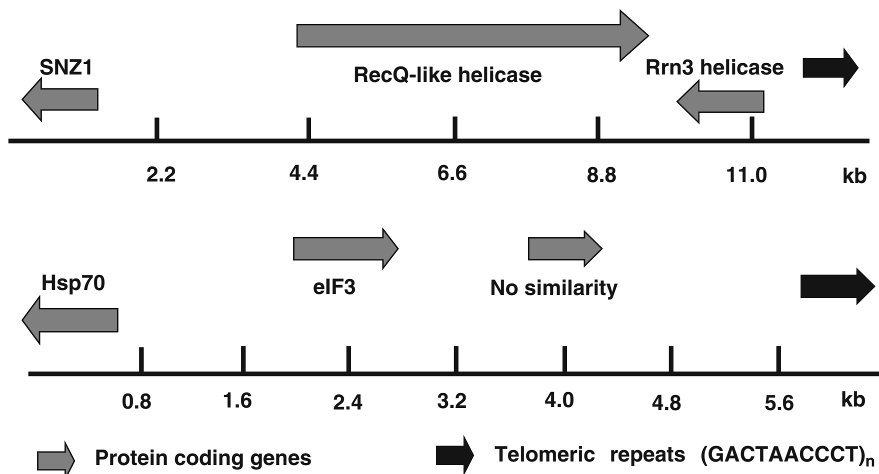


Fig. 4 Telomeric ends of *Y. lipolytica* chromosomes. Two examples of telomere ends are shown; they have not been assigned to a particular chromosome arm, due to the presence of upstream subtelomeric repeats. *Black arrows*: telomeric repeats, *grey arrows*: predicted genes. Sizes in kb

3.3 Telomeres

Telomeric sequences were not contigated into the chromosomes in the current Génolevures release due to the presence of repeated sequences in the subtelomeric regions. These consisted either of rDNA clusters or of Ylt1 transposon remnants and various types of protein coding genes repeated on each end. Two types of telomeric ends could be assembled (see Fig. 4). The telomeric repeats were identified as (5′-GGGTTAGTCA-3′) decameres tandemly repeated 40–50 times on single reads (sequence reported in (Dujon et al. 2004) is erroneous).

In *S. cerevisiae*, the telomerase activity depends in vivo on five genes encoding two subunits of the catalytic core (the reverse transcriptase Est2p and the template RNA Tlc1) and three regulatory subunits which are dispensable for in vitro activity (Est1p, Est3p and Cdc13p). The proteins of the *Y. lipolytica* telomerase complex have widely diverged from their homologues in other hemiascomycetous yeasts, and only an *EST2* homologue (YALI0D12188g) could be identified (Dujon et al. 2004). Mutants affecting *YlEST2* have been analyzed (Kinsky et al. 2010) as well as a new telomere binding protein YlTay1p with two MybB domains (Kramara et al. 2010). The gene encoding Tlc1 could not be identified unambiguously.

A highly conserved (157–159 nt/159 nt) chromosomal origin of replication (oriX009, see above) is present exclusively at ~600 bp from the ends of the assembled chromosomes when they terminate with rDNA clusters. It has been previously reported, through tagging of oriX009 with *Sce*I, that this sequence is located 20–420 kb away from chromosome ends (Vernis et al. 1999). This suggests that oriX009 is centromere proximal to the subtelomeric rDNA structures and is not an equivalent of the telomere-associated *ARS* of *S. cerevisiae* (Newlon et al. 1993).

4 Non-protein Coding Genes

Non-protein coding genes were defined by a combination of methods (similarity or pattern searches) complemented by structural predictions and sequencing of cDNA libraries. The set described below is for the time being the most complete concerning a species basal to the CTG clade, but is still incomplete.

4.1 Ribosomal DNA

Earlier studies have shown that 35S rDNA in *Y. lipolytica* is encoded by several types of units and that 5S RNA genes are not part of the rDNA repeat but present as multiple copies dispersed throughout the genome. Heterogeneity of 35S rDNA repeats is mainly due to length variations (from 1 to 2 kb) of the non-transcribed spacer (NTS) region, which exhibits a complex pattern of repetitions in periodicities of about 150 bp (van Heerikhuizen et al. 1985; see for an example gb|M14402). At least five different types of units have been characterised in the parental strain of E150, and different strains of *Y. lipolytica* were shown to harbour different types of units. Tetrad analysis and R-looping experiments strongly suggested a clustered organisation of each type of 35S unit in homogeneous, continuous tracts of different size (20 to more than 400 kb in E150), although meiotic exchanges could occur within the rDNA clusters (Casaregola et al. 1997; Fournier et al. 1986; van Heerikhuizen et al. 1985). The number of 35S rDNA units within each cluster appears to be highly variable from strain to strain (Clare et al. 1986; Le Dall et al. 1994), which may account for the pronounced chromosomal polymorphism observed in the species (Naumova et al. 1993).

Although specific 35S rDNA clusters could not be assigned to individual chromosomes in the E150 assembly, they were mapped to six ends of four chromosomes (see Fig. 2). In addition, about 110 copies of 5S RNA were found dispersed in the genome. A similar organisation of rDNA genes has been recently reported in the case of *Pichia pastoris*, where 35S rDNA genes are clustered on all four chromosomes in a subtelomeric position and 5S RNA genes are dispersed (De Schutter et al. 2009). This type of organisation of rDNA genes is thus likely to represent the common situation in some yeasts that diverged before the apparition of the CTG clade (yeasts which reassigned the CUG codon to serine instead of leucine (Fitzpatrick et al. 2006; Wang et al. 2009)), but not in all: a single rDNA locus is present in *A. adeninivorans*, but 5S RNA genes are dispersed as in *Y. lipolytica* and in *P. pastoris*.

In *Y. lipolytica*, and in the four other species of the clade that have been sequenced, an unusual organisation of 5S RNA genes was observed: about half of the 5S RNA copies existed as dicistronic tRNA-5S RNA fusions (for a more thorough description, see Neuvéglise et al. 2013). This type of tRNA fusions appears so far restricted to yeasts of the *Yarrowia* clade.

4.2 tRNA and Other Non-protein Coding Transcripts

A complete survey of PolIII transcribed genes is presented in a companion chapter in this book (Neuvéglise et al. 2013), and the reader is referred to it for more details.

The 7SL RNA, component of the signal recognition particle, is encoded by two paralogues which are both actively transcribed: *SCR1* and *SCR2* (He et al. 1990; Poritz et al. 1988). We note that the published sequences, derived from strains related to E150 but not isogenic, differ by two substitutions each from the E150 sequence and that *YlSCR1* lies only 67 nt upstream from the proposed translation start of an adjacent CDS, YALI0A01375g.

The PolII transcribed genes for the small nuclear RNAs (snRNAs) involved in spliceosomal intron excision, snRNA U1, U2, U4 and U5, have all been identified as well as the PolIII transcribed U6 (Dujon et al. 2004). There are two copies of U1 in E150 genome, both matching the most frequent 5′ splicing sites (GTGAGT and GTAAGT). The snRNA H1, a component of Rnase P, was also identified (Kachouri et al. 2005), but as said above the RNA component of the telomerase could not.

The small nucleolar RNA (sno) involved in rDNA modification and/or processing U3 is present in three copies (vs. two in *S. cerevisiae*), whereas snoR52 is split into two separate genes, snR52-A and snR52-B (see Neuvéglise et al. 2013). Other snoRNAs could not be identified until now.

As a result of our recent survey of cDNA libraries, at least 1,100 reads matching ORF-free intergenic regions have been identified (Mekouar et al. 2010). Further work and comparative studies (Steigele et al. 2007) within the *Yarrowia* clade are thus likely to expand the current repertoire of non-coding RNAs.

5 Protein Coding Genes

This set of gene was defined based on ab initio predictions (such as those based on GeneMark), ORF predictions and similarity searches. Gene models were refined manually and completed by data obtained from cDNA libraries (full-length cDNA or high-throughput RNA-seq).

5.1 Gene Counts, Gene Density and Polymorphism

6,449 protein coding genes are currently present in the database, which is slightly more than *S. cerevisiae* (5,807 CDS) for an assembly size about 1.7 times larger (20.5 Mb vs. 12.2 Mb for *Y. lipolytica* vs. *S. cerevisiae*, respectively). The average gene density (1 gene for 3 kb) is thus rather low but varies along the chromosomes: chromosomes are made of alternating blocks of densely packed regions (more than

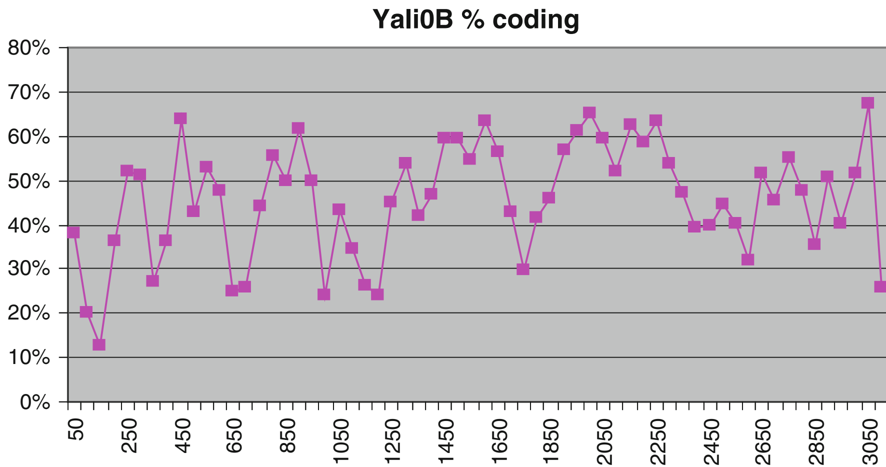


Fig. 5 Gene density along *Y. lipolytica* chromosomes. The situation observed in Yali0B is given as an example, but similar organisations are found on all chromosomes. Gene density is expressed as the percentage of ORF nucleotides in a sliding 50 kb window. Yali0B sequence is at the *bottom*. Sizes are in kb

40 % and up to 70 % coding sequences in specific 50 kb regions) and gene poor areas (less than 40 % and down to 20 % in others) (see Fig. 5).

Since CDS conservation between *Y. lipolytica* and other yeasts is often poor, the choice of the translation initiation codon is often somewhat arbitrary: in case of strong sequence conservation, the ATG was chosen by similarity; if not, the most upstream one was chosen. Comparing possibly starts in species of the *Yarrowia* clade and analysis of cDNAs will improve this definition, as well as taking into account statistics on ATG environment (see Fig. 6).

Polymorphism of protein coding sequences among different strains can be estimated from individual entries in GenBank reported independently from the Génolevures project (currently nearly 500): there are on average two SNPs/kb and more than 20 for some genes.

Similar gene contents are predicted for the four other species analysed in the *Yarrowia* clade (see Table 1) with a strong conservation of amino acid sequences (74–84 % identity on average).

One-third of all CDS are conserved among nine hemiascomycetous species taken all across the tree (*S. cerevisiae*, *C. glabrata*, *Z. rouxii*, *S. kluyveri*, *K. thermotolerans*, *K. lactis*, *E. gossypii*, *D. hansenii* and *Y. lipolytica*), and 1,024 correspond to singletons (probable direct orthologues) which have also homologues in most Euascomycetes and Basidiomycetes (P. Durrens, unpublished data).

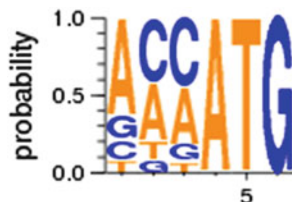


Fig. 6 ATG environment in currently defined CDS. The full set of CDS currently annotated in the genome of *Y. lipolytica* was used to compute the probability of a given nucleotide at positions -1, -2 and -3 (http://weblogo.berkeley.edu/logo.cgi). As in other hemiascomycetes, a purine (preferably an A) is present at position -3 in 80 % of the cases

5.2 Gene Models and Introns

Manual curation and alignment of cDNA reads permitted the detection of a total of 1,128 introns, including 36 present in the 5′-untranslated region (5′-UTR) of mRNAs. With 1,009 genes encoding intronic mRNAs or 16 % of its total repertoire, *Y. lipolytica* is currently the most intron-rich hemiascomycete. This seems to hold true for the four other species of the *Yarrowia* clade, where intron positions within the corresponding genes are largely conserved, although here again *C. phangngaensis* looks a bit atypical (Table 1): introns are on average shorter than in the other species and several have been apparently precisely erased.

Intron detection in *Y. lipolytica* is facilitated by a strong conservation of the intron splicing sites (Mekouar et al. 2010). The 5′ splicing (donor) site is mostly GTRRGT (GTGAGT in 87 % of the cases) instead of GTRTGT as in most other hemiascomycetes. The last base of the upstream exon is a purine in 80 % of the cases (preferentially a G). The internal splicing site is cTAAC (instead of TACTAAC in *S. cerevisiae*) and lays 0–2 bases upstream from the 3′ downstream acceptor site, most often YAG. Average size of introns is around 280 nt, but there are 16 introns larger than 1 kb (up to 3.4 kb) and around 300 that are shorter than 100 bp (see also http://genome.jouy.inra.fr/genosplicing/index.html). Introns are usually in the 5′ part of the gene and display conspicuous sequence biases in order to generate in-frame translation stop codons in case they would not be spliced out: these in-frame translation stop codon likely trigger nonsense-mediated mRNA decay (NMD) of abnormal mRNAs. Components of the NMD machinery have been identified (*UPF1* and *UPF2* homologues, *UPF3* being apparently missing) (Mekouar et al. 2010). Analysis of cDNA libraries indeed evidenced a significant proportion of unspliced mRNA as well as a few cases of alternative splicing events: exon skipping, alternative 5′ and 3′ sites and alternative transcription initiation sites associated with intron inclusion in the primary transcript (see, e.g. Fig. 7 and Neuveglise et al. 2005).

Interestingly, other yeast species that diverged before emergence of the CTG clade share related intron structures, which may thus be ancestral for hemiascomycetes. The GTRRGT consensus is predominant in *Arxula*

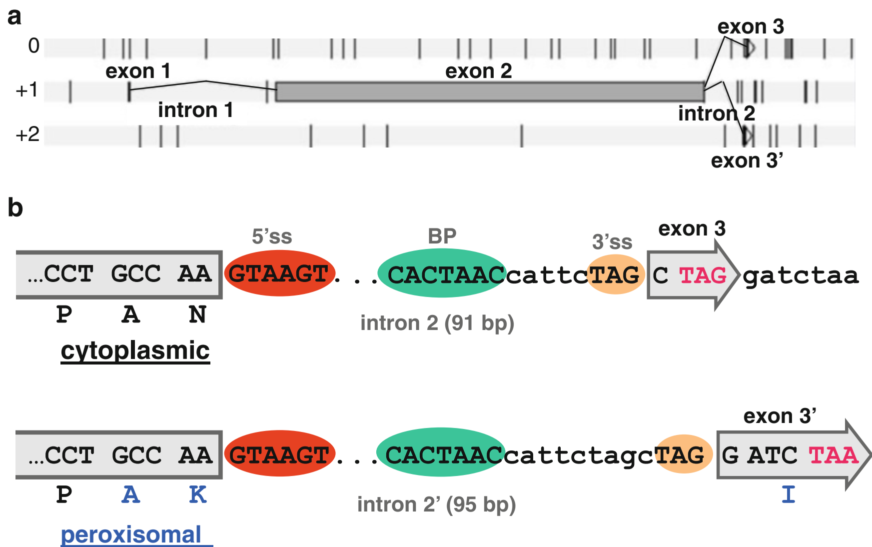


Fig. 7 Predicted localisation of YALI0E14190p in peroxisome or cytoplasm depends on intron 2 alternative splicing. (**a**) Gene models for YALI0E14190g with two introns. Exons are represented by *grey rectangles* and introns by *thin black lines*. Alternative splicing of intron 2 generates two different terminal exons (exon 3 and exon 3′) on two different phases (indicated by 0, +1 and +2 on the *left*). (**b**) Detailed representation of introns 2 and exons 3. Intron features are represented by coloured ovals: 5′ splice site in *red*, branch point in *green* and 3′ splice site in beige. When intron 2 is 91 nt long, the resulting mRNA is translated into a cytoplasmic protein; when it is 95 nt long, the resulting protein carries a C-terminal AKI signal for peroxisomal localisation

adeninivorans as in *Pichia pastoris* (mostly GTAAGT in both cases), and the distance between the internal and the 3′ splicing sites is short (2–3 nt in *A. adeninivorans* and 7 nt in *P. pastoris*). These yeasts have smaller genomes than *Y. lipolytica* and its relatives (excluding *C. phangngaensis*), and their intron sizes tend also to be shorter (average around 80 nt in *P. pastoris*). This suggests that yeasts of the *Yarrowia* clade (including *Y. lipolytica*) underwent intron extension, perhaps as part of a general relaxation of genome size constraints.

5.3 Yarrowia-*Specific Genes and Dubious Genes*

The current annotation of the genome largely reflects the 2004 situation, since the 2008 update only corrected several gene models (intron mispredictions) and introduced a few genes that had been identified by the *Yarrowia* community (Desfougeres et al. 2008; Hirakawa et al. 2009; Morgner et al. 2008). Some 1,030 predicted proteins are thus presently annotated in the Génolevures database as having no match to any published sequence ("no similarity") or as dubious genes ("no similarity, possibly non-coding"). A preliminary analysis of the *C. galli*

proteome indicates that most of these genes (838/1,030) are conserved in *C. galli* and are thus likely to represent true genes. A recent search for putative homologues of the 1,030 genes evidenced that 8 only had no matches in databases (including *Y. lipolytica* paralogues) or in *C. galli*. In conclusion, most of the 6,443 genes predicted in the *Y. lipolytica* database are likely to be real genes.

While most of the 1,030 genes previously annotated as having no similarity are still represented by singletons or paralogous families with no similarity outside of the *Yarrowia* clade, possible homologues for 148 of them can now be found in databases including 93 in fungi (mostly Euascomycetes) and 45 in other hemiascomycetes like *P. pastoris* (12 genes) or yeasts of the CTG clade (28 genes). This illustrates once again the basal position of *Y. lipolytica* in the yeast phylum. The function of most of these genes remains unknown, but YALI0C14762g, previously annotated as having no similarity, has recently been identified as encoding the transcriptional factor Yas3 regulating cytochrome P450 expression (Hirakawa et al. 2009), and the near-complete complement of nuclear-encoded Complex 1 proteins has now been identified (see Kercher and Brandt 2013).

5.4 Pseudogenes, Mutated Genes and Missing Genes

A total of 140 pseudogenes have been identified. Some are likely to result from sequencing errors since intact copies have been identified previously in *Y. lipolytica*: YALI0F00748g (*YlPAY2*; Eitzen et al. 1995) and YALI0F21054g (*YlSTE7*, see gi| AJ007393). Three pseudogenes correspond to genes that have been purposely mutated during the construction of E150 (Barth and Gaillardin 1996): YALI0E26741g (*ura3-302*, interrupted by an *XPR2::SUC2* fusion), YALI0F31889g (*xpr2-322*, internal *Apa*I deletion) and YALI0C00407g (*leu-270*, internal *Stu*I deletion). Wild-type gene sequences from other, but closely related, *Y. lipolytica* strains have been deposited: gb| U40564 for *URA3*, gb|M23353 and gb| M17741 for *XPR2* and gb|M37309 and gb|AF260230 for *LEU2*. The fourth mutation in E150 (*his1-1*) affects YALI0C07128g homologue of *HIS7* in *S. cerevisiae*, while *lys11-23* in the congenic strain E129 affects YALI0F31075g encoding homocitrate synthase (S. Kerscher, pers. comm.).

The remaining pseudogenes can be categorised into three classes. Class I (42 cases) corresponds to transposon remnants for which complete copies have been identified in the genome (Mutyl, Fotyl, Ylli, Ylt1; see below) or not: YALI0C24442g (a probable relic of a non-LTR retroposon), YALI0B07766g and YALI0B07766g (relics of two different Ty3 gypsy LTR-retrotransposon). Class II pseudogenes (47 cases) derive from genes encoding homologues of known function in other organisms. Most of these have suffered multiple deletion or frameshift events and are therefore likely to be true pseudogenes. Finally, 46 pseudogenes either share similarity with *Yarrowia*-specific genes of unknown function or have no similarity at all.

Genes for amino acid, nucleotide and vitamin biosynthesis could all be identified in the assembly, with three exceptions. No homologue of *THI80* (*S. cerevisiae* YOR143c or sp|P35202) encoding thiamine pyrophosphokinase could be found, which may account for the thiamine requirement exhibited by *Y. lipolytica* strains. Other biosynthetic genes missing in the assembly could be identified in the non-assembled set of reads. This was the case for homologues of *ARG1/ARG10* (sp|P22768 or YOL058w) encoding argininosuccinate synthetase (JM. Beckerich, unpublished) and *CDC21/TMP1* (sp|P06785 or YOR074c) encoding thymidylate synthase.

5.5 Gene Families and Family Expansions

All proteins predicted from sequencing of various yeast species genomes have been clustered in gene families (Sherman et al. 2004, 2009). This resulted in the assignation of 5,113/6,449 genes (85 %) in gene families, either *Y. lipolytica* specific or shared by two or more species among *S. cerevisiae*, *C. glabrata*, *Z. rouxii*, *S. kluyveri*, *K. thermotolerans*, *K. lactis*, *E. gossypii*, *D. hansenii*, *P. sorbitophila*, *A. adeninivorans* and *Y. lipolytica* (P. Durrens, unpublished data). By combining similarity and synteny data, orthologous series have been defined across hemiascomycetous yeasts for some families like sugar transporters, carboxylate transporters or heavy metal transporters (De Hertogh et al. 2006; Diffels et al. 2006; Lodi et al. 2007; Palma et al. 2009), thus refining function prediction for *Y. lipolytica* genes.

The general conclusion is that *Y. lipolytica* roughly displays the same level of gene duplication as other yeasts, but the mechanism of duplication seems to have been different. Most paralogues are dispersed in an apparent random manner across the genomes, ruling out known mechanisms of gene duplication like WGD as in *S. cerevisiae* and tandem duplications as in *D. hansenii* or segmental duplication (Dujon 2006).

Significantly amplified gene families correspond mainly to lipid or more generally hydrophobic compound metabolism (18 2–4 dienoyl CoA reductases, 16 acylglycerol lipases, 4 esterases , 12 cytochrome P450 of the CYP52 family, 11 sphingomyelinases, 6 acyl CoA oxidases, 14 fatty acid transporters) and to degradation of proteinaceous substrates (38 aspartyl proteases, 15 serine proteases). Amplification of these families parallels well-known phenotypic characteristics of *Y. lipolytica*. Most of these amplified gene families are also amplified in the *C. alimentaria*, *C. deformans* and *C. galli* (see Table 1). Once again, however, *C. phangngaensis* appears as an outlier and shows a drastic reduction of some families. On the contrary, *C. galli* and *C. deformans* amplified aspartyl proteases to an even higher degree (see Table 1). This suggests that gene families are still highly dynamic in the clade, with recent gene losses and amplification as exemplified in the case of lipases of *Y. lipolytica* and *C. galli* (see Fig. 8).

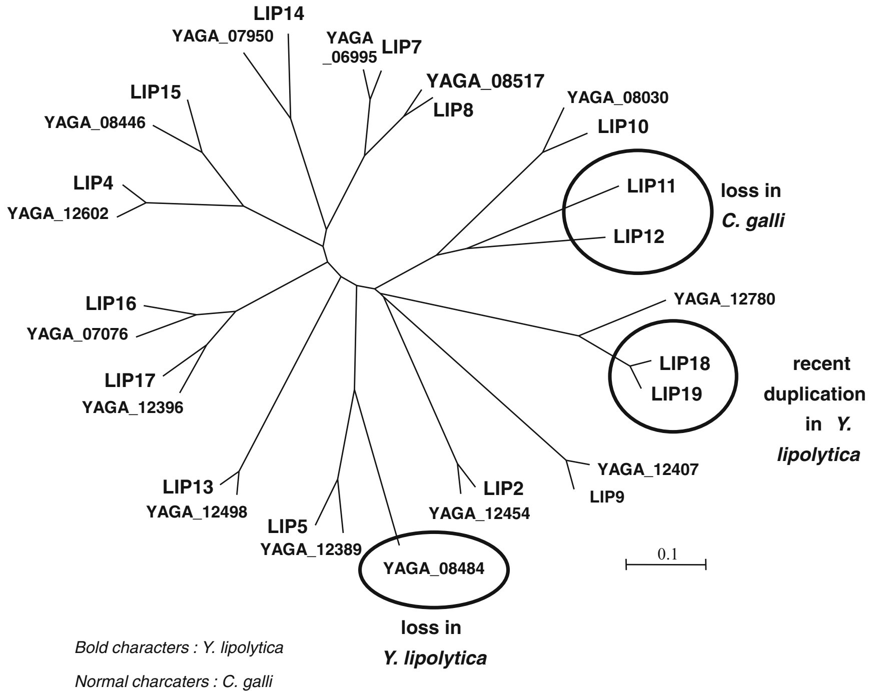


Fig. 8 Lipase families in *Y. lipolytica* and *C. galli*. All paralogues were identified by Blast searches and aligned using ClustalW. *LIP18* and *LIP19* appear as recent duplications in *Y. lipolytica*, *LIP11* and *LIP12* have apparently been lost from *C. galli* and *YAGA_08484* from *Y. lipolytica*

5.6 Horizontal Transfer

Dujon et al. (Dujon et al. 2004; Rolland et al. 2009) reported six possible cases of horizontal transfer from bacteria. Seventeen additional *Y. lipolytica* genes are currently annotated against bacterial genes in the Génolevures database, representing altogether 15 gene families. Most putative donors are soil bacteria. Revisiting annotations evidenced a maximum of ten genes of possible bacterial origin, whereas four genes initially annotated against bacteria are actually orphans and nine have more or less distant relatives in filamentous fungi (Table 2). Interestingly, the only case of intron containing gene, YALI0C22121g, occurs in a family of probable ancient fungal origin. A puzzling question is why have these genes been selected in the first place and then lost from other hemiascomycetous genomes.

Table 2 Updated list of possible cases of horizontal gene transfer

Gene (paralogues)	Current Génolevures annotation	Best fungal score	Best bacterial score	Possible origin
YALI0A15400g	tr|Q92QU2 *Rhizobium meliloti* Putative acetyltransferase protein	No match	1e-27	Bacterial
YALI0A15444g, YALI0A15466g, YALI0C03267g	gb|EET71812 *Geobacillus* sp. GCN5-related *N*-acetyltransferase	No match	5e-3	Bacterial ?
YALI0C03443g, YALI0F11605g	tr|Q89C37 *Bradyrhizobium japonicum* Blr7961 protein	1e-13	3e-15	Fungal
YALI0C15026g, YALI0F15323g	tr|Q54300 Streptomyces hygroscopicus methyltransferase	5e-15	2e-16	Fungal
YALI0C15230g, YALI0C22121g, YALI0D03597g	tr|Q8D5F7 *Vibrio vulnificus* predicted thioesterase	2e-18	5e-16	Fungal
YALI0D21582g, YALI0F01408g	tr|Q87HL8 *Pseudomonas putida* Yee/YedE family protein	3e-47	4e-17	Fungal
YALI0D06391g	tr|Q1D675 *Myxococcus xanthus* acetyltransferase, GNAT family	No match	2e-07	Bacterial
YALI0D25014g	tr|Q97P71 *Streptococcus pneumoniae* cell wall surface anchor family protein	No match	No match	Unknown
YALI0D26411g	tr|Q884V4 *Pseudomonas syringae* chemotaxis sensor histidine kinase CheA	No match	No match	Unknown
YALI0E31031g, YALI0A01089g	tr|Q88XB6 *Lactobacillus plantarum* cell surface SD repeat protein precursor	No match	No match	Unknown
YALI0E33011g	tr|P45900 *Bacillus subtilis* adenylate kinase family	No match	8e-34	Bacterial
YALI0F04290g	tr|Q987V4 Rhizobium loti D-amino peptidase	No match	5e-116	Bacterial
YALI0F05654g	tr|Q8EAT4 *Shewanella oneidensis* Conserved hypothetical protein	No match	8e-82	Bacterial
YALI0F31867g	tr|Q9I5L7 *Pseudomonas aeruginosa* conserved hypothetical protein	No match	1e-27	Bacterial
YALI0C07777g	tr|A6W0M1 *Marinomonas* sp. MWYL1 conserved hypothetical protein	No match	4e-55	Bacterial

? indicate uncertain origin

6 Miscellaneous Functions Encoded by the Genome

The metabolic pathways have been charted (http://www.genome.jp/kegg-bin/show_organism?org=yli), but no global analysis of the metabolic functions encoded by the *Y. lipolytica* genome has been reported so far. Several individual functions are the subject of different chapters of this book, and the reader is referred to these for more details. A few additional functions have been surveyed across several hemiascomycetous species including *Y. lipolytica* and will be summarised below.

6.1 Replication, Repair and Meiotic Recombination

Evolution of the pathways linked to DNA metabolism has been analysed by (Richard et al. 2005). While most of components of the mitotic replication machinery are well conserved in *Y. lipolytica*, Rfa3p, one of the three components of the single-strand binding protein complex, seems to be missing (Dna2p, involved in Okazaki fragment processing, initially considered as represented by a pseudogene, actually contains an intron). Genes involved in pyrimidine dimer or chemical damage repair are also conserved, except for RAD28 (transcription coupled repair), but only the core components of the mismatch repair pathway could be identified. Genes involved in double-strand break repair are also by and large conserved, but those involved in meiotic recombination seem to have largely diverged and some components only could be identified in *Y. lipolytica*. A single, highly conserved, copy of *RAD51* required for homologous recombination is found in *Y. lipolytica* and in *P. pastoris*, but its paralogue *DMC1* is not, although it is highly conserved from *D. hansenii* to *S. cerevisiae*. This suggests that the *RAD51* duplication occurred just before the emergence of the CTG yeasts and that, due to the absence of *DMC1* (and *HOP2* and *MND1*), pre-CTG species may not rely on DSB repair in order to achieve synapsis, contrary to *S. cerevisiae* and to most other hemiascomycetes, but behave in this respect more like *Neurospora crassa* or other eukaryotes. Absence of the Msh4-Msh5 components involved in crossover control further suggests that meiotic recombination mechanisms may have diverged significantly between pre-CTG clade yeasts like *Y. lipolytica* and other hemiascomycetes.

6.2 Sex and Mating

Y. lipolytica is a strictly heterothallic species, with two mating types (*MatA* and *MatB*) and no silent cassettes. Most natural isolates are haploids, but occasional diploids have been isolated (Bassel and Mortimer 1973). The *MATA* locus was initially cloned by complementation, for its ability to restore sporulation to a

MATB/MATB diploid (Kurischko et al. 1992). The *MAT* locus is located between two evolutionary conserved genes, *SLA2* and *APN2*, which permitted later amplification of the *MATB* idiotype (Butler et al. 2004). This organisation seems to be ancestral as it is conserved in *Neurospora crassa* and to some extent (synteny with *SLA2*) in *P. angusta* and *K. lactis* (Butler et al. 2004). Whereas *YlMATB* contains homologues of *S. cerevisiae MATα1* and *MATα2*, *YlMATA* contains a homologue of *ScMATa1* and a second gene *YlMATa2*, encoding a DNA-binding protein containing an HMG domain, unrelated to the *S. cerevisiae* pseudogene called *MATa2*. This also seems to be an ancestral feature found in Euascomycetes. No homologues of the *HMRα* and *HMRa* silent cassettes or of the *HO* inteins were found, suggesting that they were recent acquisitions in some yeast lineages (Fabre et al. 2005). A similar conclusion could be reached for the silencing complex (*SIR* genes, *RAP1*, *ABF1*,. . .) involved in shutting off the silent cassettes. It should be noted however that proteins having binding sites similar to that of Rap1 and Abf1 have been evidenced by in vivo footprinting experiments (Blanchin-Roland et al. 1994), suggesting that divergent proteins with similar functions are probably present.

A very similar organisation f the *MAT* locus was observed in the four species of the *Yarrowia* clade analysed: all strains contained a single locus (and are therefore probable haploids), two of the *MATB* type (*C. alimentaria* and *C. galli*) and two of the *MATA* type (*C. deformans* and *C. phangngaensis*). Gene models were also conserved with an intron shared by *Y. lipolytica*, *C. deformans* and *C. phangngaensis MATa1* genes.

Genes encoding the mating factor (MF) receptors for MF-α and MF-a could be identified as well as those encoding the MF-α pheromone (two genes YALI0E16533g and YALI0E30415g encoding preproMF-alpha precursors), and a candidate for the MF-a pheromone was identified on Yali0C spanning positions (2410862..2411036,2411088..2411131). An exhaustive search for genes involved in the mating signalling cascade has been conducted on several yeast genomes (Fabre et al. 2005): *Y. lipolytica* orthologues of the corresponding genes in *S. cerevisiae* could be identified in most cases, suggesting that mating regulation is conserved in all hemiascomycetes.

6.3 mRNA Metabolism

Very little is known about mRNA metabolism in hemiascomycetes in general, except of course in *S. cerevisiae* where all aspects of mRNA transcription, processing, packaging and export, translation, stability control and degradation have been dissected in great details. Many of these basic functions are highly conserved among all eukaryotes, but a number of RNA control machineries have been lost by *S. cerevisiae*. This occurred apparently either abruptly during evolution, or progressively, or independently in several lineages.

An example of independent loss was recently described in the case of RNAi control. Components of this pathway have been entirely lost by some species like

S. cerevisiae or *K. lactis*; partially lost by *P. stipitis*, *D. hansenii* or *S. bayanus*; or kept by *S. castellii*, *K. polysporus* or *C. albicans*. These last species still retain an active RNAi pathway and the two associated main components: Dicer (actually a hemiascomycete-specific paralogue of *RNT1* encoding RNaseIII) and Argonaute (Drinnenberg et al. 2009; Harrison et al. 2009). None of these components is present in *Y. lipolytica*, suggesting that the *RNT1* duplication yielding the Dicer equivalent occurred before the emergence of the CTG clade and was subsequently lost in *S. cerevisiae* or *K. lactis*, whereas Agonaute was lost independently by some lineages but not others.

An example of progressive loss resulting in regressive evolution is provided by the evolution of the exon junction complex (EJC). The EJC binds exon junction, thus resulting in a splicing imprint, and facilitates mRNA export and regulates the stability of unspliced mRNAs containing premature termination codons (PTC). In mammals, this complex acts concurrently with the NMD pathway to selectively degrade PTC-containing transcripts (Chang et al. 2007). The core of EJC (Le Hir and Andersen 2008) is made of four proteins deposited onto spliced mRNA 20–24 nt upstream of the exon–exon junctions. A pre-EJC complex, consisting of Magoh, Y14 and eIF4A3, is assembled before exon–exon junction. After exon ligation, the pre-EJC recruits the fourth protein of the core (BTZ/MLN51) and then accessory proteins, including UPF1, a component of the NMD (Gehring et al. 2009). All three components of the pre-EJC (but not MLN51) have been identified in *Y. lipolytica* and in the different species of the *Yarrowia* clade: Magoh (YALI0D26664g), Y14 (YALI0E02530g) and eIF4A3 (YALI0B20922g). All three genes have been lost from other hemiascomycetes, except from *A. adeninivorans* where a Magoh homologue can still be identified. Given that *Y. lipolytica* is the hemiascomycete with the largest number of spliceosomal introns in its genome, this finding suggests that an EJC like may still be functional and may target to the NMD pathway inefficiently spliced multi-intronic genes. NMD effectors *YlUPF1* and *YlUPF2* are indeed present in *Y. lipolytica* as in other hemiascomycetes. Recent work demonstrated that, despite the lack of a *UPF3* homologue, the NMD pathway is functional in *Y. lipolytica* (Mekouar et al. 2010). *Y. lipolytica* could thus provide a nice model to understand the exact role of EJC in NMD and to explore the role and evolution of alternative branches of NMD.

6.4 Transporters

A phylogenetic classification of hemiascomycete transporters and other membrane proteins has been proposed (see http://www.genolevures.org/yeti.html#). Transporter families have been classified as ubiquitous or species specific in hemiascomycetes (De Hertogh et al. 2006). *Y. lipolytica* (and to a lesser extent *D. hansenii*) appears unusually rich in different types of transporters, including novel ones. The anion:cation symporter subfamily involved in uptake of anionic vitamins (allantoate, nicotinate, pantothenate, biotin, ...) appears amplified in

Y. lipolytica compared to all other hemiascomycetes studied (39 members vs. 10 in *S. cerevisiae*), as well as the drug–proton antiporters, the peroxisomal protein importers, the oligopeptide transporters (17 members vs. 0–4 in other species), the ferrioxamine H1 symporters, the fatty acid transporters (14 members vs. 3–5 in other species) or the yeast metal channels (Cyt B-FRE). A major expansion of transporter families concerns two additional, novel classes of undetermined substrates comprising members unique to *Y. lipolytica*.

The distribution of transporters has been analysed for different types of families: heavy metals (Diffels et al. 2006), carboxylate (Lodi et al. 2007) and sugar transporters (Palma et al. 2007, 2009).

7 Mobile Elements

A detailed account on transposons of *Y. lipolytica* is given in Casaregola and Barth (2013). Various types of transposable elements (see Wicker et al. (2007) for classification of transposable elements) have been identified in the genome of strain E150: DNA transposons of the Mutator superfamily (Mutyl; Neuveglise et al. 2005) and of the Tc1-mariner superfamily (Fotyl; Neuvéglise, unpublished), as well as a transposase remnant of a second type of Mule element (YALI0E16731g), LTR-retrotransposons of the *gypsy* superfamily (Ylt1, Tyl3 and Tyl6) and finally LINE non-LTR-retrotransposons of the L1 superfamily (Ylli; Casaregola et al. 2002). The following complete elements were identified: 1 Fotyl (plus one relic), 5 Mutyl (plus 2 relics), 1 Tyl6, 1 relic of Ylt3, 10 Ylt1 (plus 1 relic) and 7 Ylli (plus two relics and over 100 5′-truncated elements). The distribution of the complete elements is given in Fig. 2. Solo LTRs of LTR-retroelements were also identified: 19 in the case of Ylt1 and 47 in the case of LTRyl1, an LTR which could not so far be associated to any identified transposable element (Casaregola et al. 2000; Neuveglise et al. 2002). Symmetrically, no solo LTR of the type associated with Tyl6 could be identified, suggesting that Tyl6 either is non-mobile in E150 or entered it recently, a finding that may be general for *Y. lipolytica* strains (Kovalchuk et al. 2005).

Homologues of Ylt1, Tyl6 and Ylli are present in *C. alimentaria*, *C. deformans*, *C. galli* and *C. phangngaensis*, suggesting that they were present in their last common ancestor. Fotyl is present in all species except probably in *C. phangngaensis*, whereas Mutyl could be detected only in *C. alimentaria*. In addition, a new LTR-retrotransposon of the *copia* superfamily, absent from the genome of *Y. lipolytica*, was detected in all four species, with at least one complete copy in *C. deformans* (Tyd5). These data further support the notion that some transposons have been acquired recently by yeasts of the *Y. lipolytica* clade (C. Neuvéglise, to be published).

Table 3 Mitochondrial DNA

	YALI	YAAL	YADE	YAPH
General features				
Gaps	0	2	1	2
Size (nt)	47,916	34,928	40,902	27,824
GC %	22.7	22.8	23.2	22.8
Protein encoding genes				
nd1	1	1	1	1
nd1-i1	1	0	1	0
nd2	1	1	1	1
nd3	1	1	1	1
nd4	1	1	1	1
nd4L	1	1	1	1
nd5	1	1	1	1
nd5-i1	1[a]	Ψ	1[a]	0
nd5-i2	1[a]	0	0	0
nd6	1	1	1	1
var1	1	1	1	1
cob	1	1	1	1
cob-i1	1	1	0	0
cob-i2	Ψ	1	0	0
cob-i3	Ψ	0	0	0
cob-i4	Ψ	0	0	0
cox1	1	1	1	1
cox1-i1 GIY-YIG	1	0	1	0
cox1-i2 LAGLIDADG	1	0	1	0
cox1-i3 LAGLIDADG	1	1	1	0
cox1-i4 LAGLIDADG	Ψ	1	1	0
cox1-i5 LAGLIDADG	1[a]	1	1[a]	1
cox1-i6 LAGLIDADG	1	0	1	0
cox1-i7 LAGLIDADG	1	0	Ψ[b]	0
cox1-i9 GIY-YIG	Ψ	?	1	?
atp6	1	1	1	1
atp8	1	1	1	1
atp9	1	1	1	1
cox2	1	1	1	1
cox3	1	1	1	1
cox3-i3	1	0	1	0
Intronic Orfs lost	0	9	5	14
Introns				
nd1	1	0	1	0
nd5	2	1	1	0
cox3	1	0	1	0
cox1	9	4	9	2
cob	4	2	0	0
Total introns	17	7	12	2

[a] CGC codon making the mRNA untranslatable
[b] One frameshift

8 Mitochondrial DNA

The 47.9 kb mitochondrial DNA (mitDNA, emb|AJ307410) was assembled as a follow-up of the Génolevures I programme and thus corresponds to that of W29 (Kerscher et al. 2001), one of the parent of strain E150 used to assemble the nuclear genome in Génolevures II. Later work showed that mitDNAs of both strains are quite similar, being nearly identical in the coding regions but up to 1 % divergent in the non-coding parts. Only one strand appears transcribed to yield 15 proteins, two subunits of rDNA and 27 tRNA genes. Protein encoding genes are interrupted by 17 group I introns encoding 15 endonucleases involved in mRNA maturation or intron homing. No group II intron could be identified, suggesting that intron excision is presently impossible and that only intron decay can occur. A detailed analysis of mitDNA structure and associated functions is given in Kercher and Brandt (2013).

We recently assembled the mitDNA of the three species *C. alimentaria*, *C. neoformans* and *C. phangngaensis*. The genome organisation and the protein set encoded by mitDNA in the three species are identical to those of *Y. lipolytica*. We could identify an additional gene encoding a Var1 protein as well as a supplementary Val^{UAC}tRNA, both conserved in all four species. Unexpectedly, although these species appear basal to *Y. lipolytica*, they all have smaller mitochondrial genomes and appear to have lost several group I introns, the most extreme case being represented by *C. phangngaensis* who lost all group I introns but two (see Table 3, Gaillardin et al. 2012).

9 Conclusion

The *Y. lipolytica* is presently the largest Hemiascomycete genome and has been probably one of the most difficult to annotate due to its large size, the relative abundance of introns and the strong sequence divergence of its coding regions. It is nevertheless one of the best annotated genomes among hemiascomycetes except *S. cerevisiae*, even if annotation is clearly far from being optimal and if many mistakes or erroneous annotations still remain. Our hope was to provide a reference for yeast genomes distantly related to *S. cerevisiae*, in order to get a better perspective of the diversity of yeast genome organisation and of their evolution. One of the major lessons that we derived from its analysis was that many textbook references that had been derived from *S. cerevisiae* as the yeast “par excellence” should actually be revisited for some yeasts at least and certainly for *Y. lipolytica*. Having its genome is thus just the beginning of a new story.

Acknowledgements We wish to thank all our colleagues from the Génolevures consortium and at the Génoscope, too many to be listed individually, who participated in the *Yarrowia* genome project; most of the results presented here are the result of a collective effort and are largely the fruit of their work. We are also deeply indebted to all our colleagues from the *Yarrowia*

community who participated in the launching of the *Yarrowia* genome project, by financing the first steps on their own lab money. Their input for improving the genome annotation, identifying new genes that had escaped our attention, has been invaluable. And finally, we wish to apologise to the many colleagues whose work has not been cited here for lack of space. Author's work was supported by the GDR CNRS 2354 "Génolevures-3", INRA and the ANR "Genarise" (ANR-05-BLAN-0331).

References

Barth G, Gaillardin C (1996) Yarrowia lipolytica. In: Wolf K (ed) Non-conventional yeasts in biotechnology. A handbook. Springer, Heidelberg, pp 313–388

Bassel J, Mortimer R (1973) Genetic analysis of mating type and alkane utilization in Saccharomycopsis lipolytica. J Bacteriol 114:894–896

Blanchin-Roland S, Cordero Otero RR, Gaillardin C (1994) Two upstream activation sequences control the expression of the XPR2 gene in the yeast Yarrowia lipolytica. Mol Cell Biol 14:327–338

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

Butler G, Rasmussen MD, Lin MF, Santos MA, Sakthikumar S, Munro CA, Rheinbay E, Grabherr M, Forche A, Reedy JL, Agrafioti I, Arnaud MB, Bates S, Brown AJ, Brunke S, Costanzo MC, Fitzpatrick DA, de Groot PW, Harris D, Hoyer LL, Hube B, Klis FM, Kodira C, Lennard N, Logue ME, Martin R, Neiman AM, Nikolaou E, Quail MA, Quinn J, Santos MC, Schmitzberger FF, Sherlock G, Shah P, Silverstein KA, Skrzypek MS, Soll D, Staggs R, Stansfield I, Stumpf MP, Sudbery PE, Srikantha T, Zeng Q, Berman J, Berriman M, Heitman J, Gow NA, Lorenz MC, Birren BW, Kellis M, Cuomo CA (2009) Evolution of pathogenicity and sexual reproduction in eight Candida genomes. Nature 459:657–662

Casaregola S, Barth G (2013) Transponable elements and their activities in *Yarrowia lipolytica*. In: Barth G (ed) Yarrowia lipolytica. Springer, Heidelberg

Casaregola S, Feynerol C, Diez M, Fournier P, Gaillardin C (1997) Genomic organization of the yeast Yarrowia lipolytica. Chromosoma 106:380–390

Casaregola S, Neuveglise C, Lepingle A, Bon E, Feynerol C, Artiguenave F, Wincker P, Gaillardin C (2000) Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia lipolytica. FEBS Lett 487:95–100

Casaregola S, Neuveglise C, Bon E, Gaillardin C (2002) Ylli, a non-LTR retrotransposon L1 family in the dimorphic yeast Yarrowia lipolytica. Mol Biol Evol 19:664–677

Chang YF, Imam JS, Wilkinson MF (2007) The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem 76:51–74

Clare JJ, Davidow LS, Gardner DC, Oliver SG (1986) Cloning and characterisation of the ribosomal RNA genes of the dimorphic yeast, Yarrowia lipolytica. Curr Genet 10:449–452

Cvetic C, Walter JC (2005) Eukaryotic origins of DNA replication: could you please be more specific? Semin Cell Dev Biol 16:343–353

De Hertogh B, Hancy F, Goffeau A, Baret PV (2006) Emergence of species-specific transporters during evolution of the hemiascomycete phylum. Genetics 172:771–781

De Schutter K, Lin YC, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, Rouze P, Van de Peer Y, Callewaert N (2009) Genome sequence of the recombinant protein production host Pichia pastoris. Nat Biotechnol 27:561–566

Desfougeres T, Haddouche R, Fudalej F, Neuveglise C, Nicaud JM (2008) SOA genes encode proteins controlling lipase expression in response to triacylglycerol utilization in the yeast Yarrowia lipolytica. FEMS Yeast Res 10:93–103

Diffels JF, Seret ML, Goffeau A, Baret PV (2006) Heavy metal transporters in Hemiascomycete yeasts. Biochimie 88:1639–1649

Drinnenberg IA, Weinberg DE, Xie KT, Mower JP, Wolfe KH, Fink GR, Bartel DP (2009) RNAi in budding yeast. Science 326:544–550

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

Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuveglise C, Talla E, Goffard N, Frangeul L, Aigle M, Anthouard V, Babour A, Barbe V, Barnay S, Blanchin S, Beckerich JM, Beyne E, Bleykasten C, Boisrame A, Boyer J, Cattolico L, Confanioleri F, De Daruvar A, Despons L, Fabre E, Fairhead C, Ferry-Dumazet H, Groppi A, Hantraye F, Hennequin C, Jauniaux N, Joyet P, Kachouri R, Kerrest A, Koszul R, Lemaire M, Lesur I, Ma L, Muller H, Nicaud JM, Nikolski M, Oztas S, Ozier-Kalogeropoulos O, Pellenz S, Potier S, Richard GF, Straub ML, Suleau A, Swennen D, Tekaia F, Wesolowski-Louvel M, Westhof E, Wirth B, Zeniou-Meyer M, Zivanovic I, Bolotin-Fukuhara M, Thierry A, Bouchier C, Caudron B, Scarpelli C, Gaillardin C, Weissenbach J, Wincker P, Souciet JL (2004) Genome evolution in yeasts. Nature 430:35–44

Durrens P, Sherman DJ (2005) A systematic nomenclature of chromosomal elements for hemiascomycete yeasts. Yeast 22:337–342

Eitzen GA, Aitchison JD, Szilard RK, Veenhuis M, Nuttley WM, Rachubinski RA (1995) The Yarrowia lipolytica gene PAY2 encodes a 42-kDa peroxisomal integral membrane protein essential for matrix protein import and peroxisome enlargement but not for peroxisome membrane proliferation. J Biol Chem 270:1429–1436

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

Fitzpatrick DA, Logue ME, Stajich JE, Butler G (2006) A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. BMC Evol Biol 6:99

Fournier P, Gaillardin C, Persuy MA, Klootwijk J, van Heerikhuizen H (1986) Heterogeneity in the ribosomal family of the yeast Yarrowia lipolytica: genomic organization and segregation studies. Gene 42:273–282

Fournier P, Abbas A, Chasles M, Kudla B, Ogrydziak DM, Yaver D, Xuan JW, Peito A, Ribet AM, Feynerol C et al (1993) Colocalization of centromeric and replicative functions on autonomously replicating sequences isolated from the yeast Yarrowia lipolytica. Proc Natl Acad Sci USA 90:4912–4916

Gaillardin C, Neuveglise C, Kerscher S, Nicaud JM (2012) Mitochondrial genomes of yeasts of the Yarrowia clade. FEMS Yeast Res 12:317–331

Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S, Lee SI, Basturkmen M, Spevak CC, Clutterbuck J, Kapitonov V, Jurka J, Scazzocchio C, Farman M, Butler J, Purcell S, Harris S, Braus GH, Draht O, Busch S, D'Enfert C, Bouchier C, Goldman GH, Bell-Pedersen D, Griffiths-Jones S, Doonan JH, Yu J, Vienken K, Pain A, Freitag M, Selker EU, Archer DB, Penalva MA, Oakley BR, Momany M, Tanaka T, Kumagai T, Asai K, Machida M, Nierman WC, Denning DW, Caddick M, Hynes M, Paoletti M, Fischer R, Miller B, Dyer P, Sachs MS, Osmani SA, Birren BW (2005) Sequencing of Aspergillus nidulans and comparative analysis with A. fumigatus and A. oryzae. Nature 438:1105–1115

Gehring NH, Lamprinaki S, Hentze MW, Kulozik AE (2009) The hierarchy of exon-junction complex assembly by the spliceosome explains key features of mammalian nonsense-mediated mRNA decay. PLoS Biol 7:e1000120

Harrison BR, Yazgan O, Krebs JE (2009) Life without RNAi: noncoding RNAs and their functions in Saccharomyces cerevisiae. Biochem Cell Biol 87:767–779

He F, Yaver D, Beckerich JM, Ogrydziak D, Gaillardin C (1990) The yeast Yarrowia lipolytica has two, functional, signal recognition particle 7S RNA genes. Curr Genet 17:289–292

Hegemann JH, Fleig UN (1993) The centromere of budding yeast. Bioessays 15:451–460

Heus JJ, Zonneveld BJ, Steensma HY, Van den Berg JA (1994) Mutational analysis of centromeric DNA elements of Kluyveromyces lactis and their role in determining the species specificity of the highly homologous centromeres from K. lactis and Saccharomyces cerevisiae. Mol Gen Genet 243:325–333

Hirakawa K, Kobayashi S, Inoue T, Endoh-Yamagami S, Fukuda R, Ohta A (2009) Yas3p, an Opi1 family transcription factor, regulates cytochrome P450 expression in response to n-alkanes in Yarrowia lipolytica. J Biol Chem 284:7126–7137

Kachouri R, Stribinskis V, Zhu Y, Ramos KS, Westhof E, Li Y (2005) A surprisingly large RNase P RNA in Candida glabrata. RNA 11:1064–1072

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

Kercher S, Brandt U (2013) Mitochondrial genomics and proteomics of *Yarrowia lipolytica*. In: Barth G (ed) Yarrowia lipolytica. Springer, Heidelberg

Kerscher S, Durstewitz G, Casaregola S, Gaillardin C, Brandt U (2001) The complete mitochondrial genome of Yarrowia lipolytica. Comp Funct Genomics 2:80–90

Ketel C, Wang HS, McClellan M, Bouchonville K, Selmecki A, Lahav T, Gerami-Nejad M, Berman J (2009) Neocentromeres form efficiently at multiple possible loci in Candida albicans. PLoS Genet 5:e1000400

Kinsky S, Mihalikova A, Kramara J, Nosek J, Tomaska L (2010) Lack of the catalytic subunit of telomerase leads to growth defects accompanied by structural changes at the chromosomal ends in *Yarrowia lipolytica*. Curr Genet 56:413–425

Knutsen AK, Robert V, Poot GA, Epping W, Figge M, Holst-Jensen A, Skaar I, Smith MT (2007) Polyphasic re-examination of Yarrowia lipolytica strains and the description of three novel Candida species: Candida oslonensis sp. nov., Candida alimentaria sp. nov. and Candida hollandica sp. nov. Int J Syst Evol Microbiol 57:2426–2435

Kovalchuk A, Senam S, Mauersberger S, Barth G (2005) Tyl6, a novel Ty3/gypsy-like retrotransposon in the genome of the dimorphic fungus Yarrowia lipolytica. Yeast 22:979–991

Kramara J, Willcox S, Gunisova S, Kinsky S, Nosek J, Griffith JD, Tomaska L (2010) Tay1 protein, a novel telomere binding factor from *Yarrowia lipolytica*. J Biol Chem 285:38078–38092

Kurischko C, Fournier P, Chasles M, Weber H, Gaillardin C (1992) Cloning of the mating-type gene MATA of the yeast Yarrowia lipolytica. Mol Gen Genet 232:423–426

Kurtzman CP (2005) New species and a new combination in the Hyphopichia and Yarrowia yeast clades. Antonie Van Leeuwenhoek 88:121–130

Le Dall MT, Nicaud JM, Gaillardin C (1994) Multiple-copy integration in the yeast Yarrowia lipolytica. Curr Genet 26:38–44

Le Hir H, Andersen GR (2008) Structural insights into the exon junction complex. Curr Opin Struct Biol 18:112–119

Limtong S, Youngmanitchai W, Kawasaki H, Seki T (2008) Candida phangngensis sp. nov., an anamorphic yeast species in the Yarrowia clade, isolated from water in mangrove forests in Phang-Nga Province, Thailand. Int J Syst Evol Microbiol 58:515–519

Linskens MH, Huberman JA (1988) Organization of replication of ribosomal DNA in Saccharomyces cerevisiae. Mol Cell Biol 8:4927–4935

Lodi T, Diffels J, Goffeau A, Baret PV (2007) Evolution of the carboxylate Jen transporters in fungi. FEMS Yeast Res 7:646–656

Matsuoka M, Matsubara M, Daidoh H, Imanaka T, Uchida K, Aiba S (1993) Analysis of regions essential for the function of chromosomal replicator sequences from Yarrowia lipolytica. Mol Gen Genet 237:327–333

Meraldi P, McAinsh AD, Rheinbay E, Sorger PK (2006) Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins. Genome Biol 7:R23

Mekouar M, Blanc-Lenfle I, Ozanne C, Da Silva C, Cruaud C, Wincker P, Gaillardin C, Neuveglise C (2010) Detection and analysis of alternative splicing in Yarrowia lipolytica reveal structural constraints facilitating nonsense-mediated decay of intron-retaining transscripts. Genome Biol 11:R65

Mesner LD, Crawford EL, Hamlin JL (2006) Isolating apparently pure libraries of replication origins from complex genomes. Mol Cell 21:719–726

Mishra PK, Baum M, Carbon J (2007) Centromere size and position in Candida albicans are evolutionarily conserved independent of DNA sequence heterogeneity. Mol Genet Genomics 278:455–465

Morgner N, Zickermann V, Kerscher S, Wittig I, Abdrakhmanova A, Barth HD, Brutschy B, Brandt U (2008) Subunit mass fingerprinting of mitochondrial complex I. Biochim Biophys Acta 1777:1384–1391

Naumova E, Naumov G, Fournier P, Nguyen HV, Gaillardin C (1993) Chromosomal polymorphism of the yeast Yarrowia lipolytica and related species: electrophoretic karyotyping and hybridization with cloned genes. Curr Genet 23:450–454

Neuveglise C, Feldmann H, Bon E, Gaillardin C, Casaregola S (2002) Genomic evolution of the long terminal repeat retrotransposons in hemiascomycetous yeasts. Genome Res 12:930–943

Neuveglise C, Chalvet F, Wincker P, Gaillardin C, Casaregola S (2005) Mutator-like element in the yeast Yarrowia lipolytica displays multiple alternative splicings. Eukaryot Cell 4:615–624

Neuvéglise C, Gaillardin C, Marck C (2013) Non-coding RNA genes transcribed by RNA polymerase III in *Yarrowia lipolytica*. In: Barth G (ed) Yarrowia lipolytica. Springer, Heidelberg

Newlon CS, Collins I, Dershowitz A, Deshpande AM, Greenfeder SA, Ong LY, Theis JF (1993) Analysis of replication origin function on chromosome III of Saccharomyces cerevisiae. Cold Spring Harb Symp Quant Biol 58:415–423

Palma M, Goffeau A, Spencer-Martins I, Baret PV (2007) A phylogenetic analysis of the sugar porters in hemiascomycetous yeasts. J Mol Microbiol Biotechnol 12:241–248

Palma M, Seret ML, Baret PV (2009) Combined phylogenetic and neighbourhood analysis of the hexose transporters and glucose sensors in yeasts. FEMS Yeast Res 9:526–534

Poritz MA, Siegel V, Hansen W, Walter P (1988) Small ribonucleoproteins in Schizosaccharomyces pombe and Yarrowia lipolytica homologous to signal recognition particle. Proc Natl Acad Sci USA 85:4315–4319

Richard GF, Kerrest A, Lafontaine I, Dujon B (2005) Comparative genomics of hemiascomycete yeasts: genes involved in DNA replication, repair, and recombination. Mol Biol Evol 22:1011–1023

Rolland T, Neuveglise C, Sacerdot C, Dujon B (2009) Insertion of horizontally transferred genes within conserved syntenic regions of yeast genomes. PLoS One 4:e6515

Rossignol T, Lechat P, Cuomo C, Zeng Q, Moszer I, D'Enfert C (2008) CandidaDB: a multi-genome database for Candida species and related Saccharomycotina. Nucleic Acids Res 36: D557–D561

Sanchez JA, Kim SM, Huberman JA (1998) Ribosomal DNA replication in the fission yeast, Schizosaccharomyces pombe. Exp Cell Res 238:220–230

Sanyal K, Baum M, Carbon J (2004) Centromeric DNA sequences in the pathogenic yeast Candida albicans are all different and unique. Proc Natl Acad Sci USA 101:11374–11379

Sherman D, Durrens P, Beyne E, Nikolski M, Souciet JL (2004) Genolevures: comparative genomics and molecular evolution of hemiascomycetous yeasts. Nucleic Acids Res 32: D315–D318

Sherman DJ, Martin T, Nikolski M, Cayla C, Souciet JL, Durrens P (2009) Genolevures: protein families and synteny among complete hemiascomycetous yeast proteomes and genomes. Nucleic Acids Res 37:D550–D554

Souciet J, Aigle M, Artiguenave F, Blandin G, Bolotin-Fukuhara M, Bon E, Brottier P, Casaregola S, de Montigny J, Dujon B, Durrens P, Gaillardin C, Lepingle A, Llorente B, Malpertuy A, Neuveglise C, Ozier-Kalogeropoulos O, Potier S, Saurin W, Tekaia F, Toffano-Nioche C, Wesolowski-Louvel M, Wincker P, Weissenbach J (2000) Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies. FEBS Lett 487:3–12

Souciet JL, Dujon B, Gaillardin C, Johnston M, Baret PV, Cliften P, Sherman DJ, Weissenbach J, Westhof E, Wincker P, Jubin C, Poulain J, Barbe V, Segurens B, Artiguenave F, Anthouard V, Vacherie B, Val ME, Fulton RS, Minx P, Wilson R, Durrens P, Jean G, Marck C, Martin T, Nikolski M, Rolland T, Seret ML, Casaregola S, Despons L, Fairhead C, Fischer G, Lafontaine I, Leh V, Lemaire M, de Montigny J, Neuveglise C, Thierry A, Blanc-Lenfle I, Bleykasten C, Diffels J, Fritsch E, Frangeul L, Goeffon A, Jauniaux N, Kachouri-Lafond R, Payen C, Potier S, Pribylova L, Ozanne C, Richard GF, Sacerdot C, Straub ML, Talla E (2009) Comparative genomics of protoploid Saccharomycetaceae. Genome Res 19:1696–1709

Steigele S, Huber W, Stocsits C, Stadler PF, Nieselt K (2007) Comparative analysis of structured RNAs in S. cerevisiae indicates a multitude of different functions. BMC Biol 5:25

van Heerikhuizen H, Ykema A, Klootwijk J, Gaillardin C, Ballas C, Fournier P (1985) Heterogeneity in the ribosomal RNA genes of the yeast Yarrowia lipolytica; cloning and analysis of two size classes of repeats. Gene 39:213–222

Vernis L, Abbas A, Chasles M, Gaillardin CM, Brun C, Huberman JA, Fournier P (1997) An origin of replication and a centromere are both needed to establish a replicative plasmid in the yeast Yarrowia lipolytica. Mol Cell Biol 17:1995–2004

Vernis L, Chasles M, Pasero P, Lepingle A, Gaillardin C, Fournier P (1999) Short DNA fragments without sequence similarity are initiation sites for replication in the chromosome of the yeast Yarrowia lipolytica. Mol Biol Cell 10:757–769

Vernis L, Poljak L, Chasles M, Uchida K, Casaregola S, Kas E, Matsuoka M, Gaillardin C, Fournier P (2001) Only centromeres can supply the partition system required for ARS function in the yeast Yarrowia lipolytica. J Mol Biol 305:203–217

Wang H, Xu Z, Gao L, Hao B (2009) A fungal phylogeny based on 82 complete genomes using the composition vector method. BMC Evol Biol 9:195

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982

Xu W, Aparicio JG, Aparicio OM, Tavare S (2006) Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in S. cerevisiae. BMC Genomics 7:276

Yamane T, Ogawa T, Matsuoka M (2008a) Derivation of consensus sequence for protein binding site in Yarrowia lipolytica centromere. J Biosci Bioeng 105:671–674

Yamane T, Sakai H, Nagahama K, Ogawa T, Matsuoka M (2008b) Dissection of centromeric DNA from yeast Yarrowia lipolytica and identification of protein-binding site required for plasmid transmission. J Biosci Bioeng 105:571–578

Mitochondrial Genomics and Proteomics of *Yarrowia lipolytica*

Stefan Kerscher and Ulrich Brandt

Abstract In recent years, we have developed the obligate aerobic yeast *Y. lipolytica* as a model system to analyse respiratory chain complex I (proton pumping NADH:ubiquinone oxidoreductase). Our studies were aimed at exploring structural and functional constraints for the reaction mechanism at the catalytic core of this giant enzyme, defined as the site where ubiquinone reduction couples to proton translocation. Extensive genomic and proteomic analyses of *Y. lipolytica* mitochondria resulted in the discovery of 41 different complex I subunits, encoded both by the mitochondrial and the nuclear genome. Most of the subunits of complex V (ATP synthase) and a protein specifically involved in complex I iron–sulfur cluster assembly were also described. Novel gel-electrophoretic separation techniques and analytical methods, especially laser-induced liquid bead ion desorption (LILBID) mass spectrometry, developed for these purposes, have great potential as useful tools for obtaining molecular fingerprints of large protein assemblies.

S. Kerscher
Center for Membrane Proteomics, Cluster of Excellence Frankfurt "Macromolecular Complexes", Molekulare Bioenergetik, Fachbereich Medizin, Goethe-Universität, 60590 Frankfurt am Main, Germany

U. Brandt (✉)
ZBC, Cluster of Excellence Frankfurt "Macromolecular Complexes", Molekulare Bioenergetik, Fachbereich Medizin, Goethe-Universität, Frankfurt am Main, Germany

Present address: Nijmegen Centre for Mitochondrial Disorders (NCMD), Radboud University Medical Centre Geert Groteplein-Zuid 10, 6525 GA Nijmegen, The Netherlands
e-mail: u.brandt@cukz.umcn.nl

G. Barth (ed.), *Yarrowia lipolytica*, Microbiology Monographs 24,
DOI 10.1007/978-3-642-38320-5_2, © Springer-Verlag Berlin Heidelberg 2013

Contents

1 Introduction 32
2 The Mitochondrial Respiratory Chain of *Y. lipolytica* 33
3 Characteristic Features of Complex I 35
4 A Genetic Approach to *Y. lipolytica* Complex I 41
5 The Mitochondrial Genome of *Y. lipolytica* 42
6 A Proteomic Approach to *Y. lipolytica* Complex I 43
7 Subunits Specific to *Y. lipolytica* Complex I 46
8 Exploring Human Complex I Defects 48
9 Exploring the Reaction Mechanism of Complex I 49
References 51

1 Introduction

In recent years, the obligate aerobic yeast *Yarrowia lipolytica* has been firmly established as a powerful yeast genetic model system for the analysis of respiratory chain complex I. Due to the intricate composition of this multi-subunit enzyme, this work was accompanied by large-scale genomic and proteomic analyses of *Y. lipolytica* mitochondria. In this chapter, we are going to review these efforts, with a special emphasis on novel analytical methods that were developed in the course of these studies.

Many questions on complex I remain unanswered. Only recently the complex was studied by X-ray crystallography (Hunte et al. 2010; Efremov et al. 2010; Baradaran et al. 2013), and the reaction mechanism is far from being understood. This lack of information is untenable in view of the vital position of complex I in human metabolism, as illustrated by the fact that inherited and acquired complex I defects cause severe, progressive and often fatal disorders, with Leigh syndrome as the most typical clinical manifestation (see below).

While *Saccharomyces cerevisiae* has proven an excellent model system for the genetic analysis of numerous topics in cell biology, the enzymatic machinery of this species, as a consequence of its adaptation to the fermentative lifestyle (Lagunas 1986), displays several peculiar features, including the lack of respiratory chain complex I. The resulting inapplicability of the *S. cerevisiae* model system certainly is among the reasons why the analysis of complex I is still lagging behind the work on other complexes of the respiratory chain.

We therefore set out to establish a yeast model system for the analysis of complex I. The obligate aerobic yeast *Y. lipolytica* was chosen since its genetics are quite well advanced (Barth and Gaillardin 1996), as it combines heterothallism and two stable mating types with the availability of replicative and integrative plasmids and several easily scorable metabolic and resistance markers. The genomic sequence of *Y. lipolytica* has been analysed by the Genolevures consortium (Dujon et al. 2004).

The most widely accepted reference system for eukaryotic complex I, studied extensively on the proteomic level, is complex I from bovine heart. Evolutionary

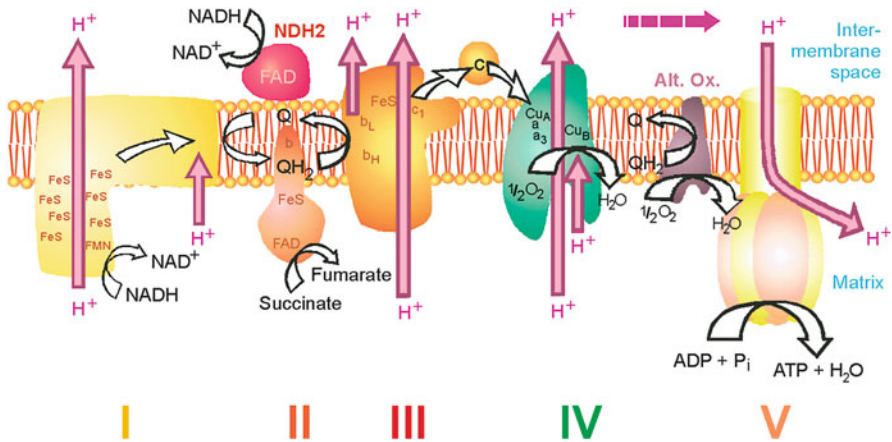


Fig. 1 Respiratory chain of *Y. lipolytica*. Schematic representation of the respiratory chain of *Y. lipolytica*. Substrate flow is indicated by *white* and proton pumping by *pink arrows*. I, II, III, IV, V, respiratory chain complexes I–V; a, a_3, heme *a* and a_3; Alt. Ox., alternative oxidase; b, heme *b*; b_L, b_H, low and high potential heme *b*; c, c_1, heme *c* and c_1; Cu_A, Cu_B, copper centre A and B; FeS, iron–sulfur cluster; Q, ubiquinone; QH_2, reduced ubiquinone; NDH2, alternative NADH dehydrogenase

conserved subunits of *Y. lipolytica* complex I were thus named after their bovine heart homologues. In addition, a few yeast-/fungus-specific subunits were found.

2 The Mitochondrial Respiratory Chain of *Y. lipolytica*

The respiratory chain of *Y. lipolytica* mitochondria does possess not only complex I but also all other respiratory chain complexes found in mammals including humans (Fig. 1), namely, complex II (succinate:ubiquinone oxidoreductase), complex III (ubiquinol:cytochrome *c* oxidoreductase), complex IV (cytochrome *c* oxidase) and complex V (ATP synthase). The ubiquinone species employed by *Y. lipolytica*, termed Q_9, has a nine-unit isoprenoid side chain (Kurtzman 1994). This is closer to $Q_{9–11}$ found in mammals than to Q_6 found in *S. cerevisiae*. Also, the phospholipid composition of mitochondrial membrane fractions (Arthur and Watson 1976) was found to be similar to that of mammalian mitochondria.

As in human mitochondria, complexes I, III and IV, by pumping protons into the intermembrane space, build up a transmembrane proton gradient which is then used to drive ATP synthesis by complex V. Besides these complexes, the respiratory chain of *Y. lipolytica* mitochondria also comprises so-called alternative enzymes, as typically found in plants and fungi. Alternative NADH dehydrogenase (NDH2) carries out the same redox reaction as complex I (Kerscher 2000; Melo et al. 2004), while alternative terminal oxidase (AOX) bypasses complexes III and IV (Joseph-Horne et al. 2001). Both enzymes consist of a single subunit only and do not pump protons. NDH2 contains a non-covalently bound FAD molecule as sole redox

prosthetic group. Alternative NADH dehydrogenases have been found in archae- and eubacteria, in protists, plants and fungi. As these enzymes can reside on either the inner or the outer face of the mitochondrial inner membrane, eukaryotic alternative NADH dehydrogenases can be classified as internal or external. For example, *S. cerevisiae* mitochondria possess two external and one internal alternative enzyme (Marres et al. 1991; Luttik et al. 1998). Only the internal form of NDH2 confers resistance to complex I inhibitors like DQA and rotenone (Kerscher et al. 2001b).

AOX which occurs in plants and fungi contains a di-iron active centre and confers resistance to inhibitors of complexes III and IV. AOX activity is detectable in *Y. lipolytica* cells growing in the presence of antimycin A or cyanide. In *Y. lipolytica* cultures growing in the absence of such poisons, AOX is expressed in stationary phase cells but in an inactive form that only becomes activated in conditions that elevate cellular AMP levels (Medentsev and Akimenko 1999).

Studying the properties of the alternative NADH dehydrogenase of *Y. lipolytica* proved to be of vital importance for establishing this yeast as a model system for complex I research (Kerscher et al. 1999). When we first tried to sporulate diploid strains heterozygous for a deletion allele for one of the seven nuclear encoded central subunits of complex I, no viable haploid strains carrying the deletion allele could be identified among the progeny. We concluded that, in contrast to other fungi like *Neurospora crassa* and *Aspergillus spec*. where complex I null mutants had been generated before, complex I function was essential for survival of standard laboratory strains of *Y. lipolytica*. At that time, it appeared that genetic analysis of complex I using *Y. lipolytica* was impossible.

However, a useful hint to a solution for this problem was found when the alternative NADH dehydrogenase activity of *Y. lipolytica* was analysed in more detail. A comparison of intact and permeabilised *Y. lipolytica* mitochondria clearly demonstrated that alternative NADH dehydrogenase activity in this yeast was exclusively due to an external enzyme. Following deletion of the NDH2 gene, alternative NADH dehydrogenase activity was completely abolished (Kerscher et al. 1999).

Next, we created a gene fusion in which a truncated NDH2 open reading frame was fused to the sequence encoding the mitochondrial targeting sequence of the 75-kDa subunit of *Y. lipolytica* complex I. The construct was randomly integrated into the *Y. lipolytica* genome, and transgenic colonies could be isolated by virtue of their ability to grow in the presence of the complex I inhibitor DQA. Thus, the internal version of *Y. lipolytica* NDH2, termed NDH2i, was able to substitute for the NADH dehydrogenase function of complex I (Kerscher et al. 2001b). Its presence permitted the deletion of nuclear genes for essential subunits of complex I, which could then be complemented with plasmid-borne site-directed mutant copies of the respective genes.

It should be noted that a further alternative NADH dehydrogenase gene (YALI0E05599g) was detected in *Y. lipolytica* during the course of the genome project. Its functional significance remains unclear since a catalytic activity corresponding to the enzyme encoded by this gene has not been reported yet.

Table 1 Assignment of EPR signals to the iron–sulfur clusters of complex I

Cluster type	Subunit name (*Bos taurus*)	EPR signal assignment	
		Traditional	Revised
Fe_2S_2	24 kDa	N1a	N1a
Fe_4S_4	51 kDa	N3	N3
Fe_2S_2	75 kDa	N1b	N1b
Fe_4S_4	75 kDa	N4	N5
Fe_4S_4	75 kDa	N5	Not detectable
Fe_4S_4	TYKY	N6a	N4
Fe_4S_4	TYKY	N6b	Not detectable
Fe_4S_4	PSST	N2	N2

Adapted from Yakovlev et al. (2007) and Roessler et al. (2010)

3 Characteristic Features of Complex I

Among the reasons why mitochondria as endosymbionts are of prime value for their hosts, their superior capability of ATP production by utilisation of oxygen holds one of the top ranks. Complex I (NADH:ubiquinone oxidoreductase) in mammals makes a major contribution (up to 40 %, depending on metabolic conditions) to the proton gradient across the inner mitochondrial membrane and hence to oxidative phosphorylation (Brandt 2006). The transfer of two electrons from NADH to ubiquinone is coupled to the translocation of four protons (Galkin et al. 2001, 2006). Complex I is a huge multiprotein assembly, composed of 45 subunits in mammals (Carroll et al. 2006) and around 40 subunits in fungi (Morgner et al. 2008), with a total molecular mass in the range of one million Dalton. The enzyme is L shaped with two arms of similar length and girth, oriented perpendicular to each other, a hydrophobic arm imbedded in the respiratory membrane and a hydrophilic peripheral arm protruding into the mitochondrial matrix (Radermacher et al. 2006). The peripheral arm carries all redox prosthetic groups, namely, one FMN and eight iron–sulfur clusters, two of which are of the binuclear and six of the tetranuclear type (Table 1; Sazanov and Hinchliffe 2006); the membrane arm harbours the proton pump.

Many bacteria, including *Escherichia coli*, possess a "minimal form" of complex I that is composed of 14 subunits only, with a total molecular mass of about 550 kDa (Friedrich 1998; Yano and Yagi 1999). Seven subunits each make up the peripheral and the membrane arm, respectively. All of them have homologues in eukaryotic complex I, termed "central subunits," since obviously they are sufficient to carry out the bioenergetic function of complex I. In mammals and fungi, the seven central subunits of the peripheral arm are nuclear coded, while the seven central subunits of the membrane arm are mitochondrially coded (Table 2). The function of the "accessory subunits" of eukaryotic complex I, all of which are nuclear coded, is largely unknown (see below).

While the subunit composition of bacterial complex I is much simpler and its molecular mass is markedly lower than that of the eukaryotic enzymes, using the

Table 2 Subunits of *Y. lipolytica* mitochondrial complex I

Subunit					M_r[c]		Prosth. groups		LILBID	
No	Gene	Protein	TMDs[a]	N-terminal sequence[b]	Precursor	Mature	Type	M_r	m/z	Remarks
Central nuclear coded										
1	NUAM	75-kDa	0/0	AEIELT[f]	78,997.2	75,198.7	$Fe_{10}S_{10}$	879.2	75.2	
2	NUBM	51-kDa	0/0	ATTQDA[f]	53,755.6	51,660.0	Fe_4S_4	351.7	51.7	R1
							FMN	456.4		
3	NUCM	49-kDa	0/0	ATTALP[f]	52,430.0	49,945.0	–	–	49.9	R3
4	NUGM[d]	30-kDa	0/0	QAAPSS[f]	33,594.9	30,476.2	–	–	30.5	
5	NUHM	24-kDa	0/0	IVSVHR[g]	27,216.4	24,068.6	Fe_2S_2	175.8	24.3	
6	NUIM	TYKY	1/0	APATDS[g]	25,651.3	22,321.3	Fe_8S_8	703.4	22.3	
7	NUKM	PSST	0/0	SAPAGT[f]	23,430.2	20,425.6	Fe_4S_4	351.7	20.2	R2,3
Sum						274,095.4		2,918.2		
Central mitochondrially coded										
8	NU1M	ND1	10/10			38,347.7			38.3	
9	NU2M	ND2	14/13			53,331.9			53.2	
10	NU3M	ND3	3/3			14,470.5			14.5	
11	NU4M	ND4	13/12			54,481.2			54.4	
12	NU5M	ND5	18/16			73,705.4			73.5	
13	NU6M	ND6	5/5			20,758.3			20.7	
14	NULM	ND4L	2/2			9,810.9			9.9	
Sum						264,905.9				
Accessory nuclear coded										
15	NUEM	0/0	MNSFEN[g]	42,705.8	40,434.0	NADPH	745.4	40.4	R4	
16	ST1	0/0			34,621.2			34.5	R5	
17	NESM	1/0	FALRAY[g]	28,435.9	23,438.1			23.4		
18	NUJM	3/3			20,827.6			20.7	R6	
19	NUZM	0/0	MLPGGP[g]		19,749.6			19.7		
20	NUPM	0/0		19,327.2	19,196.0			19.2		

21	NUXM	3/2			18,565.2			18.5	
22	N7BM	0/0			16,153.2			16.1	
23	NUYM	0/0	QKKDVP[g]	18,671.1	15,939.9			15.9	
24	NUFM	0/0	NVSKGV[g]	16,635.0	15,572.7			15.6	
25	NIAM	1/1	IRASFD[g]	17,306.6	14,642.3			14.7	
26	NB4M	0/0	AIIATA[g]	14,757.8	14,626.7			14.6	
27	NUMM	0/0			14,257.0			14.4	
28	NB6M	1/1		14,091.4	13,960.2			13.9	R7
29	NUNM	1/1		15,794.8	13,300.9			13.4	
30	NI2M	0/0			12,879.8			12.8	
31	NB8M	0/0	AEFPPL[g]	1,119.0	11,067.8			11.1	R8
32	NIDM	0/0		11,021.3	10,890.1			10.9	
33	NB5M	1/1	VELKPS[g]	10,478.9	10,347.7			10.3	
34	ACPM2	0/0		14,428.5	9,526.7	PP-HD[e]	564.7	10.1	
35	NIPM	0/0			10,018.6			10.1	
36	NIMM	1/1		9,793.3	9,662.1			9.6	
37	ACPM1	0/0		12,040.6	9,071.2	PP-HD[e]	564.7	9.6	
38	NI8M	0/0		9,604.0	9,472.8			9.5	
39	NEBM	1/1		7,836.0	7,917.2			7.9	
40	NI9M	1/1		7,836.0	7,704.8			7.7	
41	NB2M	1/1		6,935.9	6,804.7			6.8	
Sum					410,648.1		1,874.8		R9

Remarks:
R1: NADH oxidation site
R2: Ubiquinone reduction site
R3: Catalytic core of complex I (coupling ubiquinone reduction to proton translocation)
R4: NADPH serves a structural role only (Abdrakhmanova et al. 2006)
R5: Confers in vitro thiosulfate:cyanide sulfurtransferase activity (Abdrakhmanova et al. 2005)
R6: Related to Tim17, Tim22 and Tim23 proteins involved in protein translocation across the inner mitochondrial membrane (Carroll et al. 2002)

(continued)

Table 2 (continued)

Subunit					M_r[c]		Prosth. groups		LILBID	
No	Gene	Protein	TMDs[a]	N-terminal sequence[b]	Precursor	Mature	Type	M_r	m/z	Remarks

R7: The mammalian homologue is identical to GRIM19, a protein implicated in interferon-β and retinoic acid-induced apoptotic cell death (Lufei et al. 2003; Huang et al. 2004)

R8: The human homologue displays structural similarity to thioredoxin-like ferredoxins (Brockmann et al. 2004)

R9: The total molecular mass of all constituent proteins is 949,649.4 Da; including all known prosthetic groups, the total molecular mass of complex I is 954,442.4 Da

[a] Predicted using servers http://www.enzim.hu/hmmtop/ and http://www.cbs.dtu.dk/services/TMHMM/

[b] Identified by Edman degradation

[c] Average molecular masses of mature proteins, calculated using the "compute pI/Mw" tool at http://www.expasy.ch/tools/pi_tool.html

[d] Including hexa-histidine tag and hexa-alanine spacer (Kashani-Poor et al. 2001a)

[e] Phosphopantetheine-hydroxy-tetradecanoate

[f] Djafarzadeh et al. (2000)

[g] Abdrakhmanova et al. (2004)

fungal enzyme as a model system for mammalian complex I offers a number of advantages. Fungal complex I is more stable and more easily purified in active form. Also, it does allow site-directed mutagenesis not only of the nuclear coded central but also the accessory subunits, including the remodelling and in vitro analysis of genetic defects in human complex I (see below).

2D and 3D models for the overall shape of complex I have been constructed by averaging electron microscopic pictures of single particles. While the resolution of this method is limited, the most detailed three-dimensional pictures by using this method were obtained using purified *Y. lipolytica* complex I (Clason et al. 2007). The two arms of complex I from several bacterial or eukaryotic sources have been dissociated using detergents or chaotropic reagents and then purified or further split into subcomplexes. Low-resolution images of fragments corresponding to the peripheral and the membrane intrinsic arm of *N. crassa* complex I have been published as early as 1991 (Hofhaus et al. 1991).

Before of high-resolution structural images of complex I became available, the X-ray structures of evolutionary-related enzymes, namely, the water-soluble [NiFe] hydrogenases from several eu- and archaebacterial sources (Montet et al. 1997) proved to be valuable models for the part of complex I that harbours the catalytic core of the enzyme, a designation we introduced for the domains immediately involved in the reduction of the quinone substrate and in coupling this reaction to proton translocation across the respiratory membrane (Kashani-Poor et al. 2001b). Bacterial [NiFe] hydrogenases consist of a large and a small subunit. They are related to the 49-kDa and PSST subunits of complex I, respectively. By generating a large number of amino acid substitutions in complex I of *Y. lipolytica* and assessing their effects on catalytic activity, inhibitor sensitivity and the electron paramagnetic resonance (EPR) signature of iron–sulfur cluster N2 (Ahlers et al. 2000b; Kashani-Poor et al. 2001b; Grgic et al. 2004; Garofano et al. 2003), we reached the following conclusions: The overall structure of the complex I catalytic core, located at the interface between the 49-kDa and PSST subunits, is highly similar to the corresponding part of bacterial [NiFe] hydrogenases. Iron–sulfur cluster in the PSST subunit corresponds to the so-called proximal iron–sulfur cluster in the hydrogenase small subunit, while the quinone-binding site in the 49-kDa subunit corresponds to the hydrogen evolving [NiFe] centre in the hydrogenase large subunit (Brandt et al. 2005).

A wealth of complex I inhibitors have been identified and several attempts at classification have been made (Degli Esposti 1998; Friedrich et al. 1994). The fact that several mutations in the catalytic core region of complex I affected the affinity for representative inhibitors from classes A (DQA, 2-decyl-4-quinazolinyl amine) and B (rotenone) provided strong support for the hypothesis that all classes of complex I inhibitors and the hydrophobic substrate ubiquinone bind to overlapping domains within a common binding pocket, as proposed in an earlier study using direct competition assays (Okun et al. 1999).

Later, the X-ray structure of the peripheral arm of *Thermus thermophilus* complex I was solved (Sazanov and Hinchliffe 2006). In this structure, the redox prosthetic groups of complex I appear arranged in a chain starting with the

non-covalently bound FMN molecule in the 51-kDa subunit that serves as the direct oxidant of NADH. Electrons are then passed on via an unbranched chain of iron–sulfur clusters that starts with tetranuclear cluster N3 in the 51-kDa subunit and terminates with binuclear cluster N2 in the homologue of the PSST subunit of bovine complex I. Binuclear cluster N1c, which is found in some bacterial complexes I only, does not participate in electron transfer reactions and merely serves a structural role.

It should be noted that there is an ongoing controversy on the assignment of the intervening clusters to the EPR signals that contribute to the EPR signatures. The "traditional" assignments, largely based on the work of Ohnishi and co-workers (Ohnishi 1998; Sazanov and Hinchliffe 2006), was recently challenged by Hirst and co-workers (Yakovlev et al. 2007; Roessler et al. 2010). While discussing this issue is beyond the scope of the present article, the "traditional" and "revised" assignments specific EPR signals to the iron–sulfur cluster bearing subunits of complex I are presented in Table 1.

The X-ray structure of the peripheral arm of *Th. thermophilus* complex I confirmed the notion that there is a high degree of structural similarity between bacterial [NiFe] hydrogenases and complex I. The interface between the 49-kDa and the PSST subunits, including the domains that correspond to [NiFe] cluster-binding site, forms a large cavity at the periphery of the fragment. It is widely accepted that this part of the enzyme constitutes much of the ubiquinone-binding pocket, within which the substrate has to interact with its specific binding site for productive interaction, i.e. reduction, coupled to proton pumping. Further research, including site-directed mutagenesis in *Y. lipolytica*, has provided a wealth of detailed information on this part of complex I (see below).

However, the position and orientation of the peripheral arm fragment and thus of the ubiquinone-binding pocket with respect to the membrane arm remained unclear. Also the involvement of the PSST subunits in crystal contacts and the fact that some sequence stretches within the 49-kDa and PSST subunits could not be resolved in the structure, left room for speculations on the actual shape and location of the ubiquinone-binding pocket within the holoenzyme. Meanwhile, the X-ray structural analysis of both the complete bacterial and mitochondrial complex I (Hunte et al. 2010; Efremov et al. 2010; Baradaran et al. 2013) revealed that the immediate electron donor of iron-sulfur cluster N2 resides about 30 Å above the polar headgroups of the membrane surface. This confirmed several lines of evidence indicating that this distance is at least 30 Å (Clason et al. 2007, 2010; Zickermann et al. 2003). In the structural models derived from these studies, the orientation of the quinone-binding pocket correctly suggested that it opens to the bulk phase and that the extremely hydrophobic substrate has to leave the membrane level at least in part. Clearly, the X-ray structure of the complex I holoenzyme was required to unambigously elucidate the orientation of the peripheral arm with respect to the membrane arm.

4 A Genetic Approach to *Y. lipolytica* Complex I

In the following, we report on our efforts to establish *Y. lipolytica* as a model system for the genetic and proteomic analysis of complex I. Through long years of spadework, a wealth of information has been obtained and a number of methodological approaches have been developed that now allow tackling various aspects of complex I structure, function and assembly.

All seven nuclear genes encoding the central subunits of the peripheral arm of complex I, as well as several others, could be cloned by virtue of sequence similarity with their homologues in fungi like *N. crassa* and in mammals. Following completion of the *Y. lipolytica* genome project, in addition to the seven nuclear coded central, a total of 27 genes for accessory subunits of *Y. lipolytica* complex I have been identified up to now (Table 2). Evidence that the protein products of these genes are bona fide subunits of *Y. lipolytica* complex I has been provided by proteomic analysis of the purified enzyme. For a subset of genes, including all seven central nuclear encoded subunits, deletion by homologous recombination was achieved, which either completely abolished complex I assembly or resulted in the appearance of subcomplexes that most likely represent relatively stable assembly intermediates (unpublished results).

As discussed above, the generation of strains carrying deletions of individual genes for complex I subunits was only achieved after genomic integration of the artificial NDH2i gene, encoding an internal version of alternative NADH dehydrogenase. Following replacement of the genomic copy of the NUGM gene encoding the 30-kDa subunit of complex I with a version carrying a C-terminally attached His-tag, affinity purification of *Y. lipolytica* complex I could be achieved with high purity and yield. To our best knowledge, this method holds the record for the largest multiprotein assembly purified to date using a single affinity tag.

However, strain development is still an ongoing process. For example, it can be expected that deletions of genes encoding complex I specific assembly factors or chaperones will yield intermediate phenotypes, i.e. result in an incomplete loss of complex I function. In such cases, it would be desirable to carry out suppressor or enhancer screens. This and other approaches will involve classical genetic crosses, employing truly isogenic strains. Such analyses are plagued with low mating and sporulation frequencies, low spore viabilities and non-uniform genetic background of the available *Y. lipolytica* strains. It has long been speculated that there is a common reason behind these problems: Multiple genomic rearrangements between individual "wild-type" isolates might prevent isogenisation by standard inbreeding programmes. An additional problem is presented by the fact that the frequency of homologous recombination is comparably low in *Y. lipolytica*, with the vast majority of recombinants resulting from random integration events.

Recently, we have put some effort into improving this situation. First, we deleted the homologue of the *N. crassa* mus51 gene (YALI0C08701g). In strains lacking this central component of the non-homologous end joining (NHEJ) system, a strong prevalence of homologous recombination repair (HRR) was observed following

transformation: a linear DNA fragment containing a *URA3* (gene encoding orotidine-5′-phosphate decarboxylase) marked deletion of a gene for an accessory complex I subunit was integrated by homologous recombination with the genomic locus in 12 out of 12 cases, a figure which is in stark contrast to the 0.25–12.5 % of homologous recombination events that we achieved in similar gene deletion projects using standard laboratory strains of *Y. lipolytica* (unpublished observation).

To generate truly isogenic strains of *Y. lipolytica*, we have replaced the mating-type locus from MATA to MATB in a YALI0C08701g deletion strain using pop-in-pop-out recombination. However, the project has not been finished yet, and mating and sporulation frequencies of strains isogenised by this “sex-reversal” procedure are yet unknown.

5 The Mitochondrial Genome of *Y. lipolytica*

In *Y. lipolytica*, like in mammals, the seven central, highly hydrophobic subunits of the membrane arm of complex I are encoded by the mitochondrial genome. Cross-hybridisation of a *Podospora anserina* mitochondrial DNA (mtDNA) probe with *Y. lipolytica* total DNA (Nosek and Fukuhara 1994) was among the first pieces of evidence for the presence of complex I in this obligate aerobic yeast; the sequence of a 6.6-kb *Y. lipolytica* mtDNA fragment containing the ND4 gene was published shortly afterwards (Matsuoka et al. 1994). The complete sequence of the *Y. lipolytica* mitochondrial genome could be determined at an early stage of the *Y. lipolytica* genome project by combining sequence information from a total DNA “shotgun” library and from contigs obtained by cloning of “satellite” DNA purified using cesium chloride gradient ultracentrifugation (Kerscher et al. 2001a).

Fourteen respiratory chain proteins are encoded by the *Y. lipolytica* mtDNA sequence, including the seven hydrophobic subunits (ND1-6 and ND4L) of the membrane arm of NADH:ubiquinone oxidoreductase (complex I), apocytochrome *b* of the cytochrome bc_1 complex (complex III), three subunits (COX1, 2, 3) of cytochrome *c* oxidase (complex IV) and three subunits of ATP synthase (complex V). N-terminal sequencing of several mitochondrial respiratory chain proteins clearly demonstrated that the *Y. lipolytica* mitochondrial genome is translated according to the “mold mitochondrial” translation table. However, up to date (February 2010), the NCBI database erroneously assigns “Translation table 3 (Yeast Mitochondrial)” as mitochondrial genetic code of *Y. lipolytica*.

The two RNA components of the mitochondrial ribosomes and 27 mitochondrial tRNAs were also found. The analysis of the *Y. lipolytica* mtDNA sequence revealed some unexpected features: the set of mitochondrial tRNAs is incomplete there is no tRNA able to read GCN codons encoding arginine, and, as a consequence, such codons are absent from the protein coding genes listed above. A total of 17 introns were found in the mitochondrial genome of *Y. lipolytica*. All of them fall into class I of fungal introns, characterised by a common secondary structure, a unique splicing

mechanism and the presence (at least in the majority of cases) of intronic open reading frames encoding endonucleases/maturases which may assist in the splice process or in "homing" into cognate intronless alleles. In the case of the *Y. lipolytica* mitochondrial genome, most of these intronic ORFs must be regarded as functionless pseudogenes, either because they have accumulated multiple frame-shift and/or point mutations or because they contain untranslatable CGN codons.

6 A Proteomic Approach to *Y. lipolytica* Complex I

Mitochondrial membranes and purified complex I from *Y. lipolytica* were characterised extensively using several types of two-dimensional polyacrylamide gel electrophoresis (Abdrakhmanova et al. 2004, 2005; Djafarzadeh et al. 2000). For mitochondrial membranes, a combination of blue native polyacrylamide electrophoresis (BN-PAGE) and denaturing sodium dodecyl sulfate (SDS-PAGE) separation was routinely employed for assessing complex I assembly in strains carrying specific point mutation or gene deletions, as silver staining or Western blotting of such gels permits the detection of fully assembled complex I and subcomplexes that may either represent assembly intermediates or degradation products. Supercomplexes that contain complex I are also detectable but are not well resolved by this method. Analysis of purified complex I by dSDS-PAGE (in which the "d" stands for "doubled"), a combination of denaturing electrophoretic separation in the absence and presence of 16 % urea (Rais et al. 2004), proved especially useful for disclosing its constituent proteins. Due to the large number of subunits, a complex pattern of protein spots could be observed. A typical example is shown in Fig. 2. Separation of protein mixtures by dSDS-PAGE results in a typical pattern, with hydrophobic proteins running above and highly acidic proteins below the diagonal that contains the bulk of the proteins. In the case of *Y. lipolytica* complex I, six of the seven highly hydrophobic proteins encoded by the mitochondrial genome produced clearly visible spots on silver-stained dSDS gels, while the small ND4L subunit was not detectable. Assignment of individual protein spots to the corresponding genes was a formidable task that required the extensive application of various proteomic methods, most notably tandem mass spectrometry. Also, N-terminal sequences of several subunits of *Y. lipolytica* complex I were determined using Edman degradation.

However, dSDS gels do not allow accurate determinations of subunit masses. At this point, valuable insights could be gained by applying the newly developed method called laser-induced liquid bead ion desorption (LILBID) mass spectrometry (Morgner et al. 2008). This method allows the determination of molecular masses of undigested proteins and even non-covalent multiprotein complexes. As a drawback, at least at the current stage of technical development, it provides only limited mass accuracy, but this is far outweighed by the fact that it works very well with membrane proteins and can be applied to protein mixtures of which only minimal amounts are required. LILBID spectrometry of purified *Y. lipolytica*

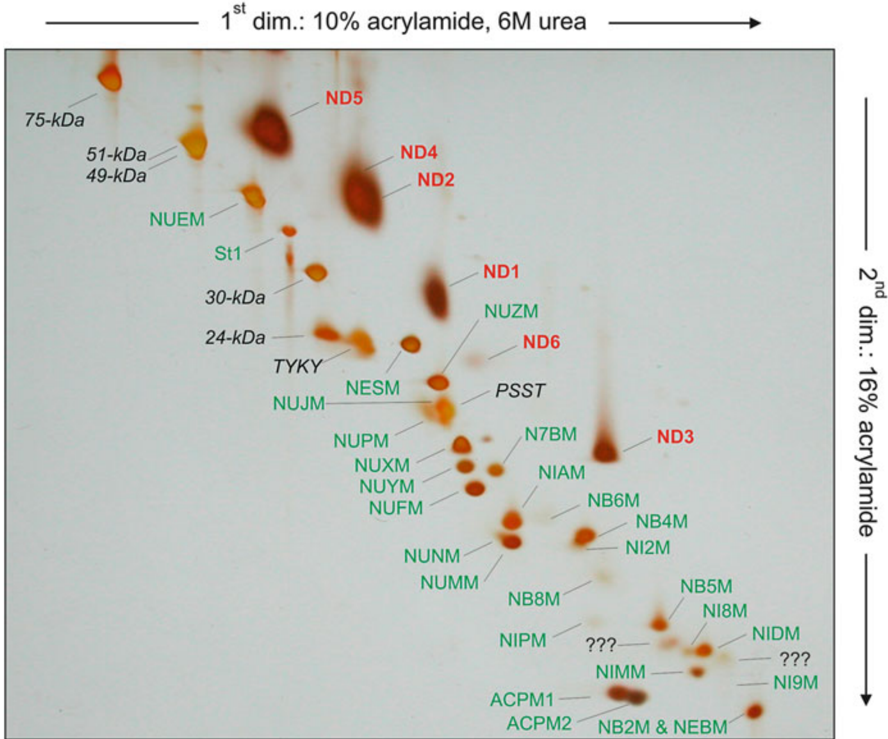


Fig. 2 dSDS-PAGE analysis of affinity-purified *Y. lipolytica* complex I. His-tagged complex I was purified using Ni-NTA Sepharose, followed by gel filtration on a TSK 4000SW column, separated on dSDS gels and stained with silver. Individual protein spots were cut out and subjected to tandem mass spectrometry. Hydrophobic, mitochondrially coded central subunits of the membrane arm (running above the diagonal) are labelled in **bold face**, hydrophilic central subunits of the peripheral arm in *italic* and accessory subunits of the peripheral arm in standard letters. All known subunits of *Y. lipolytica* complex I could be assigned to spots in the gel, with the sole exception of the small, mitochondrially coded ND4L protein. Two unidentified spots, most likely corresponding to contaminating proteins, are marked with "*question marks*"

complex I demonstrated that for seven of the small accessory subunits, peak positions were inconsistent with the masses deposited in the genome project. Careful searches for potential splice donor and acceptor sites in the corresponding genomic regions, followed by direct sequencing of cDNAs obtained by reverse transcription polymerase chain reaction (RT-PCR), indicated that the respective gene annotations had to be corrected.

All proteomic data on the 14 central and the 27 accessory subunits of *Y. lipolytica* complex I are summarised in Table 2. The roles of individual subunits in fungal and in mammalian complex I are listed as well—please note that as indicated part of this information is based on sequence similarities only and has not been confirmed experimentally. Especially, the function of most of the accessory subunits is not known. Among them, a group of small accessory subunits, all

possessing one transmembrane helix flanked by two highly charged domains, is remarkable. The number of such subunits seems to be highly variable between taxa, with 9 and 14 group members in *Y. lipolytica* and in mammals, respectively. Also, the degree of sequence similarity is rather low between taxa, and while some subunits can be assigned as homologues, their unifying feature seems to be functional analogy. We have speculated that these proteins, termed STMD (single transmembrane domain) subunits, function as assembly factors or chaperones but are retained in the fully assembled complex (Brandt et al. 2005; Zickermann et al. 2010). Interestingly, deletions of several of the corresponding genes (NIAM, NUNM, NB8M) resulted in similar complex I defects: only a very minor fraction of the enzyme was fully assembled, but the majority of the subunits, including all seven central subunits of the peripheral arm, were found within a large, membrane-associated subcomplex that migrated at an apparent molecular mass of about 700 kDa in BN-PAGE (unpublished results). A similar assembly defect was observed following deletion of the gene encoding the NUJM subunit, which has three transmembrane helices (unpublished results). In contrast, the STMD subunit NEBM seems to play a different role in complex I assembly. Following deletion of the NEBM gene, a hydrophobic subcomplex with a molecular mass of about 300 kDa was found (unpublished results). These observations support the notion that the STMD subunits, and the NUJM subunit, are involved in "joining" precursors of the peripheral and the membrane arm of complex I during assembly.

Although the proteomic analysis of *Y. lipolytica* complex I is now well advanced, we cannot be absolutely sure that all subunits have been identified. Both in the dSDS gel pattern and in the LILBID profile of *Y. lipolytica* complex I, a few unassigned features are apparent. These may represent additional, yet unidentified subunits of complex I or may arise from contaminating proteins, proteolytic breakdown or similar artefacts. However, by comparison of the subunits identified in *Y. lipolytica* with those identified in other fungi, we have reason to believe that our list is almost complete. For all subunits that were identified in complex I from *N. crassa*, the corresponding *Y. lipolytica* proteins could be identified, with one exception only: No counterpart for the NURM subunit of *N. crassa* (NCBI accession number: X71414) could be found so far in *Y. lipolytica*. Conversely, no counterpart of the *Y. lipolytica* NUNM gene (Genolevures accession number: YALI0F14003g) could be found in *N. crassa*. The fact that both genes encode STMD subunits seems to add further support to the hypothesis that there is little evolutionary conservation among this class of subunits.

Most of *Y. lipolytica* complex I seems to be present in mitochondrial membranes in monomeric form. Supercomplexes of respiratory chain complexes do occur as well but appear to be less frequent than in mammalian mitochondria (Nübel et al. 2009). In a careful study aimed at characterising complex I containing supercomplexes in mitochondrial membranes from *Y. lipolytica*, 2D BN gels were employed, using BN-PAGE as described in Wittig et al. (2006) in the first dimension and BN-PAGE with 0.02 % dodecyl-β-D-maltoside in the cathode buffer

Table 3 BN-PAGE of *Y. lipolytica* mitochondrial membranes reveals complex I in monomeric form or as part of supercomplexes

Band designation	Tentative composition	Apparent mass in BN-PAGE
I_1	I_1	950
d	I_1III_2 and I_1IV_2	1,400
e	$I_1III_2IV_1$	1,600
f	$I_1III_2IV_2$	1,800
⋆	I_1IV_1	1,150
♦	I_2	2,000

Adapted from Nübel et al. (2009)

in the second dimension. This analysis led to the identification of six such complexes, as summarised in Table 3 (Nübel et al. 2009).

A major advantage of the LILBID method is that it works with proteins eluted from BN-PAGE gels. Thus, LILBID has great potential for analysing the subunit composition of multiprotein complexes, including supercomplexes that are hard to purify with good yield and purity. This is illustrated by the fact that *Y. lipolytica* complex I prepared by the 2D BN-PAGE strategy described above is suitable for LILBID analysis (Sokolova et al. 2010). In the first dimension, using standard BN-PAGE, monomeric complex I is not well separated from dimeric complex V. In the second dimension, dimeric complex V is dissolved into monomers, and the complex I band can be excised. LILBID spectra of BN-PAGE bands were very similar to those obtained with affinity-purified complex I, but a few extra peaks were present as well which were most likely were due to complex V subunits. To identify such contaminating proteins more easily, LILBID spectra of dimeric *Y. lipolytica* complex V, isolated from BN gels in a similar fashion, were recorded (Sokolova et al. 2010). LILBID peaks could be assigned to 13 of the 15 complex V subunits identified by searching the *Y. lipolytica* genome. Again, in some cases, the splice pattern deposited in the genome project database had to be revised. The complex V data are summarised in Table 4.

7 Subunits Specific to *Y. lipolytica* Complex I

Thiosulfate:cyanide sulfurtransferase, also referred to as rhodanese, is an enzyme that is able to catalyse in vitro the transfer of a sulfane sulfur atom from thiosulfate to cyanide yielding thiocyanate and sulfite (Bordo and Bork 2002). While its in vivo function is still unknown, a number of physiological roles have been suggested, including involvement in the synthesis or repair of the iron–sulfur clusters of complex I (Pagani and Galante 1983; Ogata and Volini 1990).

Tandem mass spectrometric analysis demonstrated that rhodanese is present in *Y. lipolytica* complex I, purified by His-tag-mediated Ni^{2+}-affinity chromatography and gel filtration. The preparation exhibited a low level of in vitro rhodanese

Table 4 Subunits of *Y. lipolytica* complex V (ATP synthase)

ATP synthase subunit	Gene locus/gene name	Swiss-Prot/Uni-Prot	N-terminal sequence[a]	M_r		LILBID
				Precursor	Mature[b]	m/z
α	YALI0F03179g	Q6C326_YARLI	AEKATPEV[c]	58,078.64	55,407.47	55.5
β	YALI0B03982g	Q6CFT7_YARLI	ASSAGVGSGK[c]	54,406.97	50,894.76	51.0
γ	YALI0F02893g	Q6C338_YARLI	atLREIEMR[c]	32,336.8	30,074.18	30.1
b	YALI0F20306g	Q6C105_YARLI	sNQVDPKVKaT[c]	24,002.53	22,444.69	22.4
d	YALI0B06831g	Q6CFH9_YARLI	sVAAARSsAVK[c]	19,820.58	19,689.39	19.7
OSCP	YALI0D12584g	Q6C9B1_YARLI	ATKAAAPVKV[c]	22,776.33	20,518.47	20.5
h	YALI0F04774g	Q6C2V6_YARLI		11,636.89	9,456.33	9.5
ε	–	–	SAwKSAGFSF[d]	6,785.84	6,654.64	6.7
δ	YALI0D22022g	Q6C877_YARLI	AETASVDKLR[d,e]	17,349.51	14,720.38	14.7
i/j	YALI0D17490g	Q6C8S0_YARLI	aFGIRRAYPT[d]	6,941.05	6,809.86	6.8
g	YALI0B21527g	B5FVB8_YARLI	qstAASKAAS[d]	15,087.4	13,261.25	
e	YALI0E32164g	B5FVG3_YARLI	satLNVLR[d]	9,989.39	9,858.2	9.9
8	ATP8[f]	ATP8_YARLI	mPQLvPFyFT[d]	–	5,771.89	5.8
c	ATP9[f]	ATP9_YARLI	mQLvLA[d]	–	7,716.34	7.7
a	ATP6[f]	ATP6_YARLI		–	27,808.24	

[a]Identified by Edman degradation of Coomassie-stained bands on PVDF membranes
[b]Without posttranslational modifications
[c]From Nübel et al. (2009)
[d]From Sokolova et al. (2010)
[e]Under deformylation conditions
[f]Genes published in Kerscher et al. (2001a)

activity (Abdrakhmanova et al. 2005). These observations provided the first evidence that rhodanese forms a subunit of respiratory chain complex I. However, it seems that rhodanese is bound more loosely than other subunits, as it dissociates from the enzyme during BN-PAGE. Most likely, this interaction is even weaker and could therefore not be detected in other model systems. A genomic deletion of the *Y. lipolytica* rhodanese gene did not result in a detectable complex I defect (Abdrakhmanova et al. 2005). Activity and assembly of complex I appeared unchanged with respect to the parental strain, demonstrating that whatever in vivo function rhodanese may have, it is not essential for survival, at least under standard laboratory conditions.

Another unexpected finding with *Y. lipolytica* complex I was the presence of two subunits that are isoforms of ACPM (Dobrynin et al. 2010), mitochondrial proteins that exhibit strong sequence similarity to acyl carrier proteins, subunits of the prokaryotic type fatty acid synthase complex. In contrast, complex I from all other model systems analysed so far contains one ACPM subunit only (Cronan et al. 2005; Sackmann et al. 1991).

The function of the ACPM proteins remains elusive. However, homologues of the full set of enzymes involved in prokaryotic fatty acid synthesis have been detected in *S. cerevisiae* mitochondria (Hiltunen et al. 2005; Hiltunen et al. 2008), and several hypotheses regarding their functional roles have been put forward. These include de novo or repair synthesis of mitochondrial membrane phospholipids and synthesis of octanoic acid, a precursor of lipoic acid which acts as a redox prosthetic group in the pyruvate dehydrogenase and α-ketoglutarate dehydrogenase (Wada et al. 1997) complexes. Deletion of the genes encoding the two ACPM subunits of *Y. lipolytica* complex I also gave an unexpected result: while deletion of ACPM1 (locus name YALI0D14850g) interfered with proper assembly of complex I, deletion of ACPM2 (locus name YALI0D24629g) was incompatible with survival, indicating that this subunit, in addition to its function as a subunit of complex I, has an essential role in *Y. lipolytica* (Dobrynin et al. 2010).

8 Exploring Human Complex I Defects

Among inherited respiratory chain deficiencies, complex I deficiency is observed in many cases (Smeitink et al. 2004). A typical clinical presentation is Leigh syndrome, a severe, progressive neurodegenerative disease, with characteristic brainstem defects observed in magnetic resonance tomography (MRT) scans. A large number of mitochondrial mutations have been observed in Leigh syndrome patients, but point mutations in nuclear coded complex I subunits have also been described. In order to analyse their effects on catalytic activity, we have reconstructed several human complex I mutations using the *Y. lipolytica* system (Ahlers et al. 2000a). Mitochondrial membranes from mutant strains were isolated and their activities tested in vitro. By comparison with the enzyme from the parental strain, only rather small differences were observed. Even though great care has to

be applied when data obtained with the same enzyme from different organisms are to be compared, these results are plausible: Although human Leigh syndrome mutations have a lethal outcome, the fact that patients may survive for periods of time ranging from a few days to several years indicates that the resulting complex I defects are relatively mild. The positions of human Leigh syndrome mutations, using bacterial [NiFe] hydrogenases or the peripheral arm of *Th. thermophilus* complex I as structural models, are consistent with this view. Typically, they are found at a considerable distance from the catalytic core of complex I, at the periphery of the PSST and the 49-kDa subunits (Kerscher et al. 2004).

For a rather large fraction of Leigh syndrome patients, no mutations in complex I subunit genes could be identified, and it has been suspected that they may carry genetic defects affecting complex I specific assembly factors. Indeed, several such factors have been identified in recent years, either by the detection of the gene defects underlying familial Leigh or Leigh-like syndromes or by the identification of proteins associated with complex I or its assembly intermediates, followed by gene knockout or knockdown in a variety of model systems (Lazarou et al. 2009). The latter group constitute valuable candidate genes for elucidating the molecular defects underlying yet enigmatic cases of familial complex I defects.

Also, the method of evolutionary profiling has proven useful for the identification of complex I assembly factors. It is based on a comparison of the genetic inventories of a large number of organisms, confronting those that do possess complex I with those that do not. Genes present in most members of the first but absent in members of the second group may be regarded as candidates for complex I specific assembly factors. This approach led to the identification of the B17.2L protein, a paralogue of the complex I subunit B17.2 as a complex I specific chaperone, and to the identification of a patient with a progressive encephalopathy that carried a null mutation in the gene encoding B17.2L (Ogilvie et al. 2005).

A similar approach led to the identification of Ind1, a protein specifically required for the assembly of the iron–sulfur clusters of complex I. Gene deletion in *Y. lipolytica* (Bych et al. 2008) and siRNA-based knockdown in HeLa cells led to similar effects (Sheftel et al. 2009), demonstrating evolutionary conservation of this factor: In both cases, complex I content was reduced to bout 30 % of the original level, while other iron–sulfur cluster containing proteins, either residing in mitochondria or the cytosol, was largely unaffected.

9 Exploring the Reaction Mechanism of Complex I

The reaction mechanism of complex I is far from being understood. Two central questions have emerged in recent years: (1) how does the hydrophobic substrate ubiquinone bind and reach the active site and (2) how is electron transfer linked to proton translocation. These questions may be closely related. As discussed above, it has been shown that the orientation of the peripheral arm within the holoenzyme is such that the ubiquinone-binding site of complex I is situated about 30 Å above the

respiratory membrane (Hunte et al. 2010; Efremov et al. 2010; Baradaran et al. 2013). On the other hand, ubiquinone is a highly hydrophobic molecule that partitions almost quantitatively into the membrane phase. To reconcile this apparent contrast, we have postulated early on that a hydrophobic access path or "ramp" connects the membrane with the ubiquinone reduction site (Zickermann et al. 2003, 2009).

Evidence from our previous mutational analyses in *Y. lipolytica*, guided by the [NiFe] hydrogenase model, indicated that reduction of the hydrophobic substrate ubiquinone takes place at the interface between the 49-kDa and PSST subunits and that several complex I inhibitors bind to the same region (Kerscher et al. 2001c). This picture could be refined in a further round of mutagenesis, this time guided by the X-ray structure of the peripheral arm of *Th. thermophilus* complex I (Tocilescu et al. 2007; Fendel et al. 2008). In this work, for all amino acids with surface-exposed side chains in the proposed quinone-binding pocket of the *Th. thermophilus* enzyme, substitutions at the corresponding positions were created in *Y. lipolytica* complex I.

This round of mutagenesis yielded a number of interesting insights: The quinone-binding pocket of complex I has the shape of a funnel with a rather wide opening that is formed by the 49-kDa and PSST subunits, tapering into a narrow cleft formed by the interface of the two subunits, and terminates near iron–sulfur cluster N2 in the PSST subunit and the highly conserved Y144 residue in the 49-kDa subunit. The fact that even the most subtle change affecting the highly conserved Y144 residue in the 49-kDa subunit, namely, the Y144F replacement that merely removes the hydroxyl group, completely abolishes catalytic activity with the substrate analogue DBQ (*n*-decyl-ubiquinone) strongly suggests that this residue plays a key role in the catalytic cycle (Tocilescu et al. 2010). Complex I assembly and iron–sulfur cluster N2 EPR signals were unaffected by this and several other mutations at the same position. A more detailed analysis of the effects of the Y144F exchange became possible when it was discovered that complex I from this mutant displayed marked activity with the substrate analogues Q_1 and Q_2. The most likely explanation for this finding is that these molecules still bind to the ubiquinone reduction site. Although their hydrophobic side chains are shorter, their isoprene units may provide additional binding interactions allowing productive binding to the enzyme that is not feasible with the simple aliphatic side chain of DBQ.

Affinities for DQA and rotenone were dramatically decreased, indicating that Y144 plays a key role in the binding of inhibitors and the substrate ubiquinone. We propose that Y144 directly interacts with the headgroup of ubiquinone, most likely via a hydrogen bond between the aromatic hydroxyl and the ubiquinone carbonyl, placing the substrate in an ideal distance to its electron donor, iron–sulfur cluster N2. However, proton pumping still takes place with unchanged stoichiometry in complex I from the Y144F mutant, clearly demonstrating that Y144 is not involved in proton pumping (Tocilescu et al. 2010).

The nature of the proton pump of complex I is still unknown. Since cluster N2 of bovine heart complex I displays a marked redox-Bohr effect in the physiological pH range, it had been postulated that this cluster could be involved in proton

translocation. This hypothesis, however, was ruled out using the H226M mutation in the 49-kDa subunit of *Y. lipolytica* complex I (Zwicker et al. 2006). Mutating this highly conserved residue abolished the redox-Bohr properties of cluster N2 that resides in the immediate vicinity of H226. In the *Th. thermophilus* structure, the corresponding residue had been shown to form a hydrogen bond to cluster N2. However, in the H226M mutant, proton pumping was unaffected and occurred with unchanged stoichiometry, as shown by measurements using purified complex I reconstituted into proteoliposomes.

These observations led us to conclude that all proton pumping in complex I has to be coupled to the reduction of ubiquinone, which occurs in two steps, via formation of a semiquinone. It seems feasible that this highly charged intermediate has to bind to the active site in a conformation that differs from that in the presence of the uncharged substrate or product molecules and that an energetic barrier has to be passed for the required conformational rearrangement to occur at the substrate binding site (Zickermann et al. 2009). However, the question how this mechanism triggers proton translocation remains open. While their homology to Na^+/H^+ antiporters strongly suggests the ND2, ND4 and ND5 subunits of the membrane arm as constituents of the proton pump (Mathiesen and Hägerhall 2003), it is not easy to explain how conformational changes at the site of ubiquinone reduction could be transmitted across such distances. It is tempting to speculate that domains of several subunits, encoded by both the mitochondrial and the nuclear genomes, may serve as transmitter and trigger to drive the proton pump of complex I. Thus, at least some of these domains should be amenable to site-directed mutagenesis in *Y. lipolytica*.

Future work, involving a combination of experimental approaches to address both the number and location of semiquinone species and the conformational changes within the complex I catalytic cycle, will be required to clarify this matter.

References

Abdrakhmanova A, Zickermann V, Bostina M, Radermacher M, Schägger H, Kerscher S, Brandt U (2004) Subunit composition of mitochondrial complex I from the yeast *Yarrowia lipolytica*. Biochim Biophys Acta 1658:148–156

Abdrakhmanova A, Dobrynin K, Zwicker K, Kerscher S, Brandt U (2005) Functional sulfurtransferase is associated with mitochondrial complex I from *Yarrowia lipolytica*, but is not required for assembly of its iron-sulfur clusters. FEBS Lett 579:6781–6785

Abdrakhmanova A, Zwicker K, Kerscher S, Zickermann V, Brandt U (2006) Tight binding of NADPH to the 39-kDa subunit of complex I is not required for catalytic activity but stabilizes the multiprotein complex. Biochim Biophys Acta 1757:1676–1682

Ahlers P, Garofano A, Kerscher S, Brandt U (2000a) Application of the obligate aerobic yeast *Yarrowia lipolytica* as a eukaryotic model to analyze Leigh Syndrome mutations in the complex I core subunits PSST and TYKY. Biochim Biophys Acta 1459:258–265

Ahlers P, Zwicker K, Kerscher S, Brandt U (2000b) Function of conserved acidic residues in the PSST-homologue of complex I (NADH:ubiquinone oxidoreductase) from *Yarrowia lipolytica*. J Biol Chem 275:23577–23582

Arthur H, Watson K (1976) Thermal adaptation in yeasts: growth temperatures, membrane lipid, and cytochrome composition of psychrophilic, mesophilic, and thermophilic yeasts. J Bacteriol 128:56–68

Baradaran R, Berrisford JM, Minhas GS, Sazanov LA (2013) Crystal structure of the entire respiratory complex I. Nature 494:443–448

Barth G, Gaillardin C (1996) *Yarrowia lipolytica*. In: Wolf K (ed) Non-conventional yeasts in biotechnology. Springer, Heidelberg, pp 313–388

Bordo D, Bork P (2002) The rhodanese/Cdc25 phosphatase superfamily – sequence-structure-function relations. EMBO Rep 3:741–746

Brandt U (2006) Energy converting NADH: quinone oxidoreductase (complex I). Annu Rev Biochem 75:69–92

Brandt U, Abdrakhmanova A, Zickermann V, Galkin A, Dröse S, Zwicker K, Kerscher S (2005) Structure–function relationships in mitochondrial complex I of the strictly aerobic yeast *Yarrowia lipolytica*. Biochem Soc Trans 33:840–844

Brockmann C, Diehl A, Rehbein K, Strauss H, Schmieder P, Korn B, Kuhne R, Oschkinat H (2004) The oxidized subunit B8 from human complex I adopts a thioredoxin fold. Structure 12:1645–1654

Bych K, Kerscher S, Netz DJ, Pierik AJ, Zwicker K, Huynen MA, Lill R, Brandt U, Balk J (2008) The iron-sulphur protein Ind1 is required for effective complex I assembly. EMBO J 27:1736–1746

Carroll J, Shannon RJ, Fearnley IM, Walker JE, Hirst J (2002) Definition of the nuclear encoded protein composition of bovine heart mitochondrial complex I – identification of two new subunits. J Biol Chem 277:50311–50317

Carroll J, Fearnley IM, Skehel JM, Shannon RJ, Hirst J, Walker JE (2006) Bovine complex I is a complex of 45 different subunits. J Biol Chem 281:32724–32727

Clason T, Zickermann V, Ruiz T, Brandt U, Radermacher M (2007) Direct localization of the 51 and 24 kDa subunits of mitochondrial complex I by three-dimensional difference imaging. J Struct Biol 159:433–442

Clason T, Ruiz T, Schagger H, Peng G, Zickermann V, Brandt U, Michel H, Radermacher M (2010) The structure of eukaryotic and prokaryotic complex I. J Struct Biol 169:81–88

Cronan JE, Fearnley IM, Walker JE (2005) Mammalian mitochondria contain a soluble acyl carrier protein. FEBS Lett 579:4892–4896

Degli Esposti M (1998) Inhibitors of NADH-ubiquinone reductase: an overview. Biochim Biophys Acta 1364:222–235

Djafarzadeh R, Kerscher S, Zwicker K, Radermacher M, Lindahl M, Schägger H, Brandt U (2000) Biophysical and structural characterization of proton-translocating NADH-dehydrogenase (complex I) from the strictly aerobic yeast *Yarrowia lipolytica*. Biochim Biophys Acta 1459:230–238

Dobrynin K, Abdrakhmanova A, Richers S, Hunte C, Kerscher S, Brandt U (2010) Characterization of two different acyl carrier proteins in complex I from *Yarrowia lipolytica*. Biochim Biophys Acta 1797:152–159

Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, de Montigny J, Marck C, Neuveglise C, Talla E, Goffard N, Frangeul L, Aigle M, Anthouard V, Babour A, Barbe V, Barnay S, Blanchin S, Beckerich JM, Beyne E, Bleykasten C, Boisrame A, Boyer J, Cattolico L, Confanioleri F, De Daruvar A, Despons L, Fabre E, Fairhead C, Ferry-Dumazet H, Groppi A, Hantraye F, Hennequin C, Jauniaux N, Joyet P, Kachouri R, Kerrest A, Koszul R, Lemaire M, Lesur I, Ma L, Muller H, Nicaud JM, Nikolski M, Oztas S, Ozier-Kalegeropoulos O, Pellenz S, Potier S, Richard GF, Straub ML, Suleau A, Swennen D, Tekaia F, Wesolowski-Louvel M, Westhof E, Wirth B, Zeniou-Meyer M, Zivanovic I, Bolotin-Fukuhara M, Thierry A, Bouchier C, Caudron B, Scarpelli C, Gaillardin C, Weissenbach J, Wincker P, Souciet JL (2004) Genome evolution in yeasts. Nature 430:35–44

Efremov RG, Baradaran R, Sazanov LA (2010) The architecture of respiratory complex I. Nature 465:441–445

Fendel U, Tocilescu MA, Kerscher S, Brandt U (2008) Exploring the inhibitor binding pocket of respiratory complex I. Biochim Biophys Acta 1777:660–665
Friedrich T (1998) The NADH:ubiquinone oxidoreductase (complex I) from *Escherichia coli*. Biochim Biophys Acta 1364:134–146
Friedrich T, Ohnishi T, Forche E, Kunze B, Jansen R, Trowitzsch W, Hofle G, Reichenbach HX, Weiss H (1994) Two binding sites for naturally occurring inhibitors in mitochondrial and bacterial NADH:ubiquinone oxidoreductase (complex I). Biochem Soc Trans 22:226–230
Galkin AS, Grivennikova VG, Vinogradov AD (2001) $H^+/2e^-$ stoichiometry of the NADH: ubiquinone reductase reaction catalyzed by submitochondrial particles. Biochemistry (Mosc) 66:435–443
Galkin A, Dröse S, Brandt U (2006) The proton pumping stoichiometry of purified mitochondrial complex I reconstituted into proteoliposomes. Biochim Biophys Acta 1757:1575–1581
Garofano A, Zwicker K, Kerscher S, Okun P, Brandt U (2003) Two aspartic acid residues in the PSST-homologous NUKM subunit of complex I from *Yarrowia lipolytica* are essential for catalytic activity. J Biol Chem 278:42435–42440
Grgic L, Zwicker K, Kashani-Poor N, Kerscher S, Brandt U (2004) Functional significance of conserved histidines and arginines in the 49 kDa subunit of mitochondrial complex I. J Biol Chem 279:21193–21199
Hiltunen JK, Okubo F, Kursu VAS, Autio KJ, Kastaniotis AJ (2005) Mitochondrial fatty acid synthesis and maintenance of respiratory competent mitochondria in yeast. Biochem Soc Trans 33:1162–1165
Hiltunen JK, Schonauer MS, Kaija JA, Mittelmeier TM, Kastaniotis AJ, Dieckmann CL (2008) Mitochondrial fatty acid synthesis type II: more than just fatty acids. J Biol Chem 284:9011–9015
Hofhaus G, Weiss H, Leonard K (1991) Electron microscopic analysis of the peripheral and membrane parts of mitochondrial NADH dehydrogenase (complex I). J Mol Biol 221:1027–1043
Huang GC, Lu H, Hao AJ, Ng DCH, Ponniah S, Guo K, Lufei CC, Zeng Q, Caoj XM (2004) GRIM-19, a cell death regulatory protein, is essential for assembly and function of mitochondrial complex I. Mol Cell Biol 24:8447–8456
Hunte C, Zickermann V, Brandt U (2010) Functional modules and structural basis of conformational coupling in mitochondrial complex I. Science 329:448–451
Joseph-Horne T, Hollomon DW, Wood PM (2001) Fungal respiration: a fusion of standard and alternative components. Biochim Biophys Acta 1504:179–195
Kashani-Poor N, Kerscher S, Zickermann V, Brandt U (2001a) Efficient large scale purification of his-tagged proton translocating NADH : ubiquinone oxidoreductase (complex I) from the strictly aerobic yeast *Yarrowia lipolytica*. Biochim Biophys Acta 1504:363–370
Kashani-Poor N, Zwicker K, Kerscher S, Brandt U (2001b) A central functional role for the 49-kDa subunit within the catalytic core of mitochondrial complex I. J Biol Chem 276:24082–24087
Kerscher S (2000) Diversity and origin of alternative NADH:ubiquinone oxidoreductases. Biochim Biophys Acta 1459:274–283
Kerscher S, Okun JG, Brandt U (1999) A single external enzyme confers alternative NADH: ubiquinone oxidoreductase activity in *Yarrowia lipolytica*. J Cell Sci 112:2347–2354
Kerscher S, Durstewitz G, Casaregola S, Gaillardin C, Brandt U (2001a) The complete mitochondrial genome of *Yarrowia lipolytica*. Comp Funct Genomics 2:80–90
Kerscher S, Eschemann A, Okun PM, Brandt U (2001b) External alternative NADH:ubiquinone oxidoreductase redirected to the internal face of the mitochondrial inner membrane rescues complex I deficiency in *Yarrowia lipolytica*. J Cell Sci 114:3915–3921
Kerscher S, Kashani-Poor N, Zwicker K, Zickermann V, Brandt U (2001c) Exploring the catalytic core of complex I by *Yarrowia lipolytica* yeast genetics. J Bioenerg Biomembr 33:187–196

Kerscher S, Grgic L, Garofano A, Brandt U (2004) Application of the yeast *Yarrowia lipolytica* as a model to analyse human pathogenic mutations in mitochondrial complex I. Biochim Biophys Acta 1659:197–205

Kurtzman CP (1994) Molecular taxonomy of the yeasts. Yeast 10:1727–1740

Lagunas R (1986) Misconceptions about the energy metabolism of *Saccharomyces cerevisiae*. Yeast 2:221–228

Lazarou M, Thorburn DR, Ryan MT, McKenzie M (2009) Assembly of mitochondrial complex I and defects in disease. Biochim Biophys Acta 1793:78–88

Lufei C, Ma J, Huang G, Zhang T, Novotny-Diermayr V, Ong CT, Cao X (2003) GRIM-19, a death-regulatory gene product, suppresses Stat3 activity via functional interaction. EMBO J 22:1325–1335

Luttik MAH, Overkamp KM, Kötter P, de Vries S, van Dijken P, Pronk JT (1998) The *Saccharomyces cerevisiae NDE1 and NDE2* genes encode separate mitochondrial NADH dehydrogenases catalyzing the oxidation of cytosolic NADH. J Biol Chem 273:24529–24534

Marres CAM, de Vries S, Grivell LA (1991) Isolation and inactivation of the nuclear gene encoding the rotenone-insensitive internal NADH:ubiquinone oxidoreductase of mitochondria from *Saccharomyces cerevisiae*. Eur J Biochem 195:857–862

Mathiesen C, Hägerhall C (2003) The 'antiporter module' of respiratory chain Complex I includes the MrpC/NuoK subunit – a revision of the modular evolution scheme. FEBS Lett 549:7–13

Matsuoka M, Matsubara M, Inoue J, Kakehi M, Imanaka T (1994) Organization and transcription of the mitochondrial ATP synthase genes in the yeast *Yarrowia lipolytica*. Curr Genet 26:382–389

Medentsev AG, Akimenko VK (1999) Development and activation of cyanide-resistant respiration in the yeast *Yarrowia lipolytica*. Biochemistry 64:945–951

Melo AM, Bandeiras TM, Teixeira M (2004) New insights into type II NAD(P)H:quinone oxidoreductases. Microbiol Mol Biol Rev 68:603–616

Montet Y, Amara P, Volbeda A, Vernede X, Hatchikian EC, Field MJ, Frey M, Fontecilla-Camps JC (1997) Gas access to the active site of Ni-Fe hydrogenases probed by X-ray crystallography and molecular dynamics. Nat Struct Biol 4:523–526

Morgner N, Zickermann V, Kerscher S, Wittig I, Abdrakhmanova A, Barth HD, Brutschy B, Brandt U (2008) Subunit mass fingerprinting of mitochondrial complex I. Biochim Biophys Acta 1777:1384–1391

Nosek J, Fukuhara H (1994) NADH dehydrogenase subunit genes in the mitochondrial DNA of yeasts. J Bacteriol 176:5622–5630

Nübel E, Wittig I, Kerscher S, Brandt U, Schägger H (2009) Two-dimensional native electrophoretic analysis of respiratory supercomplexes from *Yarrowia lipolytica*. Proteomics 9(9): 2408–2418

Ogata K, Volini M (1990) Mitochondrial rhodanese – membrane-bound and complexed activity. J Biol Chem 265:8087–8093

Ogilvie I, Kennaway NG, Shoubridge EA (2005) A molecular chaperone for mitochondrial complex I assembly is mutated in a progressive encephalopathy. J Clin Invest 115:2784–2792

Ohnishi T (1998) Iron-sulfur clusters semiquinones in complex I. Biochim Biophys Acta 1364:186–206

Okun JG, Lümmen P, Brandt U (1999) Three classes of inhibitors share a common binding domain in mitochondrial complex I (NADH:ubiquinone oxidoreductase). J Biol Chem 274:2625–2630

Pagani S, Galante YM (1983) Interaction of rhodanese with mitochondrial NADH dehydrogenase. Biochim Biophys Acta 742:278–284

Radermacher M, Ruiz T, Clason T, Benjamin S, Brandt U, Zickermann V (2006) The three-dimensional structure of complex I from *Yarrowia lipolytica*: a highly dynamic enzyme. J Struct Biol 154:269–279

Rais I, Karas M, Schägger H (2004) Two-dimensional electrophoresis for the isolation of integral membrane proteins and mass spectrometric identification. Proteomics 4:2567–2571

Roessler MM, King MS, Robinson AJ, Armstrong FA, Harmer J, Hirst J (2010) Direct assignment of EPR spectra to structurally defined iron-sulfur clusters in complex I by double electron–electron resonance. Proc Natl Acad Sci USA 107:1930–1935

Sackmann U, Zensen R, Roehlen D, Jahnke U, Weiss H (1991) The acyl-carrier protein in *Neurospora crassa* mitochondria is a subunit of NADH: ubiquinone reductase (complex I). Eur J Biochem 200:463–469

Sazanov LA, Hinchliffe P (2006) Structure of the hydrophilic domain of respiratory complex I from *Thermus thermophilus*. Science 311:1430–1436

Sheftel AD, Stehling O, Pierik AJ, Netz DJA, Kerscher S, Elsasser HP, Wittig I, Balk J, Brandt U, Lill R (2009) Human Ind1, an iron-sulfur cluster assembly factor for respiratory complex I. Mol Cell Biol 29:6059–6073

Smeitink JAM, van den Heuvel LWPJ, Koopman WJH, Nijtmans LGJ, Ugalde C, Willems PHGM (2004) Cell biological consequences of mitochondrial NADH: ubiquinone oxidoreductase deficiency. Curr Neurovasc Res 1:29–40

Sokolova L, Wittig I, Barth HD, Schägger H, Brutschy B, Brandt U (2010) Laser-induced liquid bead ion desorption-MS of protein complexes from blue-native gels, a sensitive top-down proteomic approach. Proteomics 10:1401–1407

Tocilescu MA, Fendel U, Zwicker K, Kerscher S, Brandt U (2007) Exploring the ubiquinone binding cavity of respiratory complex I. J Biol Chem 282:29514–29520

Tocilescu MA, Fendel U, Zwicker K, Dröse S, Kerscher S, Brandt U (2010) The role of a conserved tyrosine in the 49-kDa subunit of complex I for ubiquinone binding and reduction. Biochim Biophys Acta 1797(6–7):625–632

Wada H, Shintani D, Ohlrogge J (1997) Why do mitochondria synthesize fatty acids? Evidence for involvement in lipoic acid production. Proc Natl Acad Sci USA 94:1591–1596

Wittig I, Braun HP, Schägger H (2006) Blue native PAGE. Nat Protoc 1:418–428

Yakovlev G, Reda T, Hirst J (2007) Reevaluating the relationship between EPR spectra and enzyme structure for the iron-sulfur clusters in NADH: quinone oxidoreductase. Proc Natl Acad Sci USA 104:12720–12725

Yano T, Yagi T (1999) H^+-translocating NADH-quinone oxidoreductase (NDH-1) of *Paracoccus denitrificans*. J Biol Chem 274:28606–28611

Zickermann V, Bostina M, Hunte C, Ruiz T, Radermacher M, Brandt U (2003) Functional implications from an unexpected position of the 49 kDa subunit of complex I. J Biol Chem 278:29072–29078

Zickermann V, Kerscher S, Zwicker K, Tocilescu MA, Radermacher M, Brandt U (2009) Architecture of complex I and its implications for electron transfer and proton pumping. Biochim Biophys Acta 1787:574–583

Zickermann V, Angerer H, Ding MG, Nübel E, Brandt U (2010) Small single transmembrane domain (STMD) proteins organize the hydrophobic subunits of large membrane protein complexes. FEBS Lett 584(12):2516–2525

Zwicker K, Galkin A, Dröse S, Grgic L, Kerscher S, Brandt U (2006) The redox-Bohr group associated with iron-sulfur cluster N2 of complex I. J Biol Chem 281:23013–23017

Transposable Elements and Their Activities in *Y. lipolytica*

Serge Casaregola and Gerold Barth

Abstract *Y. lipolytica* harbors an unusually diverse set of transposable elements among Saccharomycotina yeasts. Among them, members of both the families of transposons, retrotransposons as well as DNA transposons, are represented. Two of the LTR retrotransposons, Ylt1 and Tyl6, are members of the Ty3/*gypsy* group but have some uncommon features. Ylt1 is the largest hitherto detected fungal retrotransposon and is, in contrast to the other transposons, present in a high copy number of about 35 copies/haploid genome. Its proteins are encoded by a single ORF expressed under certain conditions resulting in transposition. Tyl6 is the only one retrotransposon among Saccharomycotina yeasts which displays a program −1 ribosomal frameshifting. The LINE-like element *Ylli* is also unique among Saccharomycotina yeasts and forms with the *C. albicans* counterparts *Zorro 1,2,3* a new family. It belongs to the L1 clade, which also contains the human LINEs. Like these element, the large majority of the *Ylli* copies are 5′ truncated, the characteristics of *Ylli* being that its copies are very short. The detected DNA transposon *Mutator* of *Y. lipolytica* (*Mutyl*) shares some similarities with several *MULE* elements found mainly in plants and in fungi. It is the first described DNA transposon in Saccharomycotina yeasts and like many of its counterparts, it may have invaded its host through horizontal transfer.

Contents

1 Introduction 58
2 Retrotransposons 59
  2.1 LTR Retrotransposons 59
  2.2 Non-LTR Retrotransposon 65

S. Casaregola (✉)
CIRM-Levures, Micalis Institute, INRA, AgroParisTech, 78850 Thiverval-Grignon, France
e-mail: Serge.Casaregola@grignon.inra.fr

G. Barth
Institut für Mikrobiologie, Technische Universität Dresden, 01062 Dresden, Germany
e-mail: gerold.barth@tu-dresden.de

G. Barth (ed.), *Yarrowia lipolytica*, Microbiology Monographs 24,
DOI 10.1007/978-3-642-38320-5_3, © Springer-Verlag Berlin Heidelberg 2013

3 DNA Transposons 69
3.1 The *Mutator*-Like DNA Transposon *Mutyl* 71
4 Conclusions 73
References 74

Abbreviations

LINE	Long interspersed elements
LTR	Long terminal repeat
RT	Reverse transcriptase
SINE	Short interspersed elements
TIR	Terminal inverted repeats
TSD	Target site duplication

1 Introduction

Transposable elements can be classified into two large families: retrotransposons (class I elements) and DNA transposons (class II elements). Class I elements can be subdivided into three main groups including long terminal repeat (LTR) retrotransposons, long interspersed elements (LINEs), and short interspersed elements (SINEs). LTR retrotransposons and LINEs are autonomous elements, which are able to catalyze their own transposition, whereas SINEs are nonautonomous elements, which rely on autonomous elements in order to transpose. Representatives of the latter group have never been found in Saccharomycotina yeasts. Transposable elements have been considered to be an important drive in eukaryotic genome evolution through the generation of allelic diversity, large-scale chromosomal rearrangements, gene transduction, duplication, and exon shuffling, domestication by exonization or exaptation of elements. The distribution of these elements in species can be informative on genome content, structure, and evolution. This is a key aspect for yeasts such as *Yarrowia lipolytica*, which are positioned in phylogenetic gaps and whose genome structure cannot be readily compared to that of other yeasts.

Y. lipolytica displays unusual features as described in a number of chapters of this monograph. The content in transposable elements of this yeast is also unusual, and hitherto known, only one yeast species, *Candida albicans*, presents an ever-larger diversity in terms of transposable elements. There exist two general ideas about transposable elements in Saccharomycotina yeasts. Retrotransposons are the only transposable elements present in yeasts and transposable elements are being lost in Saccharomycotina yeasts. The first one of these assertions is wrong: DNA transposons of the bacterial type are present in these organisms and the first DNA transposon to be described was in *Y. lipolytica* (Neuveglise et al. 2005). In addition

to the detection of DNA transposons in *C. albicans* mentioned by Goodwin and Poulter (2000), the clade of the protoploid *Saccharomycetaceae* was shown to harbor such transposons (Souciet et al. 2009). Similarly, non-LTR retrotransposons had not been found in yeast until a LINE element belonging to the same family as the LINE-1 of human was found in *Y. lipolytica* (Casaregola et al. 2000). In this respect, *Y. lipolytica* resemble euascomycetes. The second assertion is true: whatever the number of transposons found in a yeast species, its copy number per type of element is very low when compared to plants or mammals. Although *Y. lipolytica* carries a number of different transposable element families, their copy number is rather low, except a moderate number for Ylt1 (Schmid-Berger et al. 1994).

2 Retrotransposons

Retrotransposons are the most common transposons in yeasts. Nevertheless, there are very few types of retrotransposons present in the genome of *Y. lipolytica*. The presence of unusual retrotransposons and of some remnants indicates that the retrotransposon content of this yeast must have been one of the richest among Saccharomycotina yeasts.

2.1 LTR Retrotransposons

Long terminal repeat (LTR) retrotransposons transpose via a reverse transcription step, which is usually primed by annealing of a tRNA to the primer binding site, the 3′ end of the transposon RNA, followed by reverse synthesis of the first cDNA strand and then synthesis of the second DNA strand. This process occurs in the cytoplasm and the transposon is subsequently transferred to the nucleus, in which integration occurs by a mechanism similar to what is seen in class II DNA elements, with a nuclease making specific nicks at the integration site to catalyze the process (Havecker et al. 2004 and references therein). Five families of transposons have been described in *Saccharomyces cerevisiae:* the Ty1, Ty2, Ty4, and Ty5 elements, which belong to the Ty1/*copia* (Pseudovirus) type, and the Ty3 element, which belongs to the Ty3/*gypsy* (Metavirus) type. Other elements such as Tca2, defining another family in *C. albicans*, have also been described. In Ty elements, two LTRs, up to over 700 bp in size in *Kazachstania exigua* (*Saccharomyces exiguus*) (Bon et al. 2000), surround two genes *gag* and *pol* organized differently according to the element; they can be separated by a +1 frameshift in Ty1, Ty2, Ty3, and Ty4, form a single ORF in Ty5, or be separated by a stop codon in Tca2. These organizations reflect the mechanisms responsible for the regulation of the expression of the two genes and in particular *pol*, in order to prevent too high levels of transposition. It must be noticed that homologous recombination between the two LTRs of a transposon results in “popping out” of the element, leaving as a scar a solo LTR. Most of

the LTR retrotransposon copies are detected as solo LTRs (Kim et al. 1998). The number of such LTR retrotransposons varies greatly among sequenced genomes. Whereas there is a complete lack of LTR retrotransposon in the Saccharomycotina yeast *Pichia sorbitophila* (Neuveglise et al. 2002; Louis et al. 2012) and one single full-length copy of a Ty element in another Saccharomycotina *Candida glabrata*, a total of 51 Ty elements were found in the first sequenced *S. cerevisiae* genome (Kim et al. 1998), with this number also varying between strains. Ty elements tend to be clustered around tRNA genes and genes transcribed by RNA polymerase III (Pol III), this preference most likely being mediated by interactions between the Pol III complex and the integration complex (Hani and Feldmann 1998).

2.1.1 Ylt1

The first retrotransposon in *Y. lipolytica* was detected in the genome of the strain B204-12C-20 in 5′ upstream region of the open reading of the gene *GPR1* and was called *Y*arrowia *l*ipolytica *t*ransposon *1* (Ylt1) (Schmid-Berger et al. 1994). Analysis of the nucleotide sequence of Ylt1 has shown that Ylt1 has many features in common with known yeast retrotransposons. On the other hand, some of its special properties suggest an isolated position for Ylt1 among yeast retroelements.

Ylt1 has a length of 9,453 bp and is the largest fungal LTR retrotransposon reported so far (Schmid-Berger et al. 1994). The LTRs (called zeta) of Ylt1 are 714 bp long and the elements are bounded in the genome by 4-bp TGDs. Analysis of LTR containing genomic DNA fragments of strain B204-12C-20 and of all 27 Ylt1 insertions in the genome of *Y. lipolytica* strain CLIB 122 revealed the consensus sequence C/T-A-T-G/A of TGD of Ylt1 (Schmid-Berger et al. 1994). LTRs of Ylt1 are flanked by the short inverted repeats TGT…ACA which are highly conserved in Ylt1 elements (Schmid-Berger et al. 1994).

The LTR sequence contains putative transcription regulatory elements. Putative TATA and CAAT boxes were identified at the positions 214 and 190, respectively. A putative polyadenylation signal AATAAA was found at the position 496 (Schmid-Berger et al. 1994).

A polypurine tract of Ylt1 (GGGCAGG) is located immediately upstream of right LTR and a predicted primer binding site (PBS) of Ylt1 is located 2 bp downstream from the 5′LTR. The predicted PBS begins with the trinucleotide TGG, and it is suggested that reverse transcription of Ylt1 is primed by a host-encoded $tRNA^{Ala(AGC)}$ molecule, because 14 bases from the acceptor stem of the identified *Y. lipolytica* $tRNA^{Ala(AGC)}$ are complementary to the predicted PBS of Ylt1 (Kovalchuk 2005).

The retrotransposon Ylt1 contains only a single large uninterrupted ORF encoding both Gag and Pol proteins. The deduced translational product of this ORF is a polyprotein consisting of 2621 amino acid residues. Its sequence includes the regions similar to structural protein (Gag), protease (PR), reverse transcriptase (RT), RNaseH (RH), and integrase (IN) of other retroelements. The arrangement of coding regions within the ORF of Ylt1 (GAG—PR—RT—RH—IN) and the amino

acid sequence of the reverse transcriptase place Ylt1 within the Ty3/*gypsy* group of LTR retrotransposons (Kovalchuk 2005).

Gag proteins encoded by different retrotransposons show a significant sequence heterogeneity. Their only well-conserved part is an RNA-binding Zn finger motif $CX_2CX_4HX_4C$ (Covey 1986). N-terminal part of the deduced polyprotein of Ylt1 includes such a Zn finger motif of sequence CVFCGSTAHALVNC at position 2160–2199. The other parts of the putative Gag protein show no significant homologies with known proteins.

A protease homologous region follows the putative Gag protein. A putative aspartic protease motif LFDSGS was found at the position 3977. It agrees well with the consensus protease sequence (hydrophobic residue)$_2$-D-T/S-G-A/S, harboring an absolutely invariant aspartate residue, which is typical for the majority of described retroelements (Dunn et al. 2002).

The conserved motif for reverse transcriptase, the so-called YXDD box, is located at position 6032 (AVVFLDDIGV). The tyrosine residue in the conserved motif is substituted by phenylalanine in Ylt1. The conserved RNaseH motif TDAS was found further downstream at position 6549 of Ylt1. All three domains described for an integrase protein are well conserved in Ylt1. A N-terminal zinc finger domain with the HHCC motif, a core domain containing the catalytic D,D_{35}-E motif, and a C-terminal GPY/F module (Malik and Eickbush 1999) are also present.

The copy numbers of Ylt1 were calculated in the strains B204-12A-213 and B204-12C-20, from the German strain lineage, and YB423-12 and CXAU1 from the American lineage by quantification of Southern blot hybridization of their genomic DNA. In all strains of both lineages, about 30–35 entire copies of this element and of about 100 copies of solo LTRs per haploid genome were calculated. The analysis of the genomic sequence of the strain CLIB 122 of the French lineage revealed a copy number of 27 per haploid genome. These data make Ylt1 the element with the highest copy number among all transposable elements in *Y. lipolytica*.

The wild-type strains H222 and W29 which are ancestors of the German and French strain lineages, respectively, as well as strain Po1d are devoid of Ylt1. The origin of Ylt1 is the diploid strain YB423 and its haploid descendants YB423-3 and YB423-12 which were used as mating partners for construction of German and French strain lineages.

A phylogenetic analysis based on analysis of amino acid domains of 178 residues conserved in all reverse transcriptases revealed grouped Ylt1 together with the retrotransposons Tca3 and Tcd3 from the genomes of *C. albicans* and *Candida dubliniensis*, correspondingly, and with the elements Tor4a and Tor4b from the genome of the tunicate *Oikopleura dioica* (Kovalchuk 2005).

Ylt1 is an active retroelement as shown by Southern blot hybridization, Western blot hybridization of HA tagged Gag protein, as well as by β-galactosidase activity expressed by *lacZ* fused to the Ylt1 promoter (Senam 2004). Southern blot hybridization experiments have shown that new copies of Ylt1 can be detected in genomic DNA after strains were cultivated in acetate containing media. The Gag

protein tagged with a 3xHA epitope could be detected during growth of Ylt1 containing (B204-12C) and Ylt1 less (Po1d) strains transformed by a gag-HA gene containing plasmid. The Gag-HA protein is produced during growth of cells in glucose, acetate, or ethanol containing media; however, it is strongly reduced after entry of cells into the stationary phase. These data were confirmed by measurements of β-galactosidase activity in cells transformed by a plasmid harboring the Ylt1 promoter fused with the *lacZ* gene (Senam 2004). These data suggest that Ylt1 may be active in growing cells but may be silent in resting cells.

2.1.2 Tyl6

Careful analysis of the complete genome of CLIB 122 (Dujon et al. 2004) using tblastn and various yeast retrotransposon protein sequences revealed the presence of another retrotransposon with LTR, which resembled Ty3/*gypsy* elements. Surprisingly, only one copy of this transposon was found and this copy, even more surprisingly, was of full length. The entire element is 5,106 bp long and it is flanked by two identical 276-bp long repeated sequences themselves bordered by 5-bp inverted repeat characteristics of LTR retrotransposons (Kovalchuk et al. 2005). A 4-bp long TSD was also found, indicating that the presence of this transposon was not due to genomic rearrangement but that it had been active. Although this transposon is related to Ty3 element, it clearly differed from this type of elements and was therefore called Tyl6 (for Transposon of *Y. lipolytica* 6), in agreement with the proposed nomenclature for yeast LTR retrotransposons (Neuveglise et al. 2002).

The organization of the coding sequence of Tyl6 is typical for LTR retrotransposons. There are two ORFs separated by a frameshift (Kovalchuk et al. 2005). The putative Gag protein has some limited similarity to other yeast Gag and contains a conserved region of 280 bp and a Zn finger at the end of the protein. The Pol protein is more conserved and contains the expected protease/RT/RNaseH/integrase domains. Analysis of the sequence of Tyl6 revealed that this is the first Saccharomycotina yeast element to use a programmed −1 ribosomal frameshifting, whereas this is common in filamentous fungi (Gao et al. 2003) and in basidiomycetes (Goodwin and Poulter 2001). In most cases, retrotransposons use a programmed +1 ribosomal frameshifting event to lower the expression of *pol* and retrotransposition (Turkel et al. 1997). In yeasts, they use either a +1 frameshifting (like Ty1 to Ty4 in *S. cerevisiae*) or a stop codon readthrough (like Tdh2 in *D. hansenii* and Tca2 in *C. albicans*). In some instances, *gag* and *pol* are fused to form a single ORF (Ylt1 in *Y. lipolytica* and Ty5 in *S. cerevisiae*), but whereas the resulting Ty5 element is located in silenced parts of the chromosomes like the telomeres, the Ylt1 element is active and dispersed throughout the genome. The Tyl6 element in *Y. lipolytica* is an exception because the type of elements which uses −1 ribosomal frameshifting seems to be lost from the other Saccharomycotina yeasts analyzed to date.

Using hybridization to various part of Tyl6, the presence of this element was shown to be present in the American lineage and in strains resulting from the breeding programs between the American lineage strains and strains from the German or the French lineages. On the other hand, Tyl6 was not found in the strains from the German or the French lineage (Barth and Gaillardin 1996, 1997; Kovalchuk et al. 2005). This is surprising since LTR retrotransposons have a vertical propagation. A simple explanation for this observation is that the strains from the American lineage are the only ones which have retained Tyl6. Another possibility, which is at first sight less plausible, is that the presence of Tyl6 in *Y. lipolytica* is due to a horizontal transfer event from a donor species to an American *Y. lipolytica* strain. The lack of solo LTRs, which are usually more frequent than their cognate full-length elements in a genome, and the absence of Tyl6 in a number of strains from various origins support this observation. In addition, all the analyzed strains which carry Tyl6 have only one element, which seems to be located at a conserved site. The latter observation is not surprising since this element was probably inherited from the YB423-12 strain during the various inbreeding programs (Barth and Gaillardin 1996, 1997). Finally, the complete conservation of both LTRs and the TSD strongly suggests that this element has transposed recently and therefore supports the horizontal transfer hypothesis. Nevertheless, Tyl6 has not transposed since the integration of the single Tyl6 copy, unlike other *Y. lipolytica* transposons (Casaregola et al. 2002; Neuveglise et al. 2002), which transposed after the divergence of the studied *Y. lipolytica* lineages.

Tyl6 possesses a primer binding site complementary to the 3′ acceptor stem of the initiator methionine tRNA ($tRNA^{iMet}$) from its host species as do most of the *S. cerevisiae* retrotransposons. In Tyl6, the primer binding site is located 2 bp downstream of the 5′LTR. Tyl6 is integrated 15 bp upstream of a $tRNA^{iMet}$ gene (Kovalchuk et al. 2005). Although there is only one copy of Tyl6, this element therefore resembles Ty3, which has been shown to integrate preferentially in a narrow window of 13–19 bp upstream of tRNA genes (Chalker and Sandmeyer 1992, 1993). Phylogenetic analysis based on the RT domain of Pol places unambiguously Tyl6 within the sensu stricto Ty3/*gypsy* group. The closest relative of Tyl6 is the Ty3 element of *K. exigua*, Tse3. Tyl6 is also related to the *del1*-like elements of flowering plants. This phylogenetic analysis has also demonstrated that the two Ty3-like elements of *Y. lipolytica*, Tyl6 and Ylt1, are not closely related (Kovalchuk et al. 2005). A more recent phylogenetic analysis with only the Ty3-like elements of Saccharomycotina yeasts and two Ty3/*gypsy*, Cn1 of *Cryptococcus neoformans*, and an element of *Magnaporthe grisea* was performed (S. Casaregola, unpublished). The phylogenetic tree in Fig. 1 indicates that, although close to the clade of the Saccharomycotina yeasts including three elements of *K. exigua*, *C. glabrata*, *Candida tropicalis*, and *Clavispora lusitaniae*, Tyl6 is distinct from this clade and from the clade of the filamentous fungi and the basidiomycetes. This is congruent with the phylogenetic tree of the Saccharomycotina, in which *Y. lipolytica* has a basal position (Dujon 2006).

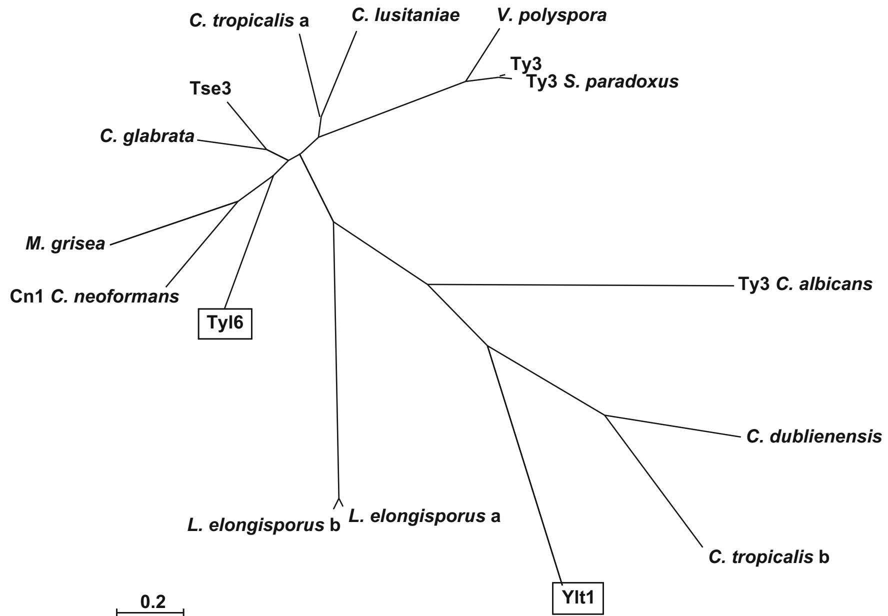


Fig. 1 Unrooted phylogenetic tree of the RT domain of the Ty3/gypsy elements in Saccharomycotina yeasts. Multiple alignments were performed using MUSCLE and phylogenetic tree was generated with PhyML using the default settings implemented in Phylogeny.fr (http://www.phylogeny.fr). The various sequences used were Ty3 (*C. albicans*), AAM439; *C. tropicalis* b, AAM439; *Candida dubliniensis*, XP0024; Ylt1 (*Y. lipolytica*), XP504795; *Lodderomyces elongisporus*, XP0015; *L. elongisporus*, XP00152; Tyl6 (*Y. lipolytica*), XP00214; *C. glabrata*, AAT76628; Tse3 (*K. exigua*), CAD29542; *C. tropicalis* a XP0025; *C. lusitaniae* b, XP00261; *Vanderwaltozyma polyspora*, XP00164; Ty3 (*S. cerevisiae*), Q99315.3; Ty3 (*Saccharomyces paradoxus*), AAO273. The Cn1 *C. neoformans* (XP571377) and a *M. grisea* element (XP001414325) were also used. The *Y. lipolytica* elements are *boxed*

2.1.3 Other Retroelements

Two sequencing projects were performed on *Y. lipolytica*. The first on the strain W29 yielded 4,940 random sequence tags (RSTs), totaling 4.9 Mb (Casaregola et al. 2000). The second project led to the complete sequence of the strain CLIB 122 (Dujon et al. 2004). During the first exploration of the genome of this yeast, a repeated sequence of 273 bp was found to be present on 18 RSTs. It contained the canonic TGTTG repeat at the 5′ end as well as CAATA at the 3′ end, indicating that it is a putative LTR. In addition, five of the 18 LTR sequences were found to be surrounded by 5-bp repeats, indicative of a duplication of the target site (Casaregola et al. 2000). This LTR, coined LTRyl1, was found, as expected, at 149 bp of a tRNA gene on one RST (Neuveglise et al. 2002), further suggesting it is an LTR of a Ty by

Table 1 Number of LTRyl1 in the strain CLIB 122

LTR	Chromosome						Total
	A	B	C	D	E	F	
Full length	5	3	2	4	1	4	19
Degenerate[a]	5	4	7	9	3	4	34
Total	10	7	9	15	4	8	49

[a]Over 80-bp long elements

inserting at the vicinity of tRNA. One to five full-length LTRyl1s are present per chromosome. A total of 49 elements including over 80-bp long remnants were detected in the complete genome of CLIB 122 (Table 1).

In strain H222, a wild-type isolate and an ancestor of the German strain lineage, a sequence of 273 bp with very high homology to the LTRyl1 sequence was detected in the 5′ upstream region of a putative *ScSTE6* (YALI0E05973g) orthologous gene. This LTR contains the canonic repeat TGTTC at the 5′ as well as the CAATA at the 3′ end and is surrounded by a target site duplication repeat of CATGT also very similar to LTRyl1 in strains W29 and CLIB 122 (Werner and Barth, unpublished). However, LTRyl1 is not present in the 5′ upstream region of the YALI0E05973 gene in both strains (Werner 2009) and vice versa no tRNA gene is neighboring this LTR in strain H222.

The presence of LTRyl1 in strains W29 and H222 is the first example of the detection of the same LTR retrotransposon in two independent wild-type isolates of *Y. lipolytica* and may be a hint for the vertical propagation of this retrotransposon, at least. Unfortunately, the complete genomic sequence of strain H222 is not available to check whether an entire copy of this retrotransposon is present as well as to estimate the copy number and the specificity of integration sites.

2.2 Non-LTR Retrotransposon

LINEs are non-LTR class I elements whose best characterized member is the LINE-1 (L1) mammalian retrotransposon. This 6- to 8-kb element contains two ORFs (ORF1 and ORF2) that are co-transcribed from the same promoter but can be processed into several distinct messenger RNAs. ORF1 encodes a trimeric nucleic acid chaperone protein that binds to L1 mRNA to form ribonucleoprotein complexes, considered to be transposition intermediates. ORF2 contains endonuclease and reverse transcriptase activities, both required for retrotransposition in wild-type cell lines. L1 elements transpose in three steps: (1) formation of a nick in double-stranded DNA at AT-rich sequences, mediated by the endonuclease activity of ORF2; (2) annealing of the L1 mRNA poly(A) tail to the 5′ poly(T) tail at the nick and cDNA synthesis by the reverse transcriptase activity of ORF2; and (3) degradation of the L1 mRNA and second-strand synthesis followed by ligation (Babushok and Kazazian 2007; Belancio et al. 2008).

2.2.1 *Ylli*, the LINE-1 of *Y. lipolytica*

LINEs are present in all eukaryotes and can constitute an important part of the genome like in human, for example, in which over 800,000 insertions make up 35 % of the genome (Lander et al. 2001). However, LINEs had not been found in *S. cerevisiae*, nor in any other Saccharomycotina yeast studied until the first coverage of *Y. lipolytica* W29 genome, which revealed the presence of a repeated sequence presenting sequence similarity with fungal reverse transcriptases (RT) (Casaregola et al. 2000). Extension of the sequencing on each side of the zone matching the fungal RT led to a sequence of 6,942 bp, which contains several features that are shared by LINEs:

- The presence of two ORFs in the same polarity (separated by two nucleotides in *Y. lipolytica*)
- The first ORF with no sequence similarity with known protein, as expected from LINE nucleic acid binding ORF1
- The presence of a Cys-rich motif in the first ORF
- The second ORF carrying endonuclease and reverse transcriptase domains in addition to a Cys-rich motif
- A poly-A tail at the 3′ end of the element
- Diverging 5′ ends among the RSTs reminiscent of the 5′ truncated forms of LINEs

In 2001, LINE-like elements were described from the analysis of the complete genome of the pathogen diploid *C. albicans* strain SC 5314. A new family with three members called *Zorro*-1, *Zorro*-2, and *Zorro*-3 was defined in which only one copy of *Zorro*-1 and one copy of *Zorro*-3 are of full length. A total of 14 elements were found, 12 being remnants. The *Zorros* resemble L1 by the presence of two ORFs, the phylogenetic proximity of ORF2, and seem to possess most the characteristics of LINEs (Goodwin et al. 2001). However, only *Zorro*-3 seems to integrate at random, whereas the two others were found in subtelomeric regions. It is not rare for LINEs to have insertion specificity, since, for instance, R1Bm in *Bombyx mori* inserts specifically in rDNA repeats (Xiong and Eickbush 1988) and *Zepp* in *Chlorella vulgaris* in telomeric regions (Higashiyama et al. 1997).

The *Y. lipolytica* element was called *Ylli* for *Y. lipolytica* LINE carries two ORFs of 714 and 1300 amino acids. ORF1 turned out to be one of the largest ORF1 found in LINEs, except that of the 775 amino acid long *Xenopus laevi Tx1*. Upstream of ORF1 and downstream of ORF2 sequence likely to correspond to a 5′UTR and a 3′UTR, respectively. A poly-A tail is found downstream the 3′UTR. It is accepted that the low processivity of the reverse transcriptase creates 5′ truncates; a little minority of LINEs are full length (Myers et al. 2002; Pavlicek et al. 2002) and even less are active (Sassaman et al. 1997). The common occurrence of *Ylli* 5′ truncation was shown by hybridization with a 5′ probe and a 3′ probe to genomic DNA. This led to a larger number of signals with the 3′ probe than with the 5′ probe. This was later confirmed by the analysis of the complete genome (see below) and in the case

of *Ylli*, this phenomenon was shown to be amplified (see Sect. 2.2.2). It was therefore difficult to define the start of the element and even the start of ORF1. The prediction for the start of ORF1 made by Casaregola et al. (2002) was recently confirmed by analyzing all the elements present in the complete genome of the strain CLIB 122. Although an ATG upstream of the chosen ATG can be found in some of the copies of *Ylli* ORF1, including that of the published sequence (Casaregola et al. 2002), no consensus was found for the presence of this ATG in the majority of the largest *Ylli*. Like most nucleic acid binding proteins, ORF1 of LINEs, also considered as *gag* of retroviruses, have a poorly conserved sequence. This is the case of *Ylli* ORF1 for which limited sequence similarity was found with various LINE ORF1s. Nevertheless, one cysteine-rich domain was found in *Ylli* ORF1, as found in most of the other LINEs.

Ylli ORF2 on the other hand displayed stronger sequence similarity with various reverse transcriptases including fungal, insect, and human L1 proteins. Careful analysis revealed that a region bearing all the conserved domains of an apurinic endonuclease was also present at the N-terminus. In addition, a cysteine-rich motif was also found near the C-terminus.

A phylogenetic analysis of the 11 motifs covering the reverse transcriptase domain of ORF2 of 12 clades of LINEs defined by Malik et al. (1999) clearly placed *Ylli* in the L1 clade with human L1Hs and mouse LIMm. *Ylli* constituted a monophyletic group with the *Zorros* of *C. albicans* and therefore form with the *C. albicans* counterparts a new family. Interestingly, all the fungal elements grouped with *Tad*-1 from *Neurospora crassa* and are phylogenetically diverged from the yeast elements. LINEs, like all retrotransposons, propagate essentially vertically, and, as expected, *Ylli* is present in all the strains tested, including the German lineage represented by H222, the French lineage by W29, and the American lineage by CX161-1B.

Finally, the highly conserved sequence of the elements found in *Y. lipolytica* indicates that *Ylli* transposed recently. This is in agreement with the recent demonstration of the retrotransposition of Zorro-3 in *C. albicans* (Goodwin et al. 2007) and in *S. cerevisiae* (Dong et al. 2009). Despite of the evolutive distance between these two species and the different genetic code used in the species of the CTG clade like *C. albicans*, *Zorro*-3 can transpose in *S. cerevisiae*, indicating that the host machinery required for non-LTR retrotransposition has been conserved in a not so closely related species.

2.2.2 *Ylli* in the Complete Genome Sequence of CLIB 122

Using the published sequence of *Ylli*, a thorough search in the complete genome of CLIB 122 was undertaken (Ozanne and Casaregola, unpublished results). One of the striking features of *Ylli* is the presence of short elements (23–134 bp long) matching the end of *Ylli*, including the poly-A tail (Casaregola et al. 2002). These short repeats are reminiscent of SINEs that are found in high copy number in mammals, but not in yeasts. SINEs derive from tRNA promoters associated with

the 3′ end of non-LTR retrotransposons and use the latter element machinery to propagate. In addition to the fact that *Ylli* short elements are much smaller than the SINEs, no conserved boxes A and B, present in the promoters of SINEs, could be found upstream the short *Ylli*s. It was in fact proposed that these short *Ylli*s were the result of abortive reverse transcription events due to a dyad symmetry with an internal energy of 3.8 kcal located 7 bp upstream from the poly-A tail which would block the transcriptase (Casaregola et al. 2002). We predicted that given the close location of the dyad symmetry to the end of the element, very short elements should be found in the *Y. lipolytica* genome. We set up a search using blastn to detect element with a size as low as 8 bp and the R'MES statistical method (Hoebeke and Schbath 2006) to estimate if shorter sequences with identity to *Ylli* were overrepresented. The threshold chosen was 7 bp long associated to a 6-bp long poly-A tail. For very short elements, we also used the presence of a target site duplication to ensure these elements were *Ylli* 5′ truncates. The only blastn search led to the detection of 153 elements from 8 to 6,549 bp with over 86 % sequence similarity. The use of the three methods led to the identification of 201 elements in the genome of CLIB 122 (Fig. 2). Except for the very large elements, all are 5′ truncated and 83 % are less than 500 bp, 67 % being less than 100 bp. Among the latter, most stop at the level of the dyad symmetry pointing out towards a role of this structure in the generation of very small *Ylli* 5′ truncates (Fig. 3). Like most transposons, LINEs create a target site duplication (TSD), which is a result from the staggered cleavage mediated by integrases and transposases, here by endonuclease and reverse transcriptase. We also studied the TSD generated by *Ylli* insertions in the genome of CLIB 122. The size of these duplications varies between 4 and 11 bp. TSDs were found in 94 of the 201 elements. For 16 of them, duplications are not perfect but rather they have a high rate of AT.

A total of ten elements were found to be potentially full length. Three of these elements have a frameshift followed by a stop codon in the 3′ of ORF1. These mutations could also be due to sequencing errors. Despite the availability of the complete genome, it is still difficult to define the 5′end of *Ylli*. It seems to be associated to A-rich region found in the elements shown in Fig. 4. Like in human, a bimodal distribution of the size of L1Hs and of *Ylli* is observed (not shown); in particular, the observed number of full-length elements is not expected if the low processivity of the reverse transcriptase associated to obstacles must be the only cause of the generation of 5′ truncates. No explanation exists for this observation.

Non-LTR retrotransposons have been described as mutagens in human by inserting into ORF (Babushok and Kazazian 2007). Several instances of "exonization" of active LINE elements, leading to the creation of new genes, have been described (see for review Burns and Boeke 2008). We surveyed the occurrence of insertion close to the start of the ORFs of *Y. lipolytica*. No *Ylli* was found inserted in an ORF, which is not surprising considering the large intergenic regions of this yeast. Although we did not find any insertion bias of *Ylli*, this may play in *Y. lipolytica*, like L1 in mammalian genomes, a role in gene expression, since we found 61 elements located less than 500 bp upstream the start of genes. An interesting case is the insertion located around 200 bp upstream of the pseudogene

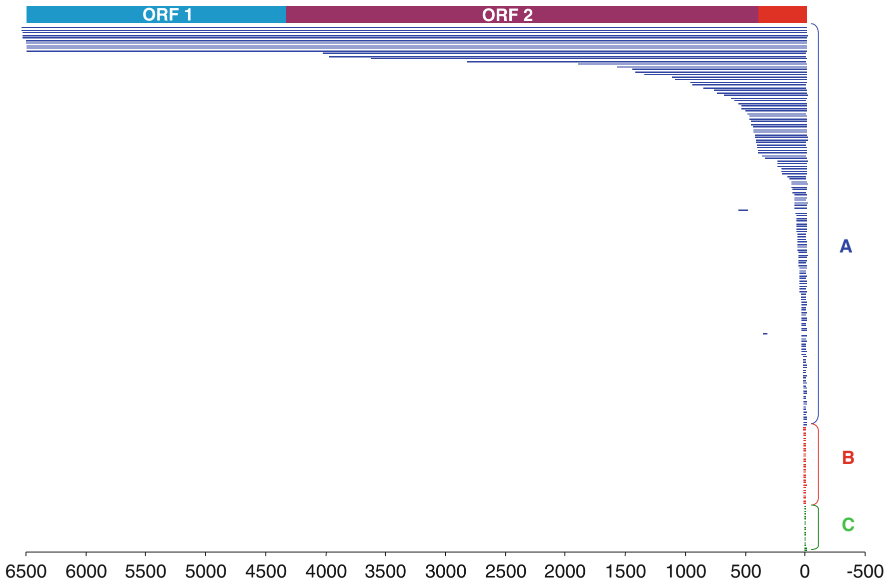


Fig. 2 Schematic representation of the CLIB 122 *Ylli* insertions in relation with a full-length *Ylli*. Each element including its poly-A tail detected in the genome is drawn to scale. *Bottom*: bp. The zero is the 3′ end of the 3′UTR. *Blue color*: elements detected using blastn. *Red color*: elements of at least 7 nucleotides detected using R'MES. *Green color*: elements of less than 7 nucleotides detected by eye and having a TSD (Ozanne and Casaregola, unpublished)

YALI0A08382g. This gene has a paralogue on the chromosome E (YALI0E18029g) and the insertion may be the cause of the pseudogenization of YALI0A08382g. In this connection, another interesting example was detected in strain H222 in which a highly homologous *Ylli* sequence is located about 200 bp upstream of the start of gene YALI0E05973g and 20 bp downstream of the above-described LTRyl1 retrotransposon (Werner 2009). This element, which is very homologous to *Ylli* of strains W29 and CLIB 122, is about 400 bp long, 5′ truncated, contains a poly-A tail of eight bp, and has inverted repeats potentially forming a stem loop structure. It is not present in the 5′ upstream regions of YALI0E05973g in strains W29 and CLIB 122. The gene YALI0E05973g has two paralogues in the genome of *Y. lipolytica*, at least, and seems, similar to YALI0A08382g, to be silent as shown by microarray analysis (Petzsch, unpublished).

3 DNA Transposons

Class II elements are mobile DNA elements that utilize a transposase and create single- or double-strand DNA breaks to transpose. They can be classified into three major subclasses: (1) elements that excise as double-stranded DNA and transpose

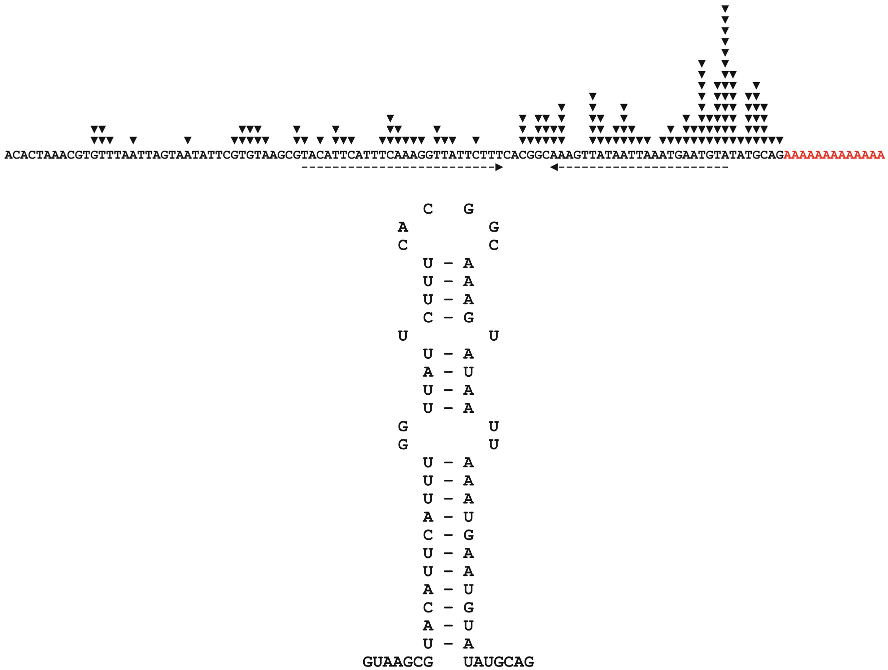


Fig. 3 Sequence of *Ylli* 3′UTR. *Top*: *Dashed underlining* indicates the dyad symmetry. The various locations of the start of the short *Ylli*s are indicated with *filled inverted triangles*. *Bottom*: the potential stem loop structure deduced from the dyad symmetry in the 3′UTR as indicated in *top* of the figure is depicted (Casaregola et al, 2002)

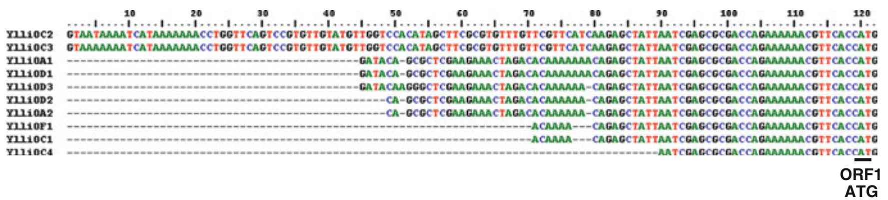


Fig. 4 Sequence alignment of the 5′UTR of putative full-length Yllis. The ORF1 ATG is indicated (Ozanne and Casaregola, unpublished)

by a classical "cut-and-paste" mechanism, such as Drosophila P elements or *hAT*; (2) elements that utilize a rolling-circle mechanism, such as *Helitrons*; and (3) elements that probably utilize a self-encoded DNA polymerase but whose transposition mechanism is not well understood, such as *Mavericks*. Based on transposase sequence similarities and phylogenetic analyses, they can be classified into ten different families (Feschotte and Pritham 2007). Their number and proportion,

compared to those of retrotransposons, are highly variable among eukaryotic genomes.

3.1 The **Mutator***-Like DNA Transposon* **Mutyl**

From the study of *S. cerevisiae* and closely related species, DNA transposons seem to be lacking in yeasts. Recent comparative genomics analysis revealed that DNA transposons have colonized Saccharomycotina yeasts and may continue to do so (Neuveglise et al. 2005; Souciet et al. 2009).

3.1.1 Description of *Mutyl*

Complete genomes are very useful since they may reveal sequences that could be overlooked, because they do not present enough overall similarity with known elements or known proteins. This is the case of *Mutyl*, a DNA transposon, which carries two ORFs, which in turn were found, repeated in the genome of CLIB 122 (Neuveglise et al. 2005).

This 7,413-bp element shares limited sequence similarity with the *Hop* element of *Fusarium oxysporum*, several putative far-red impaired response proteins of *Oryza sativa*, and a *Mutator*-related protein in *Arabidopsis thaliana*. It was called *Mutyl* for *Mutator* of *Y. lipolytica*, because of the similarity with *Mutator*-like elements, the so-called MULEs, a family of transposons found mainly in plants and known to create mutations upon insertion (Lisch 2002). *Mutyl* contains two ORFs in the same orientation putatively encoding proteins of 459 and 1178 amino acids. Interestingly, the most studied MULE, MuDR of *Zea mays*, has a very different structure, its two ORFs being transcribed in a divergent orientation and their promoters being located in the terminal inverted repeats which bound the element. *Mutyl* rather resembles DNA transposons found in bacteria and in fungi.

The strongest evidence for the belonging of *Mutyl* to the MULE family comes from the presence in the ORF *mutA* of a signature sequence of around 90 amino acids found in bacterial transposases and also found in MuDR and in the transposons of this family. In addition to the fact that this type of sequences does not yield robust phylogeny, phylogenetic analyses of MULEs are hampered by the fact that these elements are often acquired through horizontal transfer and phylogeny of the transposons and of their cognate species are not congruent. It is indeed the case for *Mutyl* that phylogenetic analysis placed between plant MULEs. The other ORF *mutB* did not show any significant sequence similarity with known proteins. Two zinc finger motifs were found in MutB that were separated by 112 amino acids. A zinc finger in the C-terminal part of the transposase MutA was also found. This motif is only present in eukaryotic transposases.

DNA transposons are also defined by the presence of repeats at each end of the element. In the case of MULE elements, the first described element MuDR contains

TIR of around 100 bp, but MULEs that were discovered since do not all carry TIRs. Some possess shorts or imperfect repeats. *Mutyl* carries a 22-bp long stretch, which is imperfectly repeated, present at 6 bp of the element extremities (CACTTCG/AG/AGTCTACAG/TACCTTA). Three bases were shown to be variable between the 5′ TIR and the 3′ TIR. Unlike MuDR for which it was shown that promoters were located within TIRs, the promoter of mutA is located within a 723-bp long region between the 3′ end of *mutB* and the 5′ end of *mutA*. The start of *mutB* is located at base 1308 and it is very likely that the promoter of *mutB* gene is located within this region and not in the TIR.

There exist five copies of *Mutyl* in the genome of CLIB 122, four being full length, whose sequences are strongly conserved. One copy has a transversion resulting in a change from a Thr to a Pro and another copy was subjected to a deletion which removed the first eight bp of the element. No TSD was found for this element, indicating the deletion encompassed more than the *Mutyl* extremity, whereas a TSD was detected for the four other elements, exact duplication of 9 bp for two elements and of 10 bp for the two others. The strong sequence conservation of *Mutyl* indicates that the element was active recently, but the existence of two sequences with similarity to *Mutyl* (80 % identity over 550 bp and 88 % over 3,352 bp) also suggests that the colonization of *Y. lipolytica* by *Mutyl* is not recent. The low copy number of *Mutyl* and the large intergenic regions of *Y. lipolytica* could explain why *Mutyl* has not inserted into an ORF. At the same time, no evidence exists for *Mutyl* having mutator activity, nor insertion specificity.

3.1.2 Peculiarities of *Mutyl*

The uniqueness of *Mutyl* among Saccharomycotina yeasts and the fact that MULEs are known to evolve horizontally hinted at the presence of *Mutyl* resulting from a horizontal transfer. The presence of *Mutyl* in the American lineage and its absence in the French and the German lineages support this view. Horizontal transfer was also proposed to explain the presence of hAT-like DNA transposon families only in the *Kluyveromyces*/*Lachancea* clade of Saccharomycotina yeasts (Souciet et al. 2009). The horizontal transfer hypothesis was also supported by the analysis of strains resulting from crosses between the various known lineages. The *Mutyl* elements present in these strains could be traced back to the American lineage, indicating that they were brought by the American strains in these crosses. Interestingly, an element was found in a haploid strain, whose CLIB 122 is derived, CBS6124-2, which itself was derived from the only *Y. lipolytica* diploid isolated to date CBS 6124. It was shown that this element was not present in this initial diploid CBS 6124. This showed that a transposition event had occurred, very likely during meiosis, which generated CBS6124-2. The region carrying this element undertook a rearrangement afterwards. De novo transposition at meiosis was already observed for Tys in *S. cerevisiae* (Neuveglise et al. 2000; Ribeiro-dos-Santos et al. 1997). This transposition event thus showed that *Mutyl* was active very recently. In agreement with this observation, we observed that one of the ORFs of

Mutyl, *mutB*, was highly expressed under normal growth conditions. On the other hand, *mutA* was not expressed, suggesting that MutB protein may act as a negative regulator of the expression of *mutA*. This is an interesting suggestion, since the non-LTR retrotransposon *Ylli* seemed to have devised an elegant way to reduce the effect of transposition through abortive reverse transcription of the element. Here, the expression of the MutA transposase gene may be turned off to avoid deleterious effect of transposition of this DNA transposon. The MULE family comprises so many different types of transposon from *Z. mays MuDR* to *F. oxysporum Hop* that the putative regulation system of *Mutyl* may not be straightforwardly comparable to any existing member of the MULE family. The situation of MULEs in plants is different, since it was shown that transcription was initiated in the TIR and processed in a convergent manner, indicating that both genes MuDRA and MuDRB were coordinately regulated (Raizada et al. 2001). In addition, it is thought that MuDRB is necessary for the integration of MULEs. Most MULEs lack MuDRB and it was proposed that the gene encoding MuDRB was acquired to increase MuDR transposition (Lisch 2002). In contrast to plant MULEs, *Mutyl* may have acquired *mutB* to reduce transposition.

Alternatively, the activity of *mutB* may be linked to another peculiarity presented by *Mutyl*.

Analysis of the sequence of *Mutyl* revealed the presence of two consensus 5′ splice sites and of two consensus 3′ splice sites, within *mutB* coding sequence, indicating that this gene may be subjected to alternative splicing. Using RT-PCR, this turned out to be the case, making mutB the first alternatively spliced gene of *Y. lipolytica*. It must be stressed that alternative splicing is rare in yeast, even in *S. cerevisiae* where alternative splicing affecting protein activities has only been recently demonstrated (Grund et al. 2008; Juneau et al. 2009). The four combinations of alternative splicing were indeed used, leading to four transcripts. Despite little changes in the mRNA size (from 796 to 811 nt), this yielded four variant proteins with important difference in size: 128, 156, 161, and 190 amino acids. Only two variants retained the detected putative zinc finger motif, suggesting that alternative splicing could modulate the activity of MutB, if, for instance, this activity was linked to DNA binding involving the zinc finger motif.

4 Conclusions

Along with *C. albicans*, *Y. lipolytica* is the one Saccharomycotina yeast, which harbors the most diverse types of transposons. Paradoxically, *Y. lipolytica* does not carry any member of the well-known transposons of the Ty1/*copia* class found in the *Saccharomyces*. Furthermore, the two detected members of the Ty3/*gypsy* class are very exceptional, because Ty6 displays a programmed −1 ribosomal frameshifting not found in any other yeast transposon and Ylt1 harbors only a single but very long open reading frame. In addition, *Y. lipolytica* shows a surprising ability at harboring and receiving various types of transposons through

horizontal transfer. Again like *C. albicans*, *Y. lipolytica* carries a LINE and a DNA transposon. It is well known that the number and the types of transposons harbored in all organisms vary enormously. The fact that *Y. lipolytica* and *C. albicans* are well apart, in the Saccharomycotina yeast phylogenetic tree, confirms this observation.

The number of full-length transposons is much reduced in *Y. lipolytica*, even in comparison with *S. cerevisiae*. It seems that *Y. lipolytica* maintains the transposons copy number as low as possible, very likely to avoid deleterious effect of insertion in genes, but considering that half of the genome of this yeast is noncoding, this observation is surprising. Various types of non-LTR and LTR retrotransposons have colonized this species: Ylli, Ylt1, Tyl6, an unidentified Ty3, and another element, for which only LTRs were identified. Taken together, this information suggests that *Y. lipolytica* was subjected to many rounds of transposon invasions and that the long intergenic regions may be the result of a number of early transposon invasion/evolution cycles. This may imply that *Y. lipolytica* does not have any constraint on the size of its genome unlike the other known Saccharomycotina yeasts. The existence of strains carrying some transposons and of strains devoid of the same transposons makes *Y. lipolytica* a very good model system for the study of mechanisms used by transposons to invade a species and to propagate themselves. Like in *C. albicans*, elements present in *Y. lipolytica* are very likely to be present in closely related species (Knutsen et al. 2007) and comparative genomics studies should shed light on some of the mechanisms used for propagation of the specific *Y. lipolytica* elements and on their regulation.

Acknowledgments S.C. thanks Christophe Ozanne for sharing unpublished results and the members of the Genolevures consortium for discussion.

References

Babushok DV, Kazazian HHJ (2007) Progress in understanding the biology of the human mutagen LINE-1. Hum Mutat 28:527–539

Barth G, Gaillardin C (1996) Yarrowia lipolytica. In: Wolf K (ed) Nonconventional yeasts in biotechnology. Springer, Berlin, pp 313–388

Barth G, Gaillardin C (1997) Physiology and genetics of the dimorphic fungus Yarrowia lipolytica. FEMS Microbiol Rev 19:219–237

Belancio VP, Hedges DJ, Deininger P (2008) Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res 18:343–358

Bon E, Neuveglise C, Lepingle A, Wincker P, Artiguenave F, Gaillardin C, Casaregola S (2000) Genomic exploration of the hemiascomycetous yeasts: 6. Saccharomyces exiguus. FEBS Lett 487:42–46

Burns KH, Boeke JD (2008) Great exaptations. J Biol 7:5

Casaregola S, Neuveglise C, Lepingle A, Bon E, Feynerol C, Artiguenave F, Wincker P, Gaillardin C (2000) Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia lipolytica. FEBS Lett 487:95–100

Casaregola S, Neuveglise C, Bon E, Gaillardin C (2002) Ylli, a non-LTR retrotransposon L1 family in the dimorphic yeast Yarrowia lipolytica. Mol Biol Evol 19:664–677

Chalker DL, Sandmeyer SB (1992) Ty3 integrates within the region of RNA polymerase III transcription initiation. Genes Dev 6:117–128

Chalker DL, Sandmeyer SB (1993) Sites of RNA polymerase III transcription initiation and Ty3 integration at the U6 gene are positioned by the TATA box. Proc Natl Acad Sci USA 90:4927–4931

Covey SN (1986) Amino acid sequence homology in gag region of reverse transcribing elements and the coat protein gene of cauliflower mosaic virus. Nucleic Acids Res 14:623–633

Dong C, Poulter RT, Han JS (2009) LINE-like retrotransposition in Saccharomyces cerevisiae. Genetics 181:301–311

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

Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuveglise C, Talla E, Goffard N, Frangeul L, Aigle M, Anthouard V, Babour A, Barbe V, Barnay S, Blanchin S, Beckerich JM, Beyne E, Bleykasten C, Boisrame A, Boyer J, Cattolico L, Confanioleri F, De Daruvar A, Despons L, Fabre E, Fairhead C, Ferry-Dumazet H, Groppi A, Hantraye F, Hennequin C, Jauniaux N, Joyet P, Kachouri R, Kerrest A, Koszul R, Lemaire M, Lesur I, Ma L, Muller H, Nicaud JM, Nikolski M, Oztas S, Ozier-Kalogeropoulos O, Pellenz S, Potier S, Richard GF, Straub ML, Suleau A, Swennen D, Tekaia F, Wesolowski-Louvel M, Westhof E, Wirth B, Zeniou-Meyer M, Zivanovic I, Bolotin-Fukuhara M, Thierry A, Bouchier C, Caudron B, Scarpelli C, Gaillardin C, Weissenbach J, Wincker P, Souciet JL (2004) Genome evolution in yeasts. Nature 430:35–44

Dunn BM, Goodenow MM, Gustchina A, Wlodawer A (2002) Retroviral proteases. Genome Biol 3:reviews3006.1–3006.7

Feschotte C, Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet 41:331–368

Gao X, Havecker ER, Baranov PV, Atkins JF, Voytas DF (2003) Translational recoding signals between gag and pol in diverse LTR retrotransposons. RNA 9:1422–1430

Goodwin TJ, Poulter RT (2000) Multiple LTR-retrotransposon families in the asexual yeast Candida albicans. Genome Res 10:174–191

Goodwin TJ, Poulter RT (2001) The diversity of retrotransposons in the yeast Cryptococcus neoformans. Yeast 18:865–880

Goodwin TJ, Ormandy JE, Poulter RT (2001) L1-like non-LTR retrotransposons in the yeast Candida albicans. Curr Genet 39:83–91

Goodwin TJ, Busby JN, Poulter RT (2007) A yeast model for target-primed (non-LTR) retrotransposition. BMC Genomics 8:263

Grund SE, Fischer T, Cabal GG, Antunez O, Perez-Ortin JE, Hurt E (2008) The inner nuclear membrane protein Src1 associates with subtelomeric genes and alters their regulated gene expression. J Cell Biol 182:897–910

Hani J, Feldmann H (1998) tRNA genes and retroelements in the yeast genome. Nucleic Acids Res 26:689–696

Havecker ER, Gao X, Voytas DF (2004) The diversity of LTR retrotransposons. Genome Biol 5:225

Higashiyama T, Noutoshi Y, Fujie M, Yamada T (1997) Zepp, a LINE-like retrotransposon accumulated in the Chlorella telomeric region. EMBO J 16:3715–3723

Hoebeke M, Schbath S (2006) R'MES: finding exceptional motifs, version 3. User guide. http://genome.jouy.inra.fr/ssb/rmes

Juneau K, Nislow C, Davis RW (2009) Alternative splicing of PTC7 in Saccharomyces cerevisiae determines protein localization. Genetics 183:185–194

Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF (1998) Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence. Genome Res 8:464–478

Knutsen AK, Robert V, Poot GA, Epping W, Figge M, Holst-Jensen A, Skaar I, Smith MT (2007) Polyphasic re-examination of Yarrowia lipolytica strains and the description of three novel

Candida species: Candida oslonensis sp. nov., Candida alimentaria sp. nov. and Candida hollandica sp. nov. Int J Syst Evol Microbiol 57:2426–2435

Kovalchuk A (2005) Molecular analysis of the LTR retrotransposon Ylt1 from the genome of the dimorphic fungus *Yarrowia lipolytica*. PhD Thesis, Technische Universität Dresden

Kovalchuk A, Senam S, Mauersberger S, Barth G (2005) Tyl6, a novel Ty3/gypsy-like retrotransposon in the genome of the dimorphic fungus Yarrowia lipolytica. Yeast 22:979–991

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Lisch D (2002) Mutator transposons. Trends Plant Sci 7:498–504

Louis VL, Despons L, Friedrich A, Martin T, Durrens P, Casarégola S, Neuvéglise C, Fairhead C, Marck C, Cruz JA, Straub ML, Kugler V, Sacerdot C, Uzunov Z, Thierry A, Weiss S, Bleykasten C, De Montigny J, Jacques N, Jung P, Lemaire M, Mallet S, Morel G, Richard GF, Sarkar A, Savel G, Schacherer J, Seret ML, Talla E, Samson G, Jubin C, Poulain J, Vacherie B, Barbe V, Pelletier E, Sherman DJ, Westhof E, Weissenbach J, Baret PV, Wincker P, Gaillardin C, Dujon B, Souciet JL (2012) Pichia sorbitophila, an interspecies yeast hybrid, reveals early steps of genome resolution after polyploidization. G3 (Bethesda) 2:299–311

Malik HS, Eickbush TH (1999) Modular evolution of the integrase domain in the Ty3/gypsy class of LTR retrotransposons. J Virol 73:5186–5190

Malik HS, Burke WD, Eickbush TH (1999) The age and evolution of non-LTR retrotransposable elements. Mol Biol Evol 16:793–805

Myers JS, Vincent BJ, Udall H, Watkins WS, Morrish TA, Kilroy GE, Swergold GD, Henke J, Henke L, Moran JV, Jorde LB, Batzer MA (2002) A comprehensive analysis of recently integrated human Ta L1 elements. Am J Hum Genet 71:312–326

Neuveglise C, Solano-Serena F, Brignon P, Gendre F, Gaillardin C, Casaregola S (2000) Homologous recombination and transposition generate chromosome I neopolymorphism during meiosis in Saccharomyces cerevisiae. Mol Gen Genet 263:722–732

Neuveglise C, Feldmann H, Bon E, Gaillardin C, Casaregola S (2002) Genomic evolution of the long terminal repeat retrotransposons in hemiascomycetous yeasts. Genome Res 12:930–943

Neuveglise C, Chalvet F, Wincker P, Gaillardin C, Casaregola S (2005) Mutator-like element in the yeast Yarrowia lipolytica displays multiple alternative splicings. Eukaryot Cell 4:615–624

Pavlicek A, Paces J, Zika R, Hejnar J (2002) Length distribution of long interspersed nucleotide elements (LINEs) and processed pseudogenes of human endogenous retroviruses: implications for retrotransposition and pseudogene detection. Gene 300:189–194

Raizada MN, Benito MI, Walbot V (2001) The MuDR transposon terminal inverted repeat contains a complex plant promoter directing distinct somatic and germinal programs. Plant J 25:79–91

Ribeiro-dos-Santos G, Schenberg AC, Gardner DC, Oliver SG (1997) Enhancement of Ty transposition at the ADH4 and ADH2 loci in meiotic yeast cells. Mol Gen Genet 254:555–561

Sassaman DM, Dombroski BA, Moran JV, Kimberland ML, Naas TP, DeBerardinis RJ, Gabriel A, Swergold GD, Kazazian HHJ (1997) Many human L1 elements are capable of retrotransposition. Nat Genet 16:37–43

Schmid-Berger N, Schmid B, Barth G (1994) Ylt1, a highly repetitive retrotransposon in the genome of the dimorphic fungus *Yarrowia lipolytica*. J Bacteriol 176:2477–2482

Senam S (2004) Molekulare Charakterisierung des Retrotransposons Ylt1 in der Hefe *Yarrowia lipolytica*. PhD Thesis, Technische Universität Dresden

Souciet JL, Dujon B, Gaillardin C, Johnston M, Baret PV, Cliften P, Sherman DJ, Weissenbach J, Westhof E, Wincker P, Jubin C, Poulain J, Barbe V, Segurens B, Artiguenave F, Anthouard V, Vacherie B, Val ME, Fulton RS, Minx P, Wilson R, Durrens P, Jean G, Marck C, Martin T, Nikolski M, Rolland T, Seret ML, Casaregola S, Despons L, Fairhead C, Fischer G, Lafontaine I, Leh V, Lemaire M, de Montigny J, Neuveglise C, Thierry A, Blanc-Lenfle I, Bleykasten C, Diffels J, Fritsch E, Frangeul L, Goeffon A, Jauniaux N, Kachouri-Lafond R, Payen C, Potier S, Pribylova L, Ozanne C, Richard GF, Sacerdot C, Straub ML, Talla E (2009) Comparative genomics of protoploid Saccharomycetaceae. Genome Res 19:1696–1709

Turkel S, Liao XB, Farabaugh PJ (1997) GCR1-dependent transcriptional activation of yeast retrotransposon Ty2-917. Yeast 13:917–930

Werner K (2009) Studies on secretion of Gpr1 protein in *Yarrowia lipolytica* and of the Gpr1p orthologues in *Saccharomyces cerevisiae*. Diploma thesis, Technische Universität Dresden (in German)

Xiong Y, Eickbush TH (1988) The site-specific ribosomal DNA insertion element R1Bm belongs to a class of non-long-terminal-repeat retrotransposons. Mol Cell Biol 8:114–123

Noncoding RNA Genes Transcribed by RNA Polymerase III in *Yarrowia lipolytica*

Cécile Neuvéglise, Claude Gaillardin, and Christian Marck

Abstract In the hemiascomycetous yeast *Yarrowia lipolytica*, the Pol III-transcribed ncRNAs display a number of unusual and remarkable features with respect to other Hemiascomycetes. With 510 bona fide genes, the genome of this organism contains the largest number of tRNA genes among yeasts. Unlike all other Hemiascomycetes which unconventionally decode the Leu CUN and Arg CGN codons like bacteria, *Y. lipolytica* follows the regular eukaryotic rule. Full genome sequencing of the first Hemiascomycetes revealed many cases of tandem tRNA genes. Such genes are present in *Y. lipolytica*, together with other unexpected cases of Pol III hybrid genes. In *Y. lipolytica*, contrary to most Hemiascomycetes and similarly to *Schizosaccharomyces pombe*, the 5S rRNA genes (transcribed by Pol III) are not embedded into the multiple rDNA units. *Y. lipolytica* appears unique with the presence of ~50 copies of hybrid dicistronic tRNA-5S rRNA genes coexisting with ~60 copies of more conventional, isolated 5S rRNA genes. These hybrid genes were shown experimentally to be actively co-transcribed in vitro and in vivo from the leader tRNA genes rendering the specific transcription factor of 5S rRNA genes, TFIIIA, dispensable. *Y. lipolytica* also contains a novel ncRNA, RUF70, expressed from multiple genes located 3′ to each of the 13 copies of the tRNA-Trp gene. Such a complex assembly of Pol III genes is not found outside the *Yarrowia* clade. Many Pol III-related features differentiate *Y. lipolytica* from other Hemiascomycetes, rendering it closer to *Schizosaccharomyces pombe*, to other ascomycetes, and to the rest of eukaryotes.

C. Neuvéglise (✉) • C. Gaillardin
INRA UMR1319, AgroParisTech, Micalis, Centre de Biotechnologie Agro-Industrielle, BP 01, 78850 Thiverval-Grignon, France
e-mail: ncecile@grignon.inra.fr; claude.gaillardin@grignon.inra.fr

C. Marck
Institut de Biologie et Technologies de Saclay (iBiTec-S), 91191 Gif-sur-Yvette Cedex, France
e-mail: christian.marck@cea.fr

G. Barth (ed.), *Yarrowia lipolytica*, Microbiology Monographs 24,
DOI 10.1007/978-3-642-38320-5_4, © Springer-Verlag Berlin Heidelberg 2013

Contents

1 *Yarrowia lipolytica*: A Distantly Related Hemiascomycete ... 80
2 Specificity of RNA Polymerase III Transcription ... 82
2.1 Genes Transcribed by Pol III in Hemiascomycetes ... 82
2.2 Basal Factors TFIIIC, TFIIIB, and TFIIIA and Transcription Initiation ... 82
2.3 Pol III Transcription Termination, Recycling, and Regulation ... 85
2.4 Dicistronic Pol III Genes ... 86
3 Transfer RNAs ... 86
3.1 Transcription of Regular tRNA Genes ... 87
3.2 Transcription of Tandem tRNA-tRNA Genes ... 89
3.3 Processing of Pre-tRNAs ... 90
3.4 Decoding Exceptions in Hemiascomycetes ... 91
4 The 5S Ribosomal RNA ... 96
4.1 TFIIIA: The 5S Ribosomal RNA Gene-Specific Factor ... 96
4.2 Different Genomic Organization of the 5S Ribosomal RNA Genes ... 96
5 Dicistronic tRNA-5S rRNA Genes ... 97
5.1 Transcription of Dicistronic tRNA-5S rRNA Genes ... 98
5.2 Specificity of *Y. lipolytica* TFIIIA ... 99
6 Dicistronic tRNA-Trp/RUF70 Genes ... 99
6.1 Transcription of Dicistronic tRNA-Trp/RUF70 Genes ... 99
7 Other Pol III ncRNAs ... 99
7.1 A Double Constraint Is Exerted on Internal Promoters ... 100
7.2 Spliceosomal U6 RNA ... 100
7.3 RPR1: The RNA Component of RNase P ... 100
7.4 7SL/scR1: The RNA Component of the Signal Recognition Particle ... 102
7.5 snR52 RNA: The Only Pol III-Produced snoRNA ... 103
7.6 Conservation of A and B Pol III Promoter Sequences in Hemiascomycetes ... 104
8 Unique Pol III Features Confirm *Y. lipolytica* as a "Different Hemiascomycetous Yeast" 104
References ... 105

1 *Yarrowia lipolytica*: A Distantly Related Hemiascomycete

The genome of the yeast *Yarrowia lipolytica* was sequenced a few years ago, together with those of *Candida glabrata*, *Kluyveromyces lactis*, and *Debaryomyces hansenii* in a joint effort to explore the molecular evolutionary mechanisms in unicellular eukaryotes (Dujon et al. 2004). In addition to its biotechnological interest, *Y. lipolytica* was chosen because of its phylogenetic position in the hemiascomycetous phylum: a preliminary exploration of this whole phylum, which covers an evolutionary range comparable to that of the entire phylum of Chordates, suggested that the phylogenetic branch leading to the modern *Y. lipolytica* diverged rapidly, probably at the basis of the phylum. Thus, *Y. lipolytica* harbors one of the most distantly related genomes to that of the model yeast *Saccharomyces cerevisiae* (Fig. 1) (Casaregola et al. 2000; Souciet et al. 2000).

This review focuses on the genes transcribed by DNA-dependent RNA polymerase III (Pol III for short), their transcription, and their RNA products (tRNA, 5S rRNA, and others). The first sections summarize the present "state of knowledge" about the specificity of Pol III transcription with respect to Pol I and Pol II. This knowledge

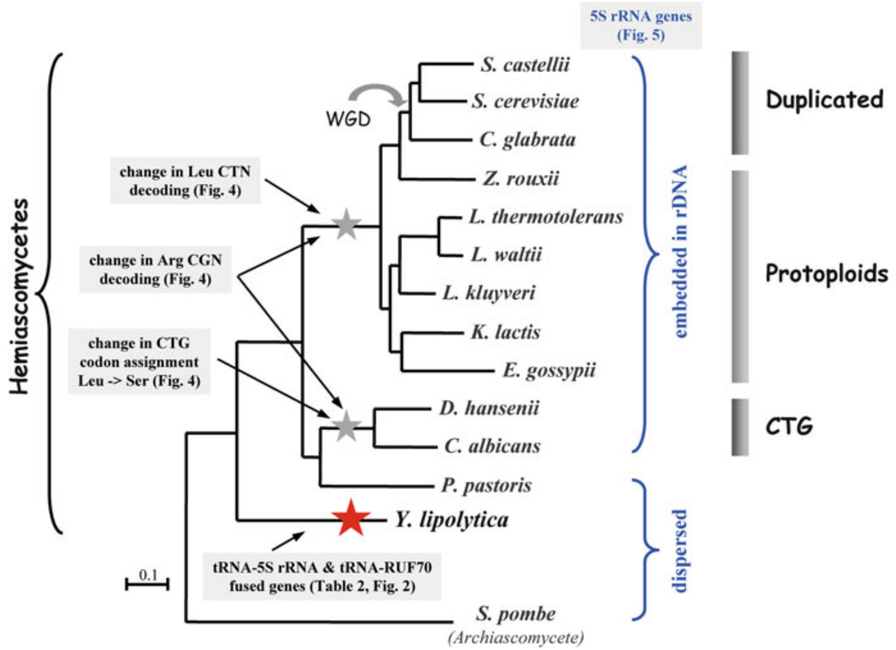


Fig. 1 Remote phylogenetic position of *Y. lipolytica* among Hemiascomycetes. The phylogenetic tree was constructed from a concatenation of 138 genes considered as singleton in their respective genomes. Amino acid sequences were aligned with MAFFT (Katoh et al. 2002) and cleaned with Gblocks (Castresana 2000) before being concatenated. Only gene families for which the ratio between the cleaned blocks and the initial alignment was higher than 70 % were considered. The tree was build from the resulting alignment (51,323 amino acids per species) by maximum likelihood using PHYML (Guindon and Gascuel 2003) with a JTT substitution model corrected for heterogeneity among sites by a Γ-law distribution using 4 different categories of evolution rates. The proportion of invariable sites and the α-parameter of the Γ-law distribution were optimized according to the data. Bootstraps were calculated from 100 replicates and all yielded 100 % values. Exceptional evolutionary events are depicted by *arrows* and *text rectangles* next to each impacted branch. *WGD* whole genome duplication

was mainly deduced from experimental and in silico studies in *S. cerevisiae* and humans during the last 30 years. Secondly, we report the in silico analyses of Pol III genes in hemiascomycetous genomes and emphasize the experimental work carried out in *Y. lipolytica* that revealed several unusual Pol III features not yet found elsewhere (e.g., dicistronic tRNA-5S rRNA genes). The unusual Pol III-related features discovered in *Y. lipolytica* further confirm that it is indeed a nonconventional hemiascomycetous yeast.

2 Specificity of RNA Polymerase III Transcription

Contrary to bacteria and archaea, eukaryotes make use of three specialized RNA polymerases. Pol I has to transcribe only one gene, although present in about hundred copies in yeasts, the gene expressing the pre-rRNA 35S, the product of which is later matured into the 18S, 5.8S, and 25S rRNAs. Pol II transcribes genes coding for proteins into messenger RNAs and also genes coding for small nuclear and small nucleolar RNAs (sn- and snoRNAs), either as individual or polycistronic genes or embedded in a spliceosomal intron. Pol III is specialized in transcribing tRNA genes, 5S rRNA genes, as well as some less abundant ncRNAs, as described below.

2.1 Genes Transcribed by Pol III in Hemiascomycetes

In *Y. lipolytica*, as well as in other hemiascomycetous yeasts, genes transcribed by Pol III include the tRNA family of genes (tDNAs) and the multicopy 5S rRNA genes (Table 1). A few RNAs encoded by a single gene are also produced by Pol III: the spliceosomal RNA U6, and the RPR1 and 7SL/scR1 RNAs which are the RNA components of RNase P and signal recognition particle (SRP), respectively. In the yeast *S. cerevisiae*, an exhaustive inventory of Pol III genes was experimentally performed by ChIP-on-Chip and revealed one additional gene, the snoRNA snR52 and a few other putative genes of unknown function (Harismendy et al. 2003; Roberts et al. 2003). Genes transcribed by Pol III share the common features of internal promoter sequences: small size (a few hundred nt) and lack of spliceosomal introns. For more detailed reviews on Pol III transcription, see Dieci et al. (2007), Geiduschek and Kassavetis (2001), and White (1998).

2.2 Basal Factors TFIIIC, TFIIIB, and TFIIIA and Transcription Initiation

Compared to Pol II which employs a bunch of basal transcription factors, Pol III uses only three factors. Two of them are general transcription factors, TFIIIC and -B, required for the transcription of all genes; the third one, TFIIIA, is specifically devoted to the recognition of the 5S ribosomal RNA genes. A most specific and unusual feature of Pol III transcription is the internal location of the promoter sequences (for most genes and at least in yeasts). In the case of the tRNA genes, these sequences, which are traditionally referred to as the “A” and “B” boxes, correspond to some of the most universally conserved bases of the tRNA molecule. These two key sequences are recognized by TFIIIC, the assembly factor, which directs the fixation of the second factor, TFIIIB upstream of the tRNA gene. Finally, TFIIIB recruits RNA Pol III to bind to upstream DNA and transcribe the gene (Fig. 2a).

Table 1 ncRNAs produced by RNA Polymerase III in *Y. lipolytica* and other Hemiascomycete yeasts

RNA product	Number of genes: *Y. lipolytica*	Number of genes: Other yeasts	Transcription factors requested	Function and specific features in *Y. lipolytica* (*Y.l.*) and other yeasts unless specified
tRNA	510	~160–274	TFIIIC, -B	– Translation – Intron (non-spliceosomal) present in some genes – Tandem co-transcribed genes often found in Hemiascomycetes – In *Y. l.* only, also found as dicistronic tRNA-5S rRNA genes
5S rRNA	109	~100	TFIIIA, -C, -B	– Ribosomal RNA, part of the large subunit (LSU) – Genes are dispersed in the genome in *Y. l.* – Embedded the rRNA unit in most other yeasts – In *Y. l.* only, about 44 % of the 5S rRNA genes are located 3′ of tRNAs and co-transcribed without the need of TFIIIA
RUF70	13	0	None	– Structural RNA of unknown function – Located 3′ of all tRNA-Trp genes – Gene co-transcribed from the leader tRNA gene – Not found outside the *Yarrowia* clade
U6 RNA	1	1	TFIIIC, -B	– Gene is named *SNR6* in *S.c.* – Spliceosomal RNA, the only one produced by Pol III – Additional TATA-type upstream promoter – B-block promoter downstream of poly-T terminator
RPR1 RNA	1	1	TFIIIC, -B	– RNA component of RNase P (5′ pre-tRNA splicing)
scR1/7SL RNA	1	1	TFIIIC, -B	– RNA component of signal recognition particle (SRP) that targets presecretory and membrane proteins to the endoplasmic reticulum (ER) – Gene is named *SCR1* in *S.c.* (small cytoplasmic RNA 1) – *Y.l.* has two genes (*ylSCR1* and *ylSCR2*)
snR52 RNA	1	1	TFIIIC, -B	– C/D box-type snoRNA responsible for one 2′-*O*-methylation in 25S rRNA in *Y. l.* and also for another one in 18S rRNA in other yeasts – Only snoRNA produced by Pol III – Deletant in *S.c.* is viable

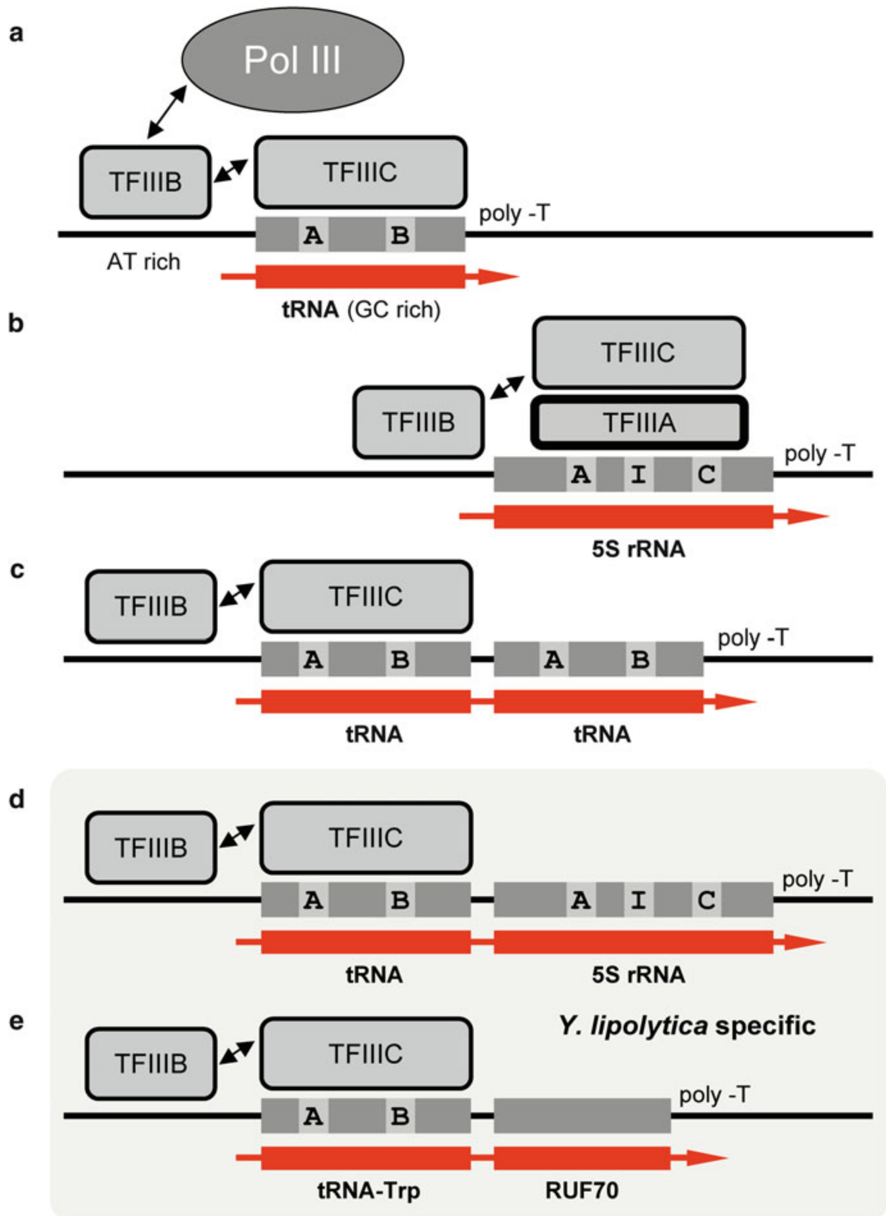


Fig. 2 Overview of transcription by RNA polymerase III in yeasts. All panels: *Thin black lines*, DNA; *heavy dark gray bars*, tRNA, 5S rRNA, and RUF70 is similar to that of genes with promoter sequence (A and B boxes for tDNAs; A, I, and C for 5S rRNA genes); "poly-T," Pol III transcription terminator; *red arrows*, primary transcripts; heavy *red bars*, matured RNA products; and *gray background*, dicistronic genes specific to *Y. lipolytica*. (**a**) In *S. cerevisiae*, regular tDNAs transcription starts with the binding of TFIIIC onto the two promoters A and B (GC-rich region). Once bound,

Both TFIIIC and TFIIIB are multisubunit factors: TFIIIC comprises six subunits amounting to 600 kDa, while TFIIIB (~170 kDa) is made of only three subunits including the TATA-box binding protein (TBP) that is also part of the Pol I and Pol II initiation complexes. Taking into account the Pol III itself, the whole basal Pol III transcriptional apparatus of yeasts reaches 1.6 mDa. Remarkably, the subunit composition of TFIIIC, TFIIIB, and Pol III remains unchanged from yeasts to human (Dumay-Odelot et al. 2007; Schramm and Hernandez 2002).

A terribly mysterious point is how TFIIIC can accommodate very different distances between the A and B boxes and thereby their different angular orientations along the DNA helix. Indeed, this basal distance of about 36 nt is more or less increased because of the variable loop present in eukaryotic Leu and Ser tRNAs (up to 5 extra nt) and even more because of the intron (see Sect. 3.3 below). In yeasts, the longest known tRNA intron is that of tDNA-Leu (CAG) in *K. lactis* (288 nt (Souciet et al. 2009)) and the longest one in *Y. lipolytica* measures 49 nt in tDNA-Ile (TAT) (Dujon et al. 2004). Use of artificial tRNA introns of variable lengths has shown that TFIIIC binds initially to the B box, irrespectively of the intron length, and enhances its ability to bind the A box (Baker et al. 1987). A scanning transmission electron microscopy (STEM) exploration of tDNA with introns of various length revealed that TFIIIC behaves as two individualized factor domains bound to the A and B blocks. Looping out of tDNA was also observed with long introns (Schultz et al. 1989).

2.3 Pol III Transcription Termination, Recycling, and Regulation

Contrary to Pol I and II, no transcription termination complex has yet been found for Pol III; in vitro transcription runs until Pol III encounters the Pol III-specific termination signal which is made up of a series of T nucleotides in the RNA-like strand (that is a series of A nucleotides in the strand read by Pol III). For example, in *Y. lipolytica* a common termination signal is “TTTTTG.” Pol III is furthermore

Fig. 2 (continued) TFIIIC binds TFIIIB upstream of tDNA (AT-rich region) and TFIIIB recruits Pol III for multiple cycles of transcription ending at the Poly-T sequence. (**b**) TFIIIA first binds to regular 5S rRNA genes and the TFIIIA-5S rRNA gene complex is bound by TFIIIC. Then, transcription proceeds similarly to tDNA transcription. A single primary transcript is produced and processed into two functional tRNAs by the usual 5′ and 3′ tRNA maturation machineries. (**c**) Transcription of a dicistronic tRNA-tRNA gene starts through the binding of a single TFIIIC molecule onto the upstream gene. The A and B promoter sequences of the second gene are not used. (**d**) TFIIIA-independent transcription of a dicistronic tRNA-5S rRNA gene proceeds similarly in *Y. lipolytica* through the recognition by TFIIIC of the A and B promoter elements of the leader tRNA gene. A single primary RNA is produced and matured into two functional products (tRNA and 5S rRNA). (**e**) Transcription of the dicistronic tRNA-Trp-RUF70 gene is similar to that of tandem tRNA genes shown in (**c**). RUF70 gene has no internal promoters. These two last types of dicistronic genes, tRNA-5S rRNA and tRNA-RUF70 genes, are specific to *Y. lipolytica*

characterized by an efficient mechanism of "recycling" or "facilitated reinitiation" that allows it to transcribe several times the same molecule (Dieci and Sentenac 1996; Ferrari et al. 2004); for this, a permanent contact of the Pol III assembly factors with the DNA is hypothesized.

In addition to the three basal transcription factors TFIIIC, -B, and -A, Pol III transcription is also regulated in yeasts and mammals by at least two effectors recently discovered: Maf1 is a negative regulator (Ciesla et al. 2007; Oficjalska-Pham et al. 2006) and Sub1 (in yeasts, named PC4 in mammals) is a positive regulator also required for efficient recycling (Tavenet et al. 2009). One more factor also implicated in Pol III transcription is the Pol II factor TFIIS known to stimulate transcript cleavage activity of arrested Pol II molecules. Unexpectedly, TFIIS and Pol III occupancies correlated well in a genome-wide analysis, and in vitro transcription assays suggested that TFIIS may affect Pol III start site selection (Ghavi-Helm et al. 2008).

2.4 Dicistronic Pol III Genes

The predominantly internal location of Pol III genes makes each of them a stand-alone transcriptional unit, and therefore, combining two or more genes appears unnecessary as well as unlikely. Nevertheless, combined co-transcribed Pol III genes are not uncommon. Dicistronic tRNA-snoRNA were first found in plants (Kruszka et al. 2003), and tandem tRNA genes are common in Hemiascomycetes (Dujon et al. 2004). tRNA-miRNA genes were also found in the human genome (Borchert et al. 2006) and in a mouse gammaherpesvirus (Pfeffer et al. 2005). *Y. lipolytica* displays two remarkable kinds of novel combined Pol III genes not found elsewhere: a large fraction of the 5S rRNA genes appears under the form of combined tRNA-5S rRNA genes, and a novel ncRNA of unknown function, RUF70, is expressed from tRNA-Trp (CCA)-RUF70 dicistronic genes (Acker et al. 2008).

3 Transfer RNAs

Transfer RNAs (tRNAs) provide the link between the messenger RNAs and the protein sequences. As a general rule, ncRNA genes are by far less easy to distinguish in genomes than the majority of Pol II genes. However, among ncRNAs, tRNA genes are the most easily detectable in genomes as they all obey a precise secondary structure known as the "Cloverleaf" model (Holley 1965; Holley et al. 1965). Several software programs are available to detect these genes (e.g., tRNAscan) as well as a number of databases to access their sequences (Abe et al. 2009; Chan and Lowe 2009; Juhling et al. 2009; Lowe and Eddy 1997). For review on various aspects of tRNA biogenesis, see Phizicky and Hopper (2010).

3.1 Transcription of Regular tRNA Genes

In the yeast *S. cerevisiae* (and probably in neighbor species), transcription of tRNA genes and early processing of the transcripts occur coordinately in the nucleolus, although tRNA genes are dispersed in the genome. For review on cellular dynamics of tRNAs and their genes, see Hopper et al. (2010).

3.1.1 Consensus Sequences of A and B Promoter Elements

Considerable past experimental work has demonstrated the crucial role of some key nucleotides for the efficient binding of TFIIIC, the Pol III assembly factor (for review, see Schramm and Hernandez (2002) and Sprague (1995)). A more recent survey of all tDNA sequences from several hemiascomycetous genomes has allowed their further delineation and led to the following consensus sequences: A box, $T_8R_9G_{10}Y_{11}nnA_{14}nnnG_{18/19}$, and B box, $G_{53}WTCRAnnC_{61}$ (R standing for purine; Y, pyrimidine; W, A, or T; n, any nucleotide; Fig. 3a, b) (Marck et al. 2006). As a matter of fact, the most conserved nucleotides correspond, in the RNA structure, to key nucleotides implicated in tRNA secondary and tertiary structures. For example, T_8 and A_{14} form a Watson–Crick base pair, and G_{53} and C_{61} make the terminal WC base pair of the T-stem. The above consensus sequences are not strictly obeyed, a few exceptions are found in *Y. lipolytica* (Fig. 3c), and still more exceptions are found in the Archiascomycete *Schizosaccharomyces pombe*, for example, A_{53}:T_{61} instead of G_{53}:C_{61} (see Fig. 2 in Marck et al. (2006) for other sequence exceptions). An attractive hypothesis suggests that TFIIIC might have been selected at an early time of eukaryotic evolution to recognize these key nucleotides at the DNA level and that, afterwards, the same consensus sequences were "adopted" by other Pol III genes.

3.1.2 Predominance of Internal Promoters vs. Upstream Promoters

Efficient transcription of most tRNA genes from yeasts relies solely on the two internal A and B promoters; however, an upstream TATA element may be present that facilitates TFIIIB binding (that contains TBP) without prior binding of TFIIIC. Indeed, in vitro TFIIIC-independent transcription of some *S. cerevisiae* tRNA genes was experimentally observed, thanks to the presence of a cryptic TATA element (Dieci et al. 2000). A similar situation occurs for the U6 RNA gene, in which an upstream TATAAATA element is present (see below Sect. 7.6). Recently, anomalous tRNA genes, so-called circularly permuted genes, were discovered in the red alga *Cyanidioschyzon merolae* (Soma et al. 2007). In such genes, the order of the A and B boxes is reversed and, here also, a predominant upstream TATA-box element accounts for transcription initiation by Pol III. Importance of possible upstream promoters (at least AT-rich sequences) should be therefore kept in mind; these upstream promoters being more strictly required in mammalian

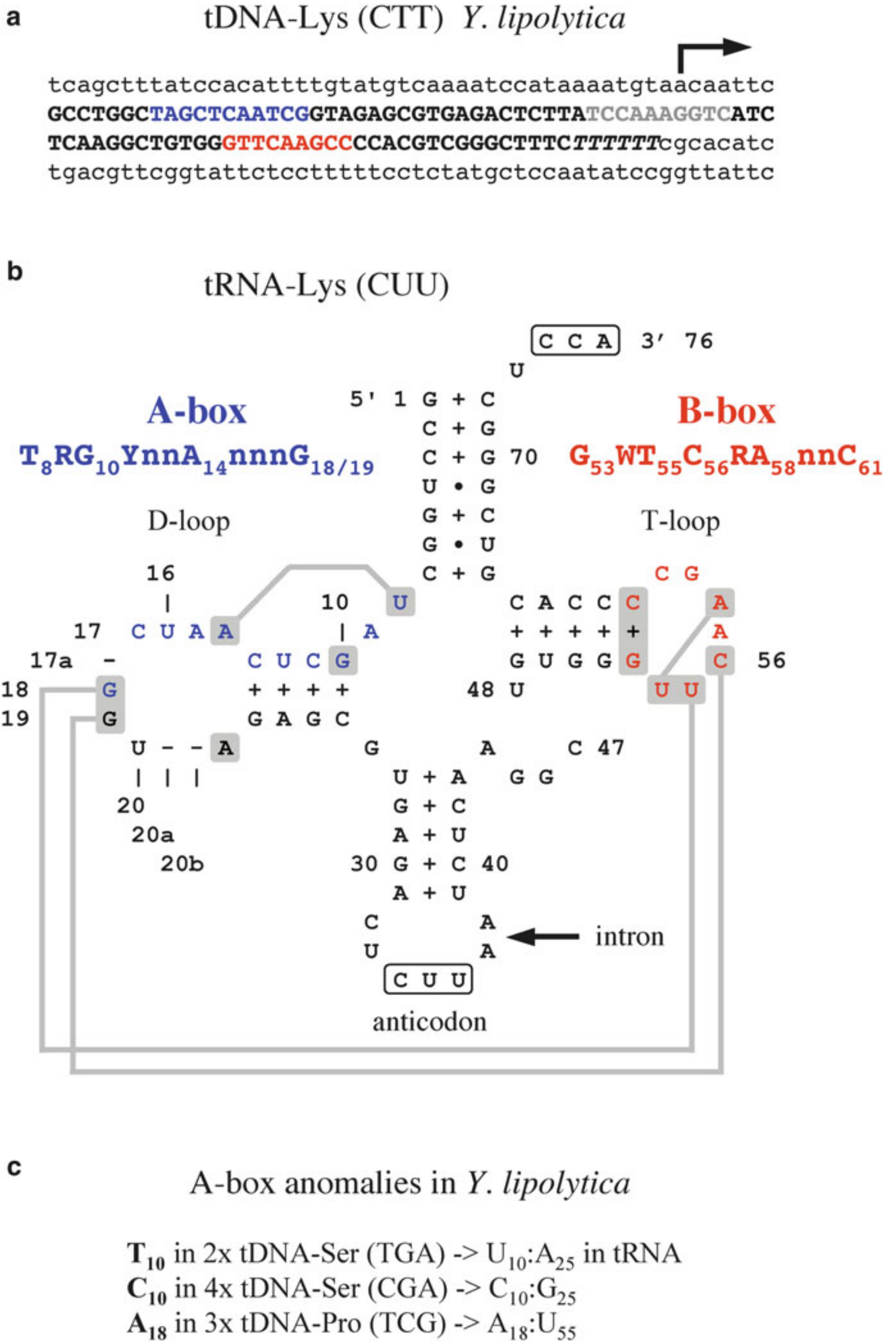


Fig. 3 tRNA gene transcription and Cloverleaf model. (**a**) One of the 34 copies of the tRNA-Lys (CUU) genes of *Y. lipolytica* is taken as an example: *higher case bold letters*, tRNA gene

genomes for the transcription of Pol III genes outside tRNA and 5S rRNA genes (see below Sects. 7.3 and 7.4).

3.1.3 tRNA Genes Are Also Chromatin Domains Insulators

A novel function for tRNA genes in compartmentalizing and organizing eukaryotic chromatin was recently elucidated. The first demonstration of a tRNA gene being involved in blocking the spread of silencing was reported in *S. cerevisiae* (Donze et al. 1999). A tRNA-Thr gene adjacent to the silenced *HMR* locus blocks the spreading of silencing while deletion of this gene increased the spreading of silencing. In various organisms (*S. cerevisiae*, *S. pombe*, *Drosophila melanogaster*), silencing is in turn mediated by different proteins but, in all cases, the barrier activity requires intact A and B tDNA promoters and intact TFIIIC and TFIIIB transcription factors, but not Pol III itself (for review, see Haldar and Kamakaka (2006)).

3.2 Transcription of Tandem tRNA-tRNA Genes

Dicistronic tRNA genes, first reported in *S. pombe*, have also been found in hemiascomycetous genomes. In such tandem genes, the sequence separating the

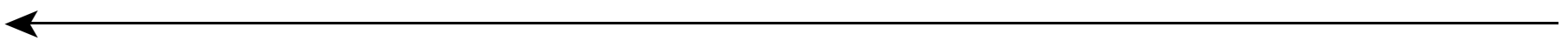


Fig. 3 (continued) (GC rich); *lowercase letters*, 5′ and 3′ flanking sequences (AT rich); *blue letters*, A-box promoter sequence; *red*, B box; *arrow*, transcription start site (about 15 nt upstream of A box); *gray letters*, intron; *italic*, poly-T transcriptional terminator (Pol III specific). (**b**) tRNA product shown under the Cloverleaf form after removing of the intron (between nt 37 and 38), trimming of the 5′ and 3′ ends, and addition of CCA 3′ terminal sequence (*boxed*, always done enzymatically in eukaryotes). Note that to become fully functional, all tRNA molecules have to undergo a number of chemical modifications (not shown). The standard tRNA numbering is used with the first ribonucleotide numbered "1" (Juhling et al. 2009), "*Plus*" signs indicate Watson–Crick base pairing, and dots indicate G:U base pairing. In *blue* and *red* are shown the observed consensus sequences for the A and B boxes of tRNA genes from nine Hemiascomycetes (including *Y. lipolytica*, numbers refer to the tRNA molecule) (Marck et al. 2006). Obviously, these sequences, recognized at the DNA level by the transcription factor TFIIIC, include many of the nearly invariable nucleotides (*gray background*) of the tRNA molecule. Most of which are concentrated in positions 8–19 (forward strand of the D-stem and D-loop) and 53–61 (terminal base pair of the T-stem and T-loop). Several of these nearly universally conserved nucleotides are also engaged in some of the 3-D base pairings that stabilize the L-shaped tRNA molecule (*gray lines*). (**c**) A few tDNAs from *Y. lipolytica* slightly deviate from the canonical Cloverleaf model at tRNA positions 10 and 18. Note that correct base pairing at base pairs 10:25 and 18:55 are still preserved. In tDNA-Ser (TGA) and tDNA-Ser (CGA), G_{10} of the A box is missed and substituted by T or C. The canonical tRNA molecule includes four optional positions: nt 17, 17a, 20a, and 20b can be present or absent. In eukaryotes, position 17 is or is not occupied, while position 17a is never occupied (contrary to bacterial and archaeal tRNAs). This eukaryote-specific pattern makes that, in all cases, the fourth base 3′ of the conserved A_{14} is always a G: G_{18} if position 17 is occupied, G_{19} if not occupied. In the tRNA-Pro (TCG), position 17 is not occupied; therefore, the fourth nucleotide after A_{14} is not A_{18} but G_{19}, thus preserving the A-box consensus. Such sequences exceptions cannot be sequencing errors as they are found in all gene copies of the same tRNA type

two genes (i.e., the distance between base 73 of the leading tRNA and base 1 of the second one) is very short (down to 3 nt) and contains no Pol III terminator that would stop transcription (Fig. 2c). The poly-T terminator is located 3′ of the second gene and the two tRNAs are correctly processed by the usual tRNA 5′ and 3′ splicing machineries as demonstrated for tandem genes of *S. pombe* (Willis et al. 1984) and *S. cerevisiae* (Engelke et al. 1985; Reyes et al. 1986). The largest reported number of tandem genes is 17 in *D. hansenii*. In closely related yeasts, the same tandems are observed: e.g., tDNA-Arg (TCT)-tDNA-Asp (GTC) in *S. cerevisiae* and *K. lactis*. In *Y. lipolytica*, nine tandem tRNA genes are present which are however not found elsewhere (Table 2A; see also Supplementary Table 5 in Dujon et al. (2004)).

3.3 Processing of Pre-tRNAs

To become fully functional, all primary tRNA molecules have to undergo a multistep maturation process mainly investigated in *S. cerevisiae*. First, unlike other ncRNAs (considering both Pol II and Pol III ncRNA products), tRNA genes do contain introns. These introns are not removed by the spliceosome but by a specific eukaryotic machinery likely to be distantly related to that of the archaeal kingdom (Belfort and Weiner 1997; Trotta et al. 1997). The eukaryotic tRNA single intron is always located between nt 37 and nt 38 when present and it contains no consensus sequence, unlike Archaea where introns are found at various locations (up to three introns within the same tRNA gene) and harbor the so-called bulge-helix-bulge (BHB) structural splicing motif (Marck and Grosjean 2003).

The role of introns in eukaryotic tRNAs is unclear, and from one yeast genome to another, different tRNA types (different anticodons) bear introns (Table 3). Intron sequences are also more or less variable in some genomes, and it also happens that some copies of the same tRNA type bear the intron while some others do not (e.g., in tRNA-His genes in *Y. lipolytica* (Dujon et al. 2004)). Contrary to spliceosomal introns, tRNA introns are much more common in yeasts than in mammals.

Maturation involves the trimming of the 5′ end by the ribonucleoprotein RNase P, the RNA component of which is also a Pol III-produced RNA (RPR1 gene, see below). For review on RNase P, see Esakova and Krasilnikov (2010). The 3′ end is cut after nucleotide 73 by several enzymes of which tRNase Z and the 3′ terminal sequence CCA are then added (in bacteria, the CCA sequence is often encoded in the genome) (Schurer et al. 2001). Histidine tRNAs contain an unusual extra G nucleotide at their 5′ end (numbered 0, encoded in the genome in bacteria). Finally, all tRNAs undergo a large number of chemical modifications, especially at the wobble first position of the anticodon that pairs to the third base of the codon.

Table 2 The tandem tRNA, 5S rRNA, and tRNA-5S rRNA genes families in *Y. lipolytica*

(A) The 9 tRNA-tRNA genes

Copy number	Gene	Intergenic distance (nt)
3	tRNA-Pro (AGG)-tRNA-Tyr (GTA)	11
3[a]	tRNA-Lys (CTT)-tRNA-Glu (CTC)	17
2	tRNA-Ile (AAT)-tRNA-Gln (CTG)	11
1	tRNA-Val (AAC)-tRNA-Gln (CTG)	15

(B) The 117 5S rRNA genes

	117	5S rRNA genes
	-8	Genes 5' and/or 3' deleted
	-1	Interrupted by a Tyl1 transposable element
=	108	Full length genes
	-60	Regular isolated genes (requiring TFIIIA for transcription)
=	48	Genes located 3' of tRNA genes (not requiring TFIIIA)

(C) The 48 tRNA-5S rRNA genes

Copy number	Gene	Total tDNA (or tandem tDNA) number	Intergenic distance (nt)
28	tRNA-Gly (GCC)-5S rRNA	30	4/5
8	tRNA-Asp (GTC)-5S rRNA	8	4/5
9	tRNA-Glu (CTC)-5S rRNA	27	4
1	tRNA-Gln (CTG)-5S rRNA	13	4
1	tRNA-Thr (AGT)-5S rRNA	23	7
1[b]	tRNA-Lys (CTT)- tRNA-Glu (CTC)-5S rRNA	3	17 4

[a]One of them followed by a 5S rRNA genes (see footnote b)
[b]Tricistronic tRNA-Lys (CTT)-tRNA-Glu (CTC)-5S rRNA gene

3.4 Decoding Exceptions in Hemiascomycetes

The three domains of life, Eukarya, Bacteria, and Archaea, differ in the way they decode the genetic information; that is to say "which tRNA reads which codon(s)." Different "sparing strategies" aimed at reducing the number of different tRNA types required to decode the genetic information are used in the three domains. All strategies rely on the property that specific chemical modifications allow a mature

Table 3 tRNAs gene usage in *Y. lipolytica* and four other selected yeasts

AA		C	AC	Sc	Lk	Dh	Yl	Sp					Sc	Lk	Dh	Yl	Sp
F	Phe	TTT	–	-	-	-	-	-	Y	Tyr	TAT	–	-	-	-	-	-
F	Phe	TTC	(GAA)	10+	8+	8+	17+	5	Y	Tyr	TAC	(GTA)	8+	7+	6+	14+	4+
L	Leu	TTA	(TAA)	7	3	9	1	2+	*	Och	TAA	–	-	-	-	-	-
L	Leu	TTG	(CAA)	10+	11	4	3+	4+	*	Amb	TAG	–	-	-	-	-	-
L	Leu	CTT	(AAG)	-	-	2	21+	5	H	His	CAT	–	-	-	-	-	-
L	Leu	CTC	(GAG)	1	1	-	-	-	H	His	CAC	(GTG)	7	8	5	12±	4
L	Leu	CTA	(TAG)	3+	3+	-	2	1+	Q	Gln	CAA	(TTG)	9	10	7	3	4
L	Leu	CTG	(CAG)	-	1+		13+	1+	Q	Gln	CAG	(CTG)	1	1+	1	15+	2
S	*Ser*	*CTG*	*(CAG)*			*1*											
I	Ile	ATT	(AAT)	13	12	9	26+	8	N	Asn	AAT	–	-	-	-	-	-
I	Ile	ATC	(GAT)	-	-	-	-	-	N	Asn	AAC	(GTT)	10	6	8	16+	6
I	Ile	ATA	(TAT)	2+	1+	2+	1+	1+	K	Lys	AAA	(TTT)	7+	4+	7+	4+	3
M	Met	ATG	(CAT)	5	5	4+	9	3+	K	Lys	AAG	(CTT)	14	9	10+	34+	9
m	iMet	ATG	(CAT)	5	5	3	9	4									
V	Val	GTT	(AAC)	14	14	10	24	9+	D	Asp	GAT	–	-	-	-	-	-
V	Val	GTC	(GAC)	-	-	-	-	-	D	Asp	GAC	(GTC)	15	9	9	28+	8
V	Val	GTA	(TAC)	2	1	1+	2	2	E	Glu	GAA	(TTC)	14	5+	9	6	4
V	Val	GTG	(CAC)	2	3	1	8±	1	E	Glu	GAG	(CTC)	2	14	1	27+	6
S	Ser	TCT	(AGA)	11	11+	6	21	7	C	Cys	TGT	–	-	-	-	-	-
S	Ser	TCC	(GGA)	-	-	-	-	-	C	Cys	TGC	(GCA)	4	5	4	8	3
S	Ser	TCA	(TGA)	3	2	3	2+	2+	*	Opa	TGA	–	-	-	-	-	-
S	Ser	TCG	(CGA)	1+	1+	1+	4+	1+	W	Trp	TGG	(CCA)	6+	5+	4	13+	3
P	Pro	CCT	(AGG)	2	1	1	21+	6	R	Arg	CGT	(ACG)	6	5+	4+	1	8
P	Pro	CCC	(GGG)	-	-	-	-	-	R	Arg	CGC	(GCG)	-	-	-	-	-
P	Pro	CCA	(TGG)	10+	10+	7+	3	2	R	Arg	CGA	(TCG)	-	-	-	25	1+
P	Pro	CCG	(CGG)	-	-	-	2+	1+	R	Arg	CGG	(CCG)	1	1	1+	-	1+
T	Thr	ACT	(AGT)	11	11	8	22	7	S	Ser	AGT	–	-	-	-	-	-

T	Thr	ACC	(GGT)	-	-	-	-	-	S	Ser	AGC	(GCT)	4+	4+	3+	6+	3+
T	Thr	ACA	(TGT)	4	2	2	3+	2	R	Arg	AGA	(TCT)	11	12	9	4	2
T	Thr	ACG	(CGT)	1	1	1	2±	1	R	Arg	AGG	(CCT)	1	1	1	1	1+
A	Ala	GCT	(AGC)	11	13	7	30	9	G	Gly	GGT	–	-	-	-	-	-
A	Ala	GCC	(GGC)	-	-	-	-	-	G	Gly	GGC	(GCC)	16	15	11	30	8
A	Ala	GCA	(TGC)	5	4	4	4	2	G	Gly	GGA	(TCC)	3	2	4	11+	3
A	Ala	GCG	(CGC)	-	-	1	2	1+	G	Gly	GGG	(CCC)	2	2	1	-	1
									Total tRNA species				42	43	43	44	46
									Total tDNAs				274	257	200	510	171
									Genome size (Mb)				12.1	11.3	12.2	20.5	12.6

This table indicates for five yeast genomes the number of tRNA genes sorted by anticodons. The first column (AA) indicates the charged amino acid (in one- and three-letter codes), the second one (C) the codon, and the third one (AC) the anticodon found (between brackets). Genomes reported are *Saccharomyces cerevisiae* (Sc), *Lachancea kluyveri* (Lk), *Debaryomyces hansenii* (Dh), *Yarrowia lipolytica* (Yl), and *Schizosaccharomyces pombe* (Sp). Data for *L. kluyveri* are taken from (Souciet et al. 2009) and for other yeasts from (Marck et al. 2006). The "*plus*" signs denote genes with intron; "*plus or minus*" denote introns of variable length among the different copies. Note that the types (anticodon) of tRNA used by *Y. lipolytica* to read Leu CUN and Arg CGN codons are identical to that of *S. pombe* and not to those of other Hemiascomycetous yeasts (see Fig. 4 for details). In *D. hansenii*, as in all yeasts from the *Candida* clade, the CTG codon (*italics*) is changed from Leu to Ser

* stop codon

tRNA to read more than one codon, if necessary (see examples in Fig. 4 and, for review, see Grosjean et al. (2010) and Marck and Grosjean (2002)). Curiously, Hemiascomycetes use some bacterial sparing strategies as detailed below.

3.4.1 Most tRNA Genes Are Multicopy Genes

Except in the case of parasites such as *Microsporidia*, most eukaryotic tRNA genes are present in several copies (dispersed in the genome) (Marck and Grosjean 2002). The number of copies is tightly correlated to the usage of the amino acid encoded and also to its number of codons. For example, in *Y. lipolytica*, only 8 tRNA-Cys (reading codons UGU and UGC) genes are present while Lys AAA and AAG codons are read by 4 and 34 tRNA-Lys, respectively. In Hemiascomycetes, the observed total number of tDNAs ranges from 162 (*K. lactis*) to 274 (*S. cerevisiae*) and to the yet unsurpassed value of 510 for *Y. lipolytica* (Table 3). Along with mammalian genomes, in which the number of active tDNAs is still debated but estimated to be around 500 tDNAs (Tang et al. 2009), *Y. lipolytica* appears as a tDNA-rich eukaryote. However, given that *Y. lipolytica* has the longest genome among hemiascomycetous genomes sequenced so far (20 Mb), the average number of tDNAs per kilobase remains in fact roughly the same (one tDNA/40 kb), as for *S. cerevisiae* (12 Mb).

3.4.2 *Y. lipolytica* Departs from Other Hemiascomycetes to Read Leu CUN and Arg CGN Codons

In any organism, the genetic code requires a tRNA set able to read all the 62 sense codons (64 codons − 3 stop codons + 1 special initiator tRNA-Met). As proposed by the “Wobble hypothesis” (Crick 1966), a given tRNA may read more than a single codon (up to three in Eukarya and Archaea, even four in Bacteria). The total number of tRNA types (each type with a different anticodon) is therefore less than 62 and the deciphering of a number of genomes has shown that in fact no more than 46 different tRNA types exist in any organism. This number of tRNA *types* should not be confused with the total number of tRNA *genes*. *Y. lipolytica* features 510 tRNA genes representing only 42 different types (anticodon) of tRNA. Remarkably, Hemiascomycetes do not decode the Leu CUN and Arg CGN codons as do other eukaryotes, but rather as do Bacteria (Fig. 4) (Marck et al. 2006). *Y. lipolytica* and also *Pichia pastoris* (De Schutter et al. 2009) depart from other Hemiascomycetes and follows the eukaryotic rules. Other non-hemiascomycetous yeasts such as the Archiascomycete *S. pombe* as well as ascomycetous filamentous fungi (e.g., *Neurospora crassa* and *Fusarium graminearum*) also follow the regular eukaryotic rule. Yeasts of the “CTG” clade (Fig. 1) have their CTG codon changed from Leu to Ser and use a special decoding scheme to read their CTN codons (Fig. 4) (Massey et al. 2003). Nowhere else in eukaryotes have such switches from the eukaryotic to the bacterial type of decoding are known; why they have occurred is not understood but should be reconsidered in the light of the theory of the regressive evolution of yeasts (Dujon 2010).

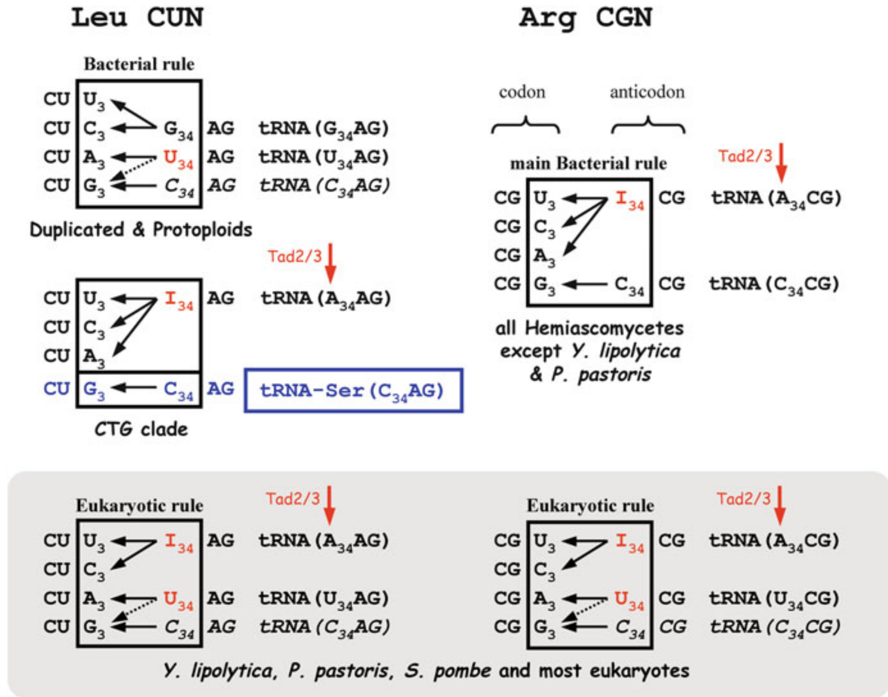


Fig. 4 Decoding of Leu CUN and Arg CGN in *Y. lipolytica*, Hemiascomycetes, and other eukaryotes. *Left*, decoding of Leu CUN codons; *right*, Arg CGN codons. Most eukaryotes read Leu CUN and Arg CGN codons as shown in the *gray box* at *bottom*. Remarkably, Hemiascomycetes use instead bacterial rules. However *Y. lipolytica* is an exception among Hemiascomycetes as it follows the standard eukaryotic rules. To read the Leu CUN codons, "duplicated" and "protoploids" yeasts use the regular bacterial rule, while yeasts of the "CTG" clade use a special decoding scheme with the CTG codon turned into a seventh Ser codon read by tRNA-Ser (CAG) (*boxed, in blue*). To read the Arg CGN codons, all Hemiascomycetes, except *Y. lipolytica*, use the main Bacterial rule. The *square boxes* symbolize the decoding of the four CUN or CGN codons with *black arrows* indicating which tRNA reads which codon(s). *Horizontal arrows* denote a Watson–Crick base pairing between the third base of the codon (at *right*) and the first base (numbered 34) of the anticodon (at *left*). Wobble base pairings are shown by *tilted black arrows* (e.g., G_{34}:U_3). All A_{34} starting anticodons are modified (*red arrows*) into I_{34} (inosine) by enzyme Tad2/3 (*S. cerevisiae* gene name) in order to read C-ending codons through I_{34}:C_3 base pairing. All U_{34}-tRNAs bear various modifications at 34 (shown as U_{34} in *red*) that restrict the reading to A_3-ending codon. If the cognate C_{34}-tRNA (in *italics*) is absent, then the U_{34}-tRNA has its U_{34} modified in such a way it can also read the G_3-ending codons through U_{34}:G_3 wobble base pairing (*dashed tilted arrow*). The bacterial typed-tRNA-Arg (A_{34}CG), with A_{34} modified into I_{34} also reads the rare CGA_3 codon in addition to CGU_3 and CGC_3. In such a situation the other tRNA reading the fourth G-ending codon has to be always present. Note the similarity between the decoding of Leu CUN in the "CTG" clade (*middle box* at *left*) and that of Arg CGN in all Hemiascomycetes except *Y. lipolytica* (*top box* at *right*)

4 The 5S Ribosomal RNA

The 5S rRNA, the smallest of the four ribosomal RNAs, is part of the large subunit of the ribosome. Whatever is the genomic organization of its genes (about hundred copies in most yeasts), the 5S rRNA requires for transcription, in addition to TFIIIC and -B, a specific zinc-finger transcription factor, TFIIIA (for review on zinc-finger proteins, see Klug (2010)).

4.1 TFIIIA: The 5S Ribosomal RNA Gene-Specific Factor

In all eukaryotes, transcription of the multicopy 5S rRNA genes is more complex than that of other Pol III genes. It requires a specific factor, TFIIIA, a nine-zinc-finger protein (Miller et al. 1985) that recognizes three internal promoter sequences referred to as A, I, and C (Fig. 2c). After binding of TFIIIA to each gene copy, the TFIIIA-5S rDNA complex is recognized by TFIIIC and transcription initiation then proceeds like it does in the case of other Pol III genes: TFIIIC binds TFIIIB upstream of the gene and TFIIIB recruits Pol III. TFIIIA specifically recognizes the 5S rRNA gene but is also able to bind the 5S rRNA, itself acting as a chaperone for nucleocytoplasmic transport and RNA storage (Guddat et al. 1990; Honda and Roeder 1980; Pelham and Brown 1980). TFIIIA is an essential protein having no other essential duty than 5S rRNA gene transcription; as a matter of fact, cells depleted of TFIIIA are viable if a single 5S rRNA gene copy is expressed by Pol III from a *RPR1* promoter (Camier et al. 1995).

4.2 Different Genomic Organization of the 5S Ribosomal RNA Genes

In Hemiascomycetes, the so-called rDNA units encoding the 35S rRNA gene are tandemly repeated about 100 times (Fig. 5) and localized at a single locus as in *S. cerevisiae* (about hundred copies on chromosome L) or at two loci, such as in *Candida glabrata* (two subtelomeric loci), or at more loci such as in *P. pastoris* (De Schutter et al. 2009) and *Y. lipolytica* (van Heerikhuizen et al. 1985). The pre-35S rRNA is processed into the three rRNAs 18S, 5.8S, and 25S. In many yeasts, the fourth rRNA, encoded by the 5S rRNA gene, is located inside each rDNA unit, 5′ from the 35S rRNA gene, and both genes are divergently transcribed by Pol III and Pol I, respectively (Paule and White 2000). In some genomes, such as *D. hansenii* (Lépingle et al. 2000) and related species (Nguyen et al. 2009) or *C. glabrata* (Dujon et al. 2004), two tandemly arranged 5S rRNA genes are present in each rDNA unit. The embedding of 5S rRNA genes into the rDNA units is not the general rule in eukaryotes: outside Hemiascomycetes, the 5S rRNA genes appear dispersed throughout the genomes. Together with *P. pastoris* which harbors 21 copies of dispersed 5S rRNA genes (De Schutter et al. 2009), *Y. lipolytica* is an exception among Hemiascomycetes as its

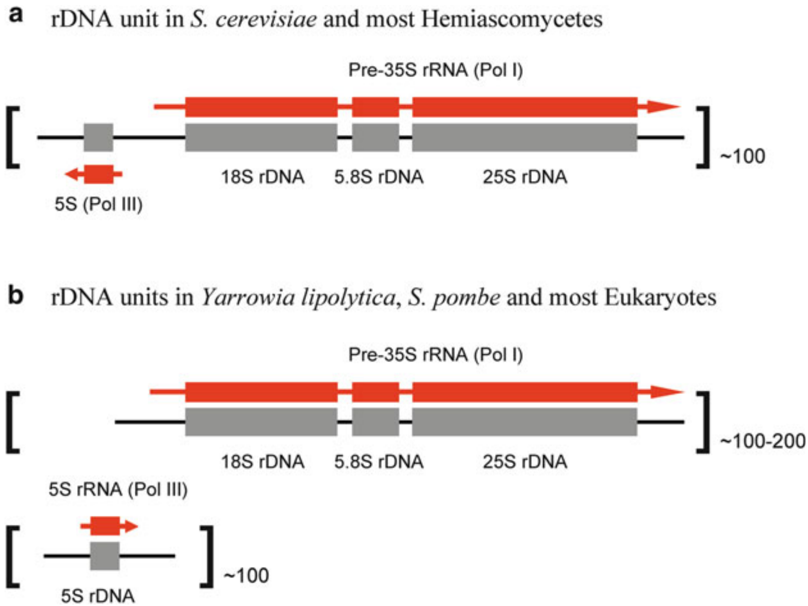


Fig. 5 The various genomic organization of 5S rRNA and other rRNA genes in *Y*. *lipolytica* and other Hemiascomycetes. (**a**) Organization in *S*. *cerevisiae* and most Hemiascomycetes except *Y*. *lipolytica*. The 5S rRNA and pre-35S rRNA are divergently transcribed by Pol III and Pol I, respectively. All units may be located within a single locus or separated in several loci often subtelomeric (for details, see Table 1 in Acker et al. (2008)). (**b**) The organization in *Y*. *lipolytica* is the same as in *S*. *pombe* as well as in most other eukaryotes (outside Hemiascomycetes). About 100 5S rRNA genes are dispersed in the genome. In *Y*. *lipolytica*, 48 out of the 108 functional 5S rRNA genes are unconventionally located 3′ to tRNA genes

117 copies of 5S rRNA gene are not included in its rDNA units (Clare et al. 1986; van Heerikhuizen et al. 1985) (Fig. 1). Among these 117 genes (Table 2B), eight appear to be incomplete and one interrupted by a Ylt1 transposable element. More surprisingly, among the remaining 108 full-length 5S rRNA genes dispersed in the genome, 48 occupy an unprecedented location as they lie immediately 3′ of tRNA genes suggesting that these combined genes might be transcribed together by Pol III in a single pass.

5 Dicistronic tRNA-5S rRNA Genes

Among the 108 complete copies of the 5S rRNA gene of *Y*. *lipolytica*, 48 are strangely located immediately 3′ of tRNA genes (Fig. 2d). The tRNAs implicated in these unprecedented Pol III gene assemblies appear not to be randomly distributed (Table 2C). For example, nearly all tRNA-Gly (GCC) (28 out of 30) are followed by a 5S rRNA gene. One or several copies of four other tRNA types are followed by 5S rRNA genes. One of the tandem tRNA genes present in three copies is followed

by a 5S rRNA gene forming a unique tricistronic gene tRNA-Lys (CTT)-tRNA-Glu (CTC)-5S rRNA gene.

Compared to other combined Pol III genes, these *Y. lipolytica* genes appear even stranger as tRNA and 5S rRNA genes do not use exactly the same transcription factors: 5S rRNA genes require, in addition to TFIIIC and TFIIIB, their specific factor TFIIIA. The interesting questions were then: (1) Are the 5S rRNA genes located 3′ to tRNA genes transcribed? (2) If transcribed, are they from the leading tRNA gene (yielding a dicistronic transcript) or from their internal promoters or, in other words, is TFIIIA required? (3) Are the products of both genes correctly expressed and matured?

5.1 Transcription of Dicistronic tRNA-5S rRNA Genes

These issues were experimentally investigated and Acker and coworkers first checked whether dicistronic transcripts were present by running RT-PCR experiments with total RNA extracted from *Y. lipolytica* (Acker et al. 2008). One copy of each tRNA type implicated in 5S rRNA fusions was investigated and composite transcripts of the predicted length were detected. Then, in vitro transcription by Pol III was investigated. For this, six fused genes bearing each a different tRNA type were cloned into pBluescript vectors and tested with two different transcription systems. One was "whole cell extract" (WCE) from either *Y. lipolytica* or *S. cerevisiae*; the other one was a nearly totally reconstituted system (with all subunits from TFIIIC and TFIIIB expressed in insect cells (Ducrot et al. 2006)). All dicistronic and the single tricistronic genes appeared to be correctly expressed and their products yielded the predicted lengths.

To answer the second question—is TFIIIA required?—the fact that TFIIIA is essential for 5S rRNA gene transcription was turned into an advantage. The *Y. lipolytica* gene coding for TFIIIA was deleted, thus preventing the expression of the 60 isolated 5S rRNA genes, and it was bet that *Y. lipolytica* might survive thanks to the expression of the 48 copies of tRNA-fused 5S rRNA genes expressed from the tRNA promoters (Table 2B, C). Strains deleted for TFIIIA genes were indeed truly viable attesting that 5S rRNA expressed from the only tRNA-5S rRNA genes are correctly spliced from the tRNA genes and produced at sufficient levels for viability. A slight phenotypic effect was observed concerning the cell morphology and the 5S rRNA was found to be slightly longer (a few nucleotides) than in the wild type strain. This last observation opens the interesting hypothesis that TFIIIA could participate in 5S rRNA maturation. Indeed, a comparable slight increase in 5S rRNA length was observed in the only other case of viable TFIIIA deletion reported so far (Camier et al. 1995).

The third question—correct expression and maturation—was answered only indirectly. All but two of the 30 tRNA-Gly (GCC) being followed by a 5S rRNA gene, tRNA-Gly (GCC) appeared, like the 5S rRNA genes, correctly expressed and matured (Acker et al. 2008).

5.2 *Specificity of* Y. lipolytica *TFIIIA*

TFIIIA is usually reported to contain nine zinc fingers but a more precise comparison throughout Hemiascomycetes and beyond suggested that in fact ten fingers are present, *S. pombe* being the only organism harboring all ten motifs. Finger 10 is not present in *Xenopus laevis*, while finger 9 is lacking in *S. cerevisiae* and *Y. lipolytica* (see Fig. 5 in Acker et al. (2008)). Interestingly, TFIIIA of *Y. lipolytica* differs from all other TFIIIA examined by an "11th" zinc finger and also an N-terminal extension of 133 residues. Whether these two unusual features are associated with any special properties of *Y. lipolytica* TFIIIA is not yet known.

6 Dicistronic tRNA-Trp/RUF70 Genes

Following the discovery of the tRNA-5S rRNA composite genes, the sequences located 3′ to all tRNA genes were examined (Acker et al. 2008). Remarkably, no Pol III terminators (poly-T) were observed on the 3′ side of all the 13 copies of the tRNA-Trp (CCA) gene. In fact, a terminator was found about 70 nt after the tRNA genes (Fig. 2e). No obvious conserved sequences appeared, but after automatic alignment and manual refinement, it appeared that these 3′ sequences were related by a common tertiary structure (see Fig. 2 in Acker et al. (2008)). A novel Pol III ncRNA was then suspected in *Y. lipolytica* and provisionally named RUF70 (RNA of unknown function about 70 nt-long).

6.1 *Transcription of Dicistronic tRNA-Trp/RUF70 Genes*

As for the tRNA-5S rRNA genes, successful RT-PCR experiments suggested the existence of dicistronic tRNA-Trp-RUF70 transcripts. Transcription of a cloned copy of one of the 13 tRNA-Trp-RU70 genes was also achieved with WCEs or a reconstituted Pol III system. Furthermore, RUF70 was detected by Northern blot. This new ncRNA gene was searched for without success in other yeasts, and its function remains yet unknown (Acker et al. 2008). A role of RUF70 in the maturation of the leader tRNA-Trp was considered but seemed unjustified because tRNA-Trp is nearly invariant throughout Hemiascomycetes and RUF70 is not present outside *Y. lipolytica,* i.e., in any other yeast whose genome is fully sequenced.

7 Other Pol III ncRNAs

In Hemiascomycetes, a few other ncRNA genes, whose RNA products have unrelated functions, are also transcribed by Pol III. Their sequences remain close enough to be discoverable with classic Blast software (Dujon et al. 2004). Throughout Hemiascomycetes, these few genes share the common features of a single copy,

A and B box promoter sequences (with a helper upstream TATA element in a single case), a short length compared to Pol II genes (however, with one remarkable exception), and a lack of intron.

7.1 A Double Constraint Is Exerted on Internal Promoters

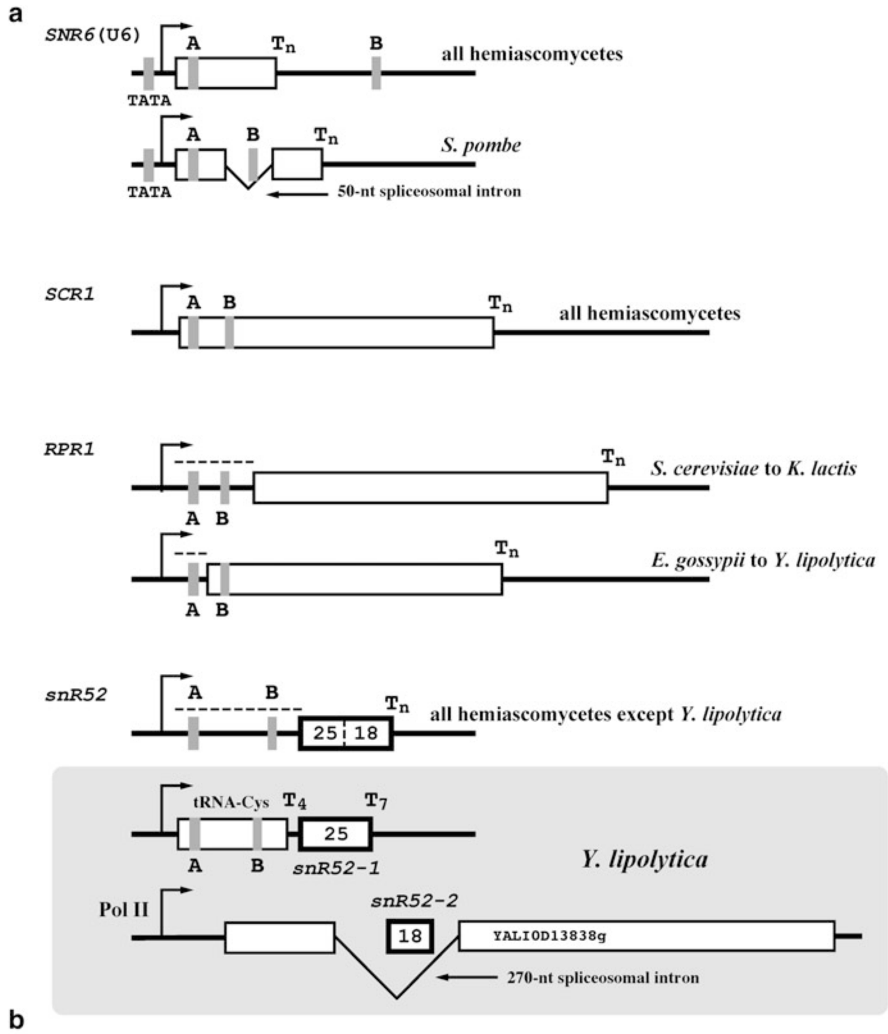
A key difference with tRNA transcription is that the A and B promoter sequences are still internal to the primary transcript; however, the RNA regions encoded by these promoter sequences do not adopt the same 3D structures as that of tRNAs. The RNA sequence encoded by the A box does not fold into a D-stem and D-loop, nor that of the B box into a T-loop. To resolve this conflict between an unconstrained 3D structure and the benefit of the A and B Pol III promoters (bound by TFIIIC) which have to remain internal to the transcript, some genes use in turn an A- and B-containing leader sequence which is posttranscriptionally cleaved from the final matured RNA product. Most of the experimental work carried out on these few genes was performed in the yeast *S. cerevisiae*, and for most other organisms, the location of the promoter sequences is only supported by comparative sequence analyses.

7.2 Spliceosomal U6 RNA

U6 RNA is the only spliceosomal RNA produced by Pol III, however, with an unusual promoter organization (its gene is named *SNR6*, Fig. 6a). As usual, the A box is internal to the transcript, but the B box is located about 100 nt downstream the Pol III terminator (Brow and Guthrie 1990). An upstream TATA element is also present in all sequenced Hemiascomycetes, including *Y. lipolytica*. In *S. cerevisiae*, it was experimentally verified that, thanks to the upstream TATA element, TFIIIB (that contains TBP) can directly bind in vitro without TFIIIC and allow transcription. Nevertheless, TFIIIC is required in vivo to relieve repression by chromatin (Burnol et al. 1993). This promoter organization is conserved throughout Hemiascomycetes, including *Y. lipolytica* (Marck et al. 2006). Interestingly, the *S. pombe* U6 RNA gene is the only Pol III gene known so far to bear a spliceosomal intron (Tani and Ohshima 1989). U6 RNA is also produced by Pol III in mammals with predominant upstream regulating elements (Kunkel and Pederson 1989).

7.3 RPR1: The RNA Component of RNase P

RPR1 is the only Pol III-produced RNA that participates in the processing of other Pol III products. It is the RNA component of RNase P that matures the 5′ end of pre-tRNAs (Lee et al. 1991). A remarkable length exception was discovered in the genome of the human pathogen *Candida glabrata* where the RPR1 RNA gene reaches an outstanding length of 1,149 nt (Dujon et al. 2004; Kachouri et al. 2005). When comparing RPR1 genes within Hemiascomycetes (Fig. 6a; see also alignment in Marck et al. (2006)),



***Y. lipolytica* genes**

genes	direction		mature product boundaries		length	A-box pos.		TRGYnnAnnnG		Δ A-B	B-box pos.		GWTCRAnnC	
SNR6(U6)	<	F	1089094	1089197	104	+21	A	TGGTcaAtttG	A	212	+244	A	GTaCAAtaC	T
SCR1	<	A	176311	176581	271	+18	T	TGtCggAgtgG	T	21	+50	C	GTTCGAgtC	T
SCR2	>	D	819163	819433	276	+18	T	TGtCggAgtgG	T	21	+50	C	GTTCGAgtt	C
RPR1	<	E	2070712	2070999	288	-28	C	TGGCctAacgG	T	68	+52	A	GTTCGAatC	C
snR52-1	>	D	677142	677224	92	-75	A	TGGCacAgcgG	T	33	-31	A	GTTCGAatC	T

Fig. 6 Variety of Pol III promoter organization in U6, RPR1, SCR1, and snR52 RNA genes in *Y. lipolytica* and other Hemiascomycetes. (**a**) Schematic representation of four Pol III genes illustrating the variable positions of A and B boxes promoter sequences found in nine

two variants of promoter organization emerge: in *S. cerevisiae*, *Saccharomyces castellii*, *C. glabrata*, *Lachancea waltii*, and *K. lactis*, both A and B promoters are included in a leader sequence whose transcript is separated from the mature RPR1 (Fig. 1). The RPR1 RNAs of *Eremothecium gossypii*, *D. hansenii*, *Candida albicans*, and *Y. lipolytica* can accommodate a B box internal to the DNA region encoding themselves (Marck et al. 2006).

7.4 7SL/scR1: The RNA Component of the Signal Recognition Particle

The SRP targets presecretory and membrane proteins to the endoplasmic reticulum (ER) (Dieci et al. 2002). The evolutionary conserved SRP is made of an RNA backbone (the 7SL/scR1 RNA encoded by the *SCR1* gene) and of 6–7 protein subunits. All examined 7SL RNAs from Hemiascomycetes can accommodate truly internal A and B sequences (Fig. 6a; see also alignment in Marck et al. (2006)). As a matter of fact, the A and B sequences are located in unpaired regions of the 3D structure. The location of these sequences, estimated by comparative sequence analysis, is compatible with experimental mutational analysis (Dieci et al. 2002). In *Y. lipolytica*, the 7SL RNA is encoded by two nearly identical copies, *YlSCR1* and *YlSCR2*. Deletions of either *SCR1* or *SCR2* are viable in *Y. lipolytica*, but deletion of both has been reported to be lethal. Systematic substitution of the two conserved nucleotides G128 and A130 in loop 1 of *SCR2*-encoded 7SL RNA resulted in thermosensitive growth phenotypes and demonstrated that *YlSCR*

Fig. 6 (continued) Hemiascomycetes (*S. cerevisiae*, *S. castellii*, *C. glabrata*, *L. waltii*, *K. lactis*, *E. gossypii*, *D. hansenii*, *C. albicans*, and *Y. lipolytica*; see Fig. 6 in Marck et al. (2006) for details). These promoters are always internal to the transcript but may be external to DNA region encoding the mature product as these sequences (*dashed lines* in *RPR1* and *snR52*) can be posttranscriptionally cleaved. *Arrows* show the transcription start site (about 15 nt upstream of A box); "Tn" indicates the poly-T Pol III terminator. In SNR6 (U6) gene, an upstream TATA element (bound by factor TFIIIB) is present. The U6 RNA gene of *S. pombe* is additionally shown; it bears the only spliceosomal intron known to interrupt a Pol III gene. In the RPR1 gene, the B box is internal to the DNA region encoding the matured RNA product in *E. gossypii*, *D. hansenii*, *C. albicans*, and *Y. lipolytica*, while both A and B boxes are external in other yeasts. The snR52 gene of *Y. lipolytica* is split into two genes; the first one, snR52-1 (labeled "25"), responsible for a modification in 25S rRNA is located downstream of a tRNA-Cys gene; the second one, snR52-2 (labeled "18"), is a Pol II-expressed gene responsible for a modification in 18S rRNA and is located within a spliceosomal intron belonging to the protein gene YALI0D13838g. (**b**) Promoter sequences of the *Y. lipolytica* genes. Positions of the A and B boxes are given with respect to the first nucleotide of the mature product (numbered +1). Positive coordinate indicates that the promoter sequence is located inside the DNA region corresponding to the mature product, negative coordinate that it is located in a leader sequence posttranscriptionally cleaved (*dashed lines* in *RPR1* and *snR52*). "ΔA-B" indicates the number of nucleotides separating the A and B boxes. Nucleotides corresponding to "n" (any base) in the consensus sequences are written with *lowercase letters*, as well as nucleotides missing the consensus. Nucleotides preceding and following the A and B promoter sequences are also shown to better enhance the actual boundaries of the DNA sequences recognized by TFIIIC

genes were indeed involved in protein secretion (He et al. 1989). Two identical copies are also found in *S. castellii* (Dujon et al. 2004). In mammals, the corresponding gene is present in several copies and upstream control elements regulate the transcription in addition to A and B promoters (Englert et al. 2004).

7.5 snR52 RNA: The Only Pol III-Produced snoRNA

This RNA was formerly believed to be a Pol II product, as are other sn- and snoRNAs (about 70 in *S. cerevisiae*) until a genome-wide mapping analysis of the Pol III machinery was performed in *S. cerevisiae* (Harismendy et al. 2003). In all Hemiascomycetes but *Y. lipolytica*, the A and B sequences are located within a leader sequence which is cleaved from the mature snR52 product (Guffanti et al. 2006) (Fig. 6). The snR52 RNA is a C/D box snoRNA responsible, in *S. cerevisiae*, for two guided 2′-*O*-methylations in the 25S rRNA (at position U2921 in the 25S rRNA of *S. cerevisiae*) and 18S rRNA (position A420) (Lowe and Eddy 1999). Around these positions, the sequence of both 25S and 18S rRNA genes are highly conserved within Hemiascomycetes and beyond; therefore, the guiding sequences (11 nt each) within *snR52* are also conserved. This allows an easy tracking of snR52 in foreign genomes.

7.5.1 A Split snR52 RNA Gene in *Y. lipolytica*

In the present work, we reexamined and corrected the promoter sequences previously determined as being those of the *Y. lipolytica* snR52 gene (Marck et al. 2006). These perfect A and B sequences (Fig. 6a, b) belong in fact to one of the four tRNA-Cys genes present in the genome (Table 3). Therefore, a novel third possible type of combined Pol III gene might be present in *Y. lipolytica* (in addition to tRNA-5S rRNA and tRNA-Trp-RUF70 genes). We also reexamined the sequence of the *Y. lipolytica* snR52 RNA gene with respect to that of other Hemiascomycetous yeasts and observed that the sequence element targeting the modification of 25S rRNA was present alone. The sequence targeting the other modification, in the 18S rRNA, was searched for in the *Y. lipolytica* genome and found located within a spliceosomal intron interrupting the protein gene YALI0D13838g which is likely to be a Pol II transcript. Such a situation is common, a number of snoRNAs are intron-encoded Pol II genes (Lowe and Eddy 1999).

For clarity, we now refer to the two separated parts of the *Y. lipolytica* “snR52 RNA gene” as snR52-1 and snR52-2 (Fig. 6a). Whether the combined tRNA-Cys-snR52-1 is a true dicistronic gene expressed as a single primary transcript remains an open question that deserves experimental investigation. We observed the presence of a four-T sequence between the tRNA-Cys gene and snR52-1 RNA gene, while the final Pol III terminator is made up of a seven-T sequence. It may be that the four-T sequence does not act as a strong Pol III terminator as the strength of a poly-T terminator may be largely influenced by the context (Braglia et al. 2005).

Such poly-T sequences do not prevent Pol III from moving further downstream (Guffanti et al. 2006). This sequence might also acts as a weak terminator in order to lower the expression of the snR52-1 gene with respect to that of the leader tRNA-Cys gene.

Our preliminary investigations suggest that snR52-1 is absent from *S. pombe* and therefore the 2′-*O*-methylation at position equivalent to U2921 of *S. cerevisiae* in 25S rRNA does not take place or is performed enzymatically without the help of a guiding C/D box snoRNA. A Pol II gene bearing the 18S guiding sequence, not located within an intron, seems to be present in *S. pombe*.

7.6 Conservation of A and B Pol III Promoter Sequences in Hemiascomycetes

The promoter sequence of the four genes coding for U6, RPR1, 7SL/scR1, and snR52 RNAs were previously identified in nine Hemiascomycetes and shown to obey the consensus determined from the comparative analyses of the tRNA gene sequences (Fig. 3b), although with some mistakes (Marck et al. 2006). The A and B promoters sequences of the four genes in *Y. lipolytica* (shown in Fig. 6b) display such "mistakes," e.g., the conserved G at the third position of the A box is missing in the 7SL RNA genes (*SCR1* and *SCR2*) and the T at the third position of the B box is missing in the U6 RNA gene (*SNR6*). How factor TFIIIC accommodates these deviations remains unclear as nothing is known at the molecular level about how TFIIIC recognizes the A and B boxes. Excepting the snR52 RNA gene, the three other genes considered here remain Pol III dependent for their transcription from yeasts to mammals, probably to benefit from the high expression level of Pol III transcription and possibly from its regulation.

8 Unique Pol III Features Confirm *Y. lipolytica* as a "Different Hemiascomycetous Yeast"

Exploration of *Y. lipolytica* at the genomic level started more than 10 years ago (Casaregola et al. 2000; Souciet et al. 2000). Achievement of its complete genome in 2004 confirmed its status as distantly related to *S. cerevisiae* (Dujon et al. 2004). Both in silico analyses (Marck et al. 2006) and further experimental work on *Y. lipolytica* (Acker et al. 2008) confirmed this yeast differs in many respects from other Hemiascomycetes: 5S rRNA genes not embedded into the rDNA unit; decoding of Leu CUN and Arg CGN codons similar to other eukaryotes and not similar to that of other Hemiascomycetes; and split snR52 gene, dicistronic tRNA-5S rRNA, and tRNA-Trp-RUF70 genes not found anywhere.

Roughly, the hemiascomycetous tree (Fig. 1) can be subdivided into (1) the "duplicated" clade harboring the model yeast *S. cerevisiae* and other species known as *Saccharomyces* sensu stricto *and Saccharomyces sensu lato*, recently renamed

Nakaseomyces, *Kazachstania*, and *Naumovia* (Kurtzman 2003) whose ancestor underwent a whole genome duplication; (2) the “protoploid” clade as defined by the Génolevures consortium (Souciet et al. 2009) including the *Kluyveromyces*, the *Zygosaccharomyces*, and the *Lachancea* clades; and (3) the “CTG” clade where this codon was converted from Leu to Ser and the rest of the Hemiascomycetes which corresponds to very few sequenced genomes (e.g., *Y. lipolytica* and *P. pastoris*) compared to the number of clades still unexplored. The representation of the phylogenetic tree of Hemiascomycetes through sequenced genomes is indeed highly biased. Therefore, the specific Pol III gene-related properties listed above that render *Y. lipolytica* so different from other Hemiascomycetes are thought to be mainly due to its remote key phylogenetic location. Indeed, *Y. lipolytica* is as different from *S. cerevisiae* as are ascidians from human (Dujon 2006). Assuming that we lack phylogenetic intermediates in the neighborhood of *Y. lipolytica*, the sequencing of more genomes within the *Yarrowia* clade will allow us to determine whether all the “break points” we saw between *Y. lipolytica* and the rest of Hemiascomycetes occurred all at once as a unique genomic “big bang” or much more progressively one after the other.

Acknowledgements We wish to thank Joël Acker for helpful discussions, Carl Mann for correcting the manuscript, and all our colleagues from the Génolevures consortium. This work was supported by the GDR CNRS 2354 “Génolevures-3,” the ANR grant “Genarise” (ANR-05-BLAN-0331), and the ANR grant “RegPolStress” (ANR-07-BLAN-0039).

References

Abe T, Ikemura T, Ohara Y, Uehara H, Kinouchi M, Kanaya S, Yamada Y, Muto A, Inokuchi H (2009) tRNADB-CE: tRNA gene database curated manually by experts. Nucleic Acids Res 37: D163–D168

Acker J, Ozanne C, Kachouri-Lafond R, Gaillardin C, Neuvéglise C, Marck C (2008) Dicistronic tRNA-5S rRNA genes in *Yarrowia lipolytica*: an alternative TFIIIA-independent way for expression of 5S rRNA genes. Nucleic Acids Res 36:5832–5844

Baker RE, Camier S, Sentenac A, Hall BD (1987) Gene size differentially affects the binding of yeast transcription factor τ to two intragenic regions. Proc Natl Acad Sci USA 84:8768–8772

Belfort M, Weiner A (1997) Another bridge between kingdoms: tRNA splicing in archaea and eukaryotes. Cell 89:1003–1006

Borchert GM, Lanier W, Davidson BL (2006) RNA polymerase III transcribes human microRNAs. Nat Struct Mol Biol 13:1097–1101

Braglia P, Percudani R, Dieci G (2005) Sequence context effects on oligo(dT) termination signal recognition by *Saccharomyces cerevisiae* RNA polymerase III. J Biol Chem 280:19551–19562

Brow DA, Guthrie C (1990) Transcription of a yeast U6 snRNA gene requires a polymerase III promoter element in a novel position. Genes Dev 4:1345–1356

Burnol A-F, Margottin F, Huet J, Almouzni G, Prioleau M-N, Mechali M, Sentenac A (1993) TFIIIC relieves repression of U6 snRNA transcription by chromatin. Nature 362:475–477

Camier S, Dechampesme AM, Sentenac A (1995) The only essential function of TFIIIA in yeast is the transcription of 5S rRNA genes. Proc Natl Acad Sci USA 92:9338–9342

Casaregola S, Neuvéglise C, Lepingle A, Bon E, Feynerol C, Artiguenave F, Wincker P, Gaillardin C (2000) Genomic exploration of the hemiascomycetous yeasts: 17. *Yarrowia lipolytica*. FEBS Lett 487:95–100

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540–552

Chan PP, Lowe TM (2009) GtRNAdb: a database of transfer RNA genes detected in genomic sequence. Nucleic Acids Res 37:D93–D97

Ciesla M, Towpik J, Graczyk D, Oficjalska-Pham D, Harismendy O, Suleau A, Balicki K, Conesa C, Lefebvre O, Boguta M (2007) Maf1 is involved in coupling carbon metabolism to RNA polymerase III transcription. Mol Cell Biol 27:7693–7702

Clare JJ, Davidow LS, Gardner DC, Oliver SG (1986) Cloning and characterisation of the ribosomal RNA genes of the dimorphic yeast, *Yarrowia lipolytica*. Curr Genet 10:449–452

Crick FH (1966) Codon–anticodon pairing: the wobble hypothesis. J Mol Biol 19:548–555

De Schutter K, Lin YC, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, Rouze P, Van de Peer Y, Callewaert N (2009) Genome sequence of the recombinant protein production host *Pichia pastoris*. Nat Biotechnol 27:561–566

Dieci G, Sentenac A (1996) Facilitated recycling pathway for RNA polymerase III. Cell 84: 245–252

Dieci G, Percudani R, Giuliodori S, Bottarelli L, Ottonello S (2000) TFIIIC-independent in vitro transcription of yeast tRNA genes. J Mol Biol 299:601–613

Dieci G, Giuliodori S, Catellani M, Percudani R, Ottonello S (2002) Intragenic promoter adaptation and facilitated RNA polymerase III recycling in the transcription of SCR1, the 7SL RNA gene of *Saccharomyces cerevisiae*. J Biol Chem 277:6903–6914

Dieci G, Fiorino G, Castelnuovo M, Teichmann M, Pagano A (2007) The expanding RNA polymerase III transcriptome. Trends Genet 23:614–622

Donze D, Adams CR, Rine J, Kamakaka RT (1999) The boundaries of the silenced HMR domain in *Saccharomyces cerevisiae*. Genes Dev 13:698–708

Ducrot C, Lefebvre O, Landrieux E, Guirouilh-Barbat J, Sentenac A, Acker J (2006) Reconstitution of the yeast RNA polymerase III transcription system with all recombinant factors. J Biol Chem 281:11685–11692

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

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

Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuvéglise C, Talla E, Goffard N, Frangeul L, Aigle M, Anthouard V, Babour A, Barbe V, Barnay S, Blanchin S, Beckerich JM, Beyne E, Bleykasten C, Boisrame A, Boyer J, Cattolico L, Confanioleri F, De Daruvar A, Despons L, Fabre E, Fairhead C, Ferry-Dumazet H, Groppi A, Hantraye F, Hennequin C, Jauniaux N, Joyet P, Kachouri R, Kerrest A, Koszul R, Lemaire M, Lesur I, Ma L, Muller H, Nicaud JM, Nikolski M, Oztas S, Ozier-Kalogeropoulos O, Pellenz S, Potier S, Richard GF, Straub ML, Suleau A, Swennen D, Tekaia F, Wesolowski-Louvel M, Westhof E, Wirth B, Zeniou-Meyer M, Zivanovic I, Bolotin-Fukuhara M, Thierry A, Bouchier C, Caudron B, Scarpelli C, Gaillardin C, Weissenbach J, Wincker P, Souciet JL (2004) Genome evolution in yeasts. Nature 430:35–44

Dumay-Odelot H, Marck C, Durrieu-Gaillard S, Lefebvre O, Jourdain S, Prochazkova M, Pflieger A, Teichmann M (2007) Identification, molecular cloning, and characterization of the sixth subunit of human transcription factor TFIIIC. J Biol Chem 282:17179–17189

Engelke DR, Gegenheimer P, Abelson J (1985) Nucleolytic processing of a tRNAArg-tRNAAsp dimeric precursor by a homologous component from *Saccharomyces cerevisiae*. J Biol Chem 260:1271–1279

Englert M, Felis M, Junker V, Beier H (2004) Novel upstream and intragenic control elements for the RNA polymerase III-dependent transcription of human 7SL RNA genes. Biochimie 86: 867–874

Esakova O, Krasilnikov AS (2010) Of proteins and RNA: the RNase P/MRP family. RNA 16: 1725–1747

Ferrari R, Rivetti C, Acker J, Dieci G (2004) Distinct roles of transcription factors TFIIIB and TFIIIC in RNA polymerase III transcription reinitiation. Proc Natl Acad Sci USA 3:13442–13447

Geiduschek EP, Kassavetis GA (2001) The RNA polymerase III transcription apparatus. J Mol Biol 310:1–26

Ghavi-Helm Y, Michaut M, Acker J, Aude JC, Thuriaux P, Werner M, Soutourina J (2008) Genome-wide location analysis reveals a role of TFIIS in RNA polymerase III transcription. Genes Dev 22:1934–1947

Grosjean H, de Crecy-Lagard V, Marck C (2010) Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. FEBS Lett 584: 252–264

Guddat U, Bakken AH, Pieler T (1990) Protein-mediated nuclear export of RNA: 5S rRNA containing small RNPs in xenopus oocytes. Cell 60:619–628

Guffanti E, Percudani R, Harismendy O, Soutourina J, Werner M, Iacovella MG, Negri R, Dieci G (2006) Nucleosome depletion activates poised RNA polymerase III at unconventional transcription sites in *Saccharomyces cerevisiae*. J Biol Chem 281:29155–29164

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704

Haldar D, Kamakaka RT (2006) tRNA genes as chromatin barriers. Nat Struct Mol Biol 13: 192–193

Harismendy O, Gendrel CG, Soularue P, Gidrol X, Sentenac A, Werner M, Lefebvre O (2003) Genome-wide location of yeast RNA polymerase III transcription machinery. EMBO J 22: 4738–4747

He F, Beckerich JM, Ribes V, Tollervey D, Gaillardin CM (1989) Two genes encode 7SL RNAs in the yeast *Yarrowia lipolytica*. Curr Genet 16:347–350

Holley RW (1965) Structure of an alanine transfer ribonucleic acid. JAMA 194:868–871

Holley RW, Everett GA, Madison JT, Zamir A (1965) Nucleotides sequences in the yeast alanine transfer ribonucleic acid. J Biol Chem 240:2122–2128

Honda BM, Roeder RG (1980) Association of a 5S gene transcription factor with 5S RNA and altered levels of the factor during cell differentiation. Cell 22:119–126

Hopper AK, Pai DA, Engelke DR (2010) Cellular dynamics of tRNAs and their genes. FEBS Lett 584:310–317

Juhling F, Morl M, Hartmann RK, Sprinzl M, Stadler PF, Putz J (2009) tRNAdb 2009: compilation of tRNA sequences and tRNA genes. Nucleic Acids Res 37:D159–D162

Kachouri R, Stribinskis V, Zhu Y, Ramos KS, Westhof E, Li Y (2005) A surprisingly large RNase P RNA in *Candida glabrata*. RNA 11:1064–1072

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066

Klug A (2010) The discovery of zinc fingers and their applications in gene regulation and genome manipulation. Annu Rev Biochem 79:213–231

Kruszka K, Barneche F, Guyot R, Ailhas J, Meneau I, Schiffer S, Marchfelder A, Echeverria M (2003) Plant dicistronic tRNA-snoRNA genes: a new mode of expression of the small nucleolar RNAs processed by RNase Z. EMBO J 22:621–632

Kunkel GR, Pederson T (1989) Transcription of a human U6 small nuclear RNA gene in vivo withstands deletion of intragenic sequences but not of an upstream TATATA box. Nucleic Acids Res 17:7371–7379

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

Lee JY, Rohlman CE, Molony LA, Engelke DR (1991) Characterization of RPR1, an essential gene encoding the RNA component of *Saccharomyces cerevisiae* nuclear RNase P. Mol Cell Biol 11:721–730

Lépingle A, Casaregola S, Neuvéglise C, Bon E, Nguyen H, Artiguenave F, Wincker P, Gaillardin C (2000) Genomic exploration of the hemiascomycetous yeasts: 14. *Debaryomyces hansenii* var. *hansenii*. FEBS Lett 487:82–86

Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964

Lowe TM, Eddy SR (1999) A computational screen for methylation guide snoRNAs in yeast. Science 283:1168–1171

Marck C, Grosjean H (2002) tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. RNA 8:1189–1232

Marck C, Grosjean H (2003) Identification of BHB splicing motifs in intron-containing tRNAs from 18 archaea: evolutionary implications. RNA 9:1516–1531

Marck C, Kachouri-Lafond R, Lafontaine I, Westhof E, Dujon B, Grosjean H (2006) The RNA polymerase III-dependent family of genes in hemiascomycetes: comparative RNomics, decoding strategies, transcription and evolutionary implications. Nucleic Acids Res 34:1816–1835

Massey SE, Moura G, Beltrao P, Almeida R, Garey JR, Tuite MF, Santos MA (2003) Comparative evolutionary genomics unveils the molecular mechanism of reassignment of the CTG codon in *Candida* spp. Genome Res 13:544–557

Miller J, McLachlan AD, Klug A (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. EMBO J 4:1609–1614

Nguyen HV, Gaillardin C, Neuvéglise C (2009) Differentiation of *Debaryomyces hansenii* and *Candida famata* by rRNA gene intergenic spacer fingerprinting and reassessment of phylogenetic relationships among *D. hansenii*, *C. famata*, *D. fabryi*, *C. flareri* (=*D. subglobosus*) and *D. prosopidis*: description of *D. vietnamensis* sp. nov. closely related to *D. nepalensis*. FEMS Yeast Res 9:641–662

Oficjalska-Pham D, Harismendy O, Smagowicz WJ, Gonzalez de Peredo A, Boguta M, Sentenac A, Lefebvre O (2006) General repression of RNA polymerase III transcription is triggered by protein phosphatase type 2A-mediated dephosphorylation of Maf1. Mol Cell 22: 623–632

Paule MR, White RJ (2000) Survey and summary: transcription by RNA polymerases I and III. Nucleic Acids Res 28:1283–1298

Pelham HRB, Brown DD (1980) A specific transcription factor that can bind either the 5S RNA gene or 5S RNA. Proc Natl Acad Sci USA 77:4170–4174

Pfeffer S, Sewer A, Lagos-Quintana M, Sheridan R, Sander C, Grasser FA, van Dyk LF, Ho CK, Shuman S, Chien M, Russo JJ, Ju J, Randall G, Lindenbach BD, Rice CM, Simon V, Ho DD, Zavolan M, Tuschl T (2005) Identification of microRNAs of the herpesvirus family. Nat Methods 2:269–276

Phizicky EM, Hopper AK (2010) tRNA biology charges to the front. Genes Dev 24:1832–1860

Reyes VM, Newman A, Abelson J (1986) Mutational analysis of the coordinate expression of the yeast tRNAArg-tRNAAsp gene tandem. Mol Cell Biol 6:2436–2442

Roberts DN, Stewart AJ, Huff JT, Cairns BR (2003) The RNA polymerase III transcriptome revealed by genome-wide localization and activity-occupancy relationships. Proc Natl Acad Sci USA 100:14695–14700

Schramm L, Hernandez N (2002) Recruitment of RNA polymerase III to its target promoters. Genes Dev 16:2593–2620

Schultz P, Marzouki N, Marck C, Ruet A, Oudet P, Sentenac A (1989) The two DNA-binding domains of yeast transcription factor tau as observed by scanning transmission electron microscopy. EMBO J 8:3815–3824

Schurer H, Schiffer S, Marchfelder A, Morl M (2001) This is the end: processing, editing and repair at the tRNA 3′-terminus. Biol Chem 382:1147–1156

Soma A, Onodera A, Sugahara J, Kanai A, Yachie N, Tomita M, Kawamura F, Sekine Y (2007) Permuted tRNA genes expressed via a circular RNA intermediate in *Cyanidioschyzon merolae*. Science 318:450–453

Souciet J, Aigle M, Artiguenave F, Blandin G, Bolotin-Fukuhara M, Bon E, Brottier P, Casaregola S, de Montigny J, Dujon B, Durrens P, Gaillardin C, Lepingle A, Llorente B, Malpertuy A, Neuvéglise C, Ozier-Kalogeropoulos O, Potier S, Saurin W,

Tekaia F, Toffano-Nioche C, Wesolowski-Louvel M, Wincker P, Weissenbach J (2000) Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies. FEBS Lett 487:3–12

Souciet JL, Dujon B, Gaillardin C, Johnston M, Baret PV, Cliften P, Sherman DJ, Weissenbach J, Westhof E, Wincker P, Jubin C, Poulain J, Barbe V, Segurens B, Artiguenave F, Anthouard V, Vacherie B, Val ME, Fulton RS, Minx P, Wilson R, Durrens P, Jean G, Marck C, Martin T, Nikolski M, Rolland T, Seret ML, Casaregola S, Despons L, Fairhead C, Fischer G, Lafontaine I, Leh V, Lemaire M, de Montigny J, Neuvéglise C, Thierry A, Blanc-Lenfle I, Bleykasten C, Diffels J, Fritsch E, Frangeul L, Goeffon A, Jauniaux N, Kachouri-Lafond R, Payen C, Potier S, Pribylova L, Ozanne C, Richard GF, Sacerdot C, Straub ML, Talla E (2009) Comparative genomics of protoploid *Saccharomycetaceae*. Genome Res 19:1696–1709

Sprague KU (1995) Transcription of eukaryotic tRNA genes. In: Söll D, RajBhandary U (eds) tRNA: structure, biosynthesis, and function. ASM, Washington, DC, pp 31–50

Tang DT, Glazov EA, McWilliam SM, Barris WC, Dalrymple BP (2009) Analysis of the complement and molecular evolution of tRNA genes in cow. BMC Genomics 10:188

Tani T, Ohshima Y (1989) The gene for the U6 small nuclear RNA in fission yeast has an intron. Nature 337:87–90

Tavenet A, Suleau A, Dubreuil G, Ferrari R, Ducrot C, Michaut M, Aude JC, Dieci G, Lefebvre O, Conesa C, Acker J (2009) Genome-wide location analysis reveals a role for Sub1 in RNA polymerase III transcription. Proc Natl Acad Sci USA 106:14265–14270

Trotta CR, Miao F, Arn EA, Stevens SW, Ho CK, Rauhut R, Abelson JN (1997) The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. Cell 89:849–858

van Heerikhuizen H, Ykema A, Klootwijk J, Gaillardin C, Ballas C, Fournier P (1985) Heterogeneity in the ribosomal RNA genes of the yeast *Yarrowia lipolytica*; cloning and analysis of two size classes of repeats. Gene 39:213–222

White RJ (1998) RNA polymerase III transcription, 2nd edn. Springer/Landes Bioscience, New York

Willis I, Hottinger H, Pearson D, Chisholm V, Leupold U, Söll D (1984) Mutations affecting excision of the intron from eukaryotic dimeric tRNA precursor. EMBO J 3:1573–1580

Utilization of Hydrophobic Substrate by *Yarrowia lipolytica*

Ryouichi Fukuda and Akinori Ohta

Abstract Utilization of *n*-alkane and its oxidized derivatives as sources of carbon and energy is one of the important properties that characterize the yeast *Yarrowia lipolytica* and has been a subject of intensive study. We describe a progress that has been made in the last decade, mainly on the transcriptional control of *ALK* genes encoding 12 CYP52 family cytochromes P450ALK that work in initial hydroxylation of *n*-alkanes. Among 12 *ALK* genes, *ALK1* is most prominently expressed upon the addition of *n*-alkanes, where two positive regulators, Yas1p and Yas2p, bind together to the promoter element ARE1 and highly activate transcription of *ALK1*. The control by *n*-alkane depends on the localization of a negative regulator Yas3p that binds to Yas2p in nucleus when *n*-alkanes are absent but remains binding to ER when *n*-alkanes are present in culture medium.

Contents

1 Utilization of Fatty Acids 112
  1.1 Uptake and Activation of Fatty Acids 112
2 Utilization of *n*-Alkanes 113
  2.1 Incorporation of *n*-Alkanes 114
  2.2 Terminal Hydroxylation of *n*-Alkanes by Cytochrome P450 114
  2.3 Further Hydroxylation of Long-Chain Alcohols to Fatty Acids 115
  2.4 Transcriptional Control of Metabolism of *n*-Alkanes 116
References 118

R. Fukuda • A. Ohta (✉)
Department of Biotechnology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan
e-mail: afukuda@mail.ecc.u-tokyo.ac.jp; aaohta@mail.ecc.u-tokyo.ac.jp

G. Barth (ed.), *Yarrowia lipolytica*, Microbiology Monographs 24,
DOI 10.1007/978-3-642-38320-5_5, © Springer-Verlag Berlin Heidelberg 2013

1 Utilization of Fatty Acids

Among the substrates that *Y. lipolytica* is able to assimilate, fatty acids are favorable energy and carbon source for this a bit strange nonconventional yeast (Barth and Gaillardin 1997). This property has collected interests of a certain group of microbiologists, especially of those undertaking its industrial application. However, decades after the beginning of analysis, our knowledge on how it takes up and metabolizes fatty acids, and how it organizes and controls the flow of fatty acid metabolites for cellular activity, is yet far under the level of satisfaction. The latter aspects of knowledge are of increasing importance in development of *Y. lipolytica* for industrial application. In this section, our current understandings on fatty acid uptake and acyl-CoA formation by *Y. lipolytica* are summarized.

1.1 Uptake and Activation of Fatty Acids

Yarrowia lipolytica grows well on fatty acid as a sole carbon source. Fatty acid uptake by *Saccharomycopsis lipolytica* (*Y. lipolytica*) cells seems to be constituted of at least two processes: one is rapid but saturable process and the other is slow process but independent of metabolic energy and membrane potential (Kohlwein and Paltauf 1984). Competition assay suggested the presence of two types of carriers: one prefers long-chain fatty acids, such as palmitic acid, stearic acid, and oleic acid; the other prefers lauric acid and myristic acid. As to the utilization of shorter-chain fatty acids, uptake of octanoic acid and decanoic acid were inefficient, and octanoic acid did not support growth of *Y. lipolytica*. The molecular process of fatty acid transport in this organism has not been much studied so far. A vectorial acylation process has been proposed in transport and activation of exogenous fatty acids in *Saccharomyces cerevisiae*, in which a transporter protein Fat1p and primary acyl-CoA synthetase (ACS) Faa1p function in complex (Zou et al. 2003), but ACS activity of Fat1p itself for long-chain fatty acids (*ibid.*) and residual transport activity of Δ*fat1* mutant (Færgeman et al. 1997) obscure understanding of the transport system. Orthologs of *FAT1* and *FAA1* are also present in *Y. lipolytica*, but Fat1p ortholog has peroxisome-targeting signal one (PTS1) at its C-terminal end, and their functions are yet the subject to be clarified.

Y. lipolytica is able to utilize long-chain alcohols and aldehydes such as dodecanol and dodecanal, as a sole carbon source, but not shorter alcohols and aldehydes such as decanol and decanal, probably because of their cytotoxicity (Takai et al. unpublished observation). The mechanisms for their transport and further oxidation to dodecanoic acid have not been much studied so far.

Fatty acids that are transported from outside or generated by *n*-alkane oxidation are converted to acyl-CoA by ACS and oxidized in peroxisome or utilized for the synthesis of membrane and storage lipids. Early biochemical and genetic studies by

Kamiryo et al. revealed that there are at least two ACSs, ACS I and ACS II, in *Candida lipolytica* (Kamiryo et al. 1977). ACS I was mainly in the cytosol fraction and ACS II was in the peroxisome fraction. Mutants defective in ACS I were unable to grow on a medium containing palmitic acid and oleic acid and 68 μM cerulenin. When grown on fatty acids with odd-numbered carbons as a sole carbon source, the wild type *C. lipolytica* contained considerable amounts of fatty acids of odd-numbered carbons, whereas the mutant strains far less contained them, suggesting that the ACS I-deficient mutants mainly degraded exogenous fatty acids by β-oxidation and did not utilize them for membrane and storage lipids. Endogenously synthesized fatty acids are released from fatty acid synthase as acyl-CoA, and they are not activated by ACS I nor ACS II (Black and DiRusso 2007).

By searching *Y. lipolytica* genome database, we found three putative genes coding for ACSs in *Y. lipolytica*, namely, *YALI0D17864g*, *YALI0E16016g*, and *YALI0E12859g*, in contrast to six genes in *S. cerevisiae* (Black and DiRusso 2007). *YALI0D17864g* is homologous to *FAA1*, *FAA2*, *FAA3*, and *FAA4* of *S. cerevisiae*. *YALI0E16016g* and *YALI0E12859g* are highly homologous to *FAT1* and *FAT2* of *S. cerevisiae*, respectively. The translation product of *YALI0E16016g* is predicted to have a PTS1 at its C-terminus, and the disruption mutant of *YALI0E16016g* did not grow on *n*-decane and *n*-dodecane and grew slowly on *n*-tetradecane and *n*-hexadecane (Park et al. unpublished observation), suggesting that the product of *YALI0E16016g* resides in peroxisome and essential for β-oxidation of fatty acids of short or medium carbon chain lengths. The gene product of *YALI0E16016g* could be ACS II. The disruption mutant of *YALI0D17864g* showed diminished growth on oleic acid and *n*-hexadecane, suggesting that the product of *YALI0D17864g* may function in activation of longer fatty acids. Disruption of *YALI0E12859g* had no effect on the growth on any carbon sources. We named *YALI0D17864g*, *YALI0E16016g*, and *YALI0E12859g* as *YlFAA1*, *YlFAT1*, and *YlFAT2*, respectively. *YALI0C09284g* and *YALI0B05456g* had weak homology to *S. cerevisiae FAT2*, but single disruption of these genes did not show any growth phenotypes (Park et al. unpublished observation).

2 Utilization of *n*-Alkanes

The ability to assimilate *n*-alkane is a property not belonging to only limited species of fungi, which is often overlooked by scientists who are studying laboratory yeast as a model eukaryotic system. This ability actually spreads over many genera of yeasts and filamentous fungi, but most of them are not easy to handle or not developed for molecular genetic studies. *Y. lipolytica* turned out to be an excellent organism for analyzing how *n*-alkanes are utilized by fungi at molecular level.

2.1 Incorporation of n-Alkanes

Utilization of *n*-alkanes starts with incorporation and subsequent transport to the ER, where the initial hydroxylation of *n*-alkanes is executed. Although the molecular mechanisms underlying these processes still remain unclear, two models are considered to account for the uptake of *n*-alkanes by alkane-assimilating yeasts: a passive, diffusion-like mechanism and an active, energy-dependent mechanism. While the former model has been appreciated based on the highly hydrophobic properties of *n*-alkanes, experimental results in favor of the latter model were reported. The uptake of ^{14}C-labeled *n*-hexadecane by *Y. lipolytica* was inhibited by addition of 2,4-dinitrophenol and KCN, and it was induced by *n*-decane and repressed by glycerol (Bassel and Mortimer 1985; Sumita et al. unpublished observation). Since uptakes were measured within the range of 20–100 min, the observed values should have been the sum of transported *n*-alkanes and their oxidized derivatives. Bassel and Mortimer isolated and characterized *Y. lipolytica* mutants defective in the utilization on *n*-alkane and showed that 16 mutations caused a significant reduction in *n*-alkane uptake (Bassel and Mortimer 1982, 1985). Nicaud and colleagues showed that *Y. lipolytica* strains with insertion mutations in *ABC1* encoding an ABC transporter are defective in the growth on *n*-hexadecane, but not on *n*-decane (Thevenieau et al. 2006). Although it is not clear whether Abc1p participates in the uptake of *n*-alkanes to assimilate or in the excretion of excess *n*-alkanes to maintain the cellular integrity, these results imply the involvement of the energy-dependent transporter in the transport of *n*-alkane of a certain carbon chain length across the plasma membrane.

2.2 Terminal Hydroxylation of n-Alkanes by Cytochrome P450

The terminal hydroxylation of *n*-alkanes to fatty alcohols is catalyzed by cytochrome P450ALK, which is classified into CYP52 family in the ER (Fig. 1). P450ALKs constitute multigene families in various alkane-assimilating yeasts, including *Candida tropicalis*, *Candida maltosa*, *Debaryomyces hansenii*, and *Y. lipolytica* (Ohkuma et al. 1995; Seghezzi et al. 1992; Yadav and Loper. 1999). Eight genes encoding P450ALKs, *ALK1* to *ALK8*, were isolated in *Y. lipolytica* (Iida et al. 1998, 2000). In addition, four additional *ALK* genes, *ALK9* to *ALK12*, were identified on the genome of *Y. lipolytica* (Fickers et al. 2005; Hirakawa et al. 2009). The deletion mutant of *ALK1* exhibited a severe growth defect on the medium containing *n*-decane, a shorter carbon chain *n*-alkane, as a sole carbon source, but not on *n*-hexadecane, a longer carbon chain *n*-alkane. Simultaneous deletion of *ALK1* and *ALK2* caused severe growth defect on *n*-hexadecane. In contrast, individual deletions of *ALK2*, *ALK3*, *ALK4*, and *ALK6* did not confer any growth defects on *n*-decane or *n*-hexadecane (Iida et al. 2000). These results

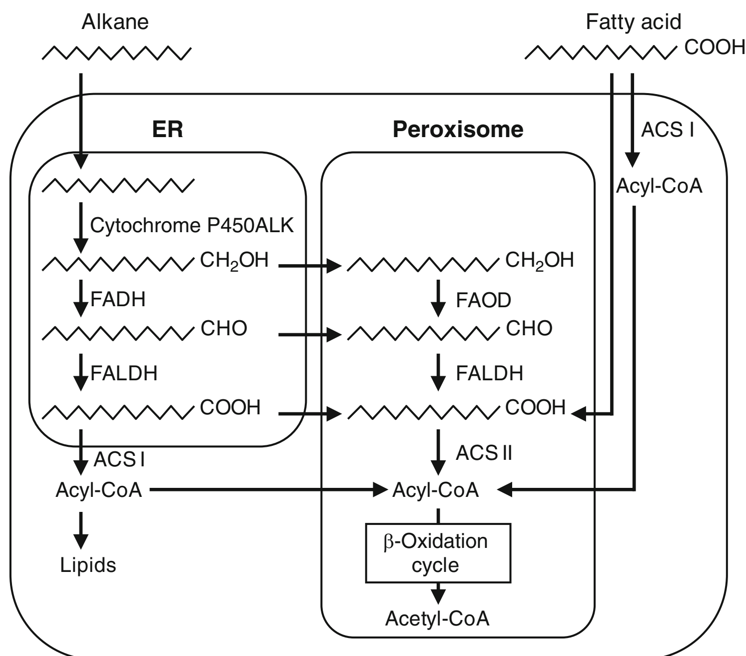


Fig. 1 Degradation pathway of *n*-alkane and fatty acids in *Y. lipolytica*. See text for detail

indicate the critical functions of Alk1p and Alk2p in the assimilation of *n*-alkanes. Alk3p, Alk5p, and Alk7p have ω-hydroxylation activities against lauric acid when expressed in leaves of *Nicotiana benthamiana* (Hanley et al. 2003). These results illustrate the distinct substrate specificities of P450ALK species in *Y. lipolytica*. Recently, we have successfully constructed a *Y. lipolytica* mutant deleted in all 12 genes encoding P450ALKs (Takai et al. unpublished observation). This mutant totally lost the ability to grow on *n*-decane and *n*-hexadecane. Individual expression of each P450ALK in this mutant will contribute to elucidation of its function in vivo and in vitro.

2.3 Further Hydroxylation of Long-Chain Alcohols to Fatty Acids

Fatty alcohols are oxidized to fatty aldehyde either by fatty alcohol dehydrogenase (FADH) in the ER or by fatty alcohol oxidase (FAOD) in peroxisome (Fig. 1). Although several genes encoding proteins homologous to alcohol dehydrogenases are found in the genome sequence of *Y. lipolytica*, their functions remain unclear

(Fickers et al. 2005). Fatty alcohols are then oxidized by as-yet-unidentified fatty aldehyde dehydrogenase (FALDH) in the ER or peroxisome to fatty acids, which are activated by ACS I in cytosol or ACS II in peroxisome, and are used for lipid synthesis or metabolized by peroxisomal β-oxidation (Fig. 1). Interestingly, P450ALK 52A3 of *C. maltosa* catalyzes complete oxidation of *n*-alkane to fatty acids in vitro (Scheller et al. 1998), raising the possibility that *n*-alkanes are oxidized to fatty acids solely by P450ALKs in the ER. In *Y. lipolytica*, however, the mutant deleted in all 12 *ALK* genes grew as well as wild-type strain on the medium containing dodecanol or dodecanal, the intermediate metabolites of *n*-dodecane, as a sole carbon source (Takai et al. unpublished observation). Thus, P450ALKs do not appear to be essential for the further oxidation of fatty alcohols or fatty aldehydes in the process of *n*-alkane assimilation.

The sequential engagement of ER and peroxisome in the metabolism of *n*-alkanes involves the transport of their metabolites from ER to peroxisome. Elucidation of the spatial distribution of *n*-alkane metabolites and their transport mechanism, in addition to identification of the enzymes involved in their metabolism, is critical for a better understanding of utilization of *n*-alkane by *Y. lipolytica*.

2.4 Transcriptional Control of Metabolism of n-Alkanes

The expression of *ALK1* and a subset of *ALK* genes is highly induced by *n*-alkanes and repressed strongly by glycerol and weakly by glucose (Hirakawa et al. 2009; Iida et al. 1998, 2000). Two basic helix-loop-helix (bHLH) type transcriptional activators, Yas1p and Yas2p, and an Opi1-family transcriptional repressor, Yas3p, are involved in the regulation of *ALK1* transcription in response to *n*-alkane (Endoh-Yamagami et al. 2007; Hirakawa et al. 2009; Yamagami et al. 2004). Yas1p and Yas2p form a complex and bind to alkane-responsive element 1 (ARE1), an E-box-like sequence important for the transcriptional activation in response to *n*-alkane, in *ALK1* promoter (Fig. 2) (Endoh-Yamagami et al. 2007; Sumita et al. 2002; Yamagami et al. 2004). Deletion of *YAS1* or *YAS2* abolishes the transcriptional activation of *ALK1* and the growth on the medium containing *n*-alkane as a carbon source. Yas3p binds to Yas2p in vitro, and deletion of *YAS3* causes remarkable overexpression of *ALK1* on *n*-alkane, glucose, or even glycerol (Hirakawa et al. 2009).

Interestingly, *YAS1*, *YAS2*, and *YAS3* are orthologs of *INO4*, *INO2*, and *OPI1*, respectively, which are involved in the transcriptional regulation of phospholipid synthetic genes in response to *myo*-inositol in *S. cerevisiae*. The heterodimer formed by Ino2p and Ino4p binds to a DNA element called UAS_{INO}/ICRE in the promoters of genes and activates their transcription in the absence of *myo*-inositol (Greenberg and Lopes 1996). According to the model proposed by Levine and colleagues, Opi1p is sequestered on the ER membrane through binding to the integral membrane protein, Scs2p or Scs22p, and to phosphatidic acid (PA). When *myo*-inositol is supplemented, PA is consumed for synthesis of phosphatidylinositol. Decrease of

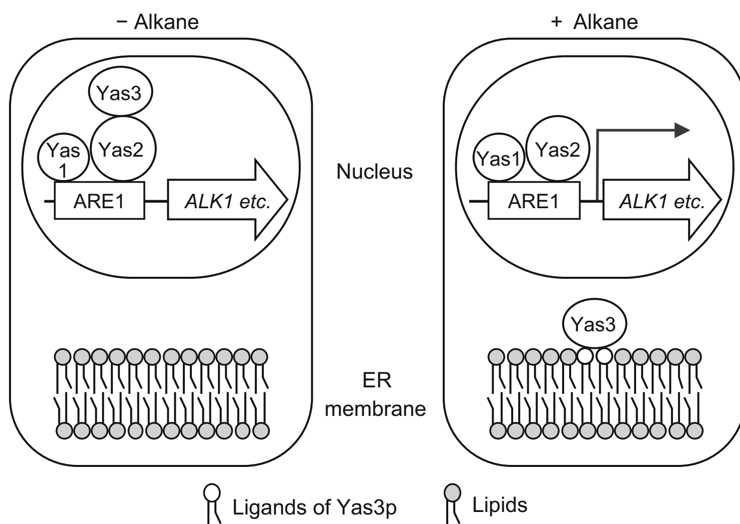


Fig. 2 Transcriptional regulation by Yas1p, Yas2p, and Yas3p. The transcriptional activator complex composed of Yas1p and Yas2p binds ARE1 in the promoter of target genes including *ALK1*. In the absence of *n*-alkane, the transcriptional repressor, Yas3p, translocates into the nucleus, binds to Yas2p, and represses transcription of target genes. When cells are grown on *n*-alkane as a carbon source, Yas3p is sequestered on the ER membrane through binding to its ligands, probably lipophilic compounds, in the membrane, and liberated Yas1p–Yas2p complex activates transcription

PA in the ER membrane causes release of Opi1p from the ER, and Opi1p translocates into the nucleus and represses the transcription of target genes through binding to Ino2p (Loewen et al. 2004). Correspondingly, Yas3p localizes in the nucleus on glucose or glycerol and to the ER on *n*-alkane, while Yas1p and Yas2p constitutively localize in the nucleus (Yamagami et al. 2004; Hirakawa et al. 2009). In addition, recombinant Yas3p binds to PA in vitro (Kobayashi et al. unpublished observation). Thus, Yas3p probably functions as a master regulator of the transcriptional response to *n*-alkane, and its intracellular localization possibly regulated by phospholipids should be vital for its mode of action (Fig. 2). However, since deletion of *SCS2* and/or *SCS22* ortholog(s) does not affect the transcription of *ALK1* or the localization of Yas3p (Kobayashi et al. 2008; Kobayashi et al. unpublished observation), regulatory mechanisms distinct from those in *S. cerevisiae* can be involved in the regulation of the localization of Yas3p.

Opi1-family proteins, most of whose functions remain uncharacterized, are widely and specifically encoded on the genome of a variety of fungi. Yas3p does not appear to be involved in the transcriptional regulation of phospholipid synthetic genes, because deletion of *YAS3* did not affect the expression of *INO1* ortholog in *Y. lipolytica* (Hirakawa et al. 2009). Opi1-family proteins may be involved in the regulation of diverse processes in synthesis or metabolism of hydrophobic compounds in fungi.

References

Barth G, Gaillardin C (1997) Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica*. FEMS Microbiol Rev 19:219–237

Bassel JB, Mortimer RK (1982) Genetic and biochemical studies of *n*-alkane non-utilizing mutants of *Saccharomycopsis lipolytica*. Curr Genet 5:77–88

Bassel JB, Mortimer RK (1985) Identification of mutations preventing *n*-hexadecane uptake among 26 *n*-alkane non-utilizing mutants of *Yarrowia* (*Saccharomycopsis*) *lipolytica*. Curr Genet 9:579–586

Black PN, DiRusso CC (2007) Yeast acyl-CoA synthetases at the crossroads of fatty acid metabolism and regulation. Biochim Biophys Acta 1771:286–298

Endoh-Yamagami S, Hirakawa K, Morioka D, Fukuda R, Ohta A (2007) Basic helix-loop-helix transcription factor heterocomplex of Yas1p and Yas2p regulates cytochrome P450 expression in response to alkanes in the yeast *Yarrowia lipolytica*. Eukaryot Cell 6:734–743

Færgeman NJ, DiRusso CC, Elberger A, Knudsen J, Black PN (1997) Disruption of the *Saccharomyces cerevisiae* homologue to the murine fatty acid transport protein impairs uptake and growth on long-chain fatty acids. J Biol Chem 272:8531–8538

Fickers P, Benetti PH, Wache Y, Marty A, Mauersberger S, Smit MS, Nicaud JM (2005) Hydrophobic substrate utilisation by the yeast *Yarrowia lipolytica*, and its potential applications. FEMS Yeast Res 5:527–543

Greenberg ML, Lopes JM (1996) Genetic regulation of phospholipid biosynthesis in *Saccharomyces cerevisiae*. Microbiol Rev 60:1–20

Hanley K, Nguyen LV, Khan F, Pogue GP, Vojdani F, Panda S, Pinot F, Oriedo VB, Rasochova L, Subramanian M, Miller B, White EL (2003) Development of a plant viral-vector-based gene expression assay for the screening of yeast cytochrome p450 monooxygenases. Assay Drug Dev Technol 1:147–160

Hirakawa K, Kobayashi S, Inoue T, Endoh-Yamagami S, Fukuda R, Ohta A (2009) Yas3p, an Opi1 family transcription factor, regulates cytochrome P450 expression in response to *n*-alkanes in *Yarrowia lipolytica*. J Biol Chem 284:7126–7137

Iida T, Ohta A, Takagi M (1998) Cloning and characterization of an *n*-alkane-inducible cytochrome P450 gene essential for *n*-decane assimilation by *Yarrowia lipolytica*. Yeast 14:1387–1397

Iida T, Sumita T, Ohta A, Takagi M (2000) The cytochrome P450ALK multigene family of an *n*-alkane-assimilating yeast, *Yarrowia lipolytica*: cloning and characterization of genes coding for new CYP52 family members. Yeast 16:1077–1087

Kamiryo T, Mishina M, Tashiro SI, Numa S (1977) *Candida lipolytica* mutants defective in an acyl-coenzyme A synthetase: isolation and fatty acid metabolism. Proc Natl Acad Sci USA 74:4947–4950

Kobayashi S, Hirakawa K, Fukuda R, Ohta A (2008) Disruption of the *SCS2* ortholog in the alkane-assimilating yeast *Yarrowia lipolytica* impairs its growth on *n*-decane, but does not impair inositol prototrophy. Biosci Biotechnol Biochem 72:2219–2223

Kohlwein SD, Paltauf F (1984) Uptake of fatty acids by the yeasts, *Saccharomyces uvarum* and *Saccharomycopsis lipolytica*. Biochim Biophys Acta 792:310–317

Loewen CJ, Gaspar ML, Jesch SA, Delon C, Ktistakis NT, Henry SA, Levine TP (2004) Phospholipid metabolism regulated by a transcription factor sensing phosphatidic acid. Science 304:1644–1647

Ohkuma M, Muraoka S, Tanimoto T, Fujii M, Ohta A, Takagi M (1995) CYP52 (cytochrome P450alk) multigene family in *Candida maltosa*: identification and characterization of eight members. DNA Cell Biol 14:163–173

Scheller U, Zimmer T, Becher D, Schauer F, Schunck WH (1998) Oxygenation cascade in conversion of n-alkanes to alpha, omega-dioic acids catalyzed by cytochrome P450 52A3. J Biol Chem 273:32528–32534

Seghezzi W, Meili C, Ruffiner R, Kuenzi R, Sanglard D, Fiechter A (1992) Identification and characterization of additional members of the cytochrome P450 multigene family CYP52 of *Candida tropicalis*. DNA Cell Biol 11:767–780

Sumita T, Iida T, Yamagami S, Horiuchi H, Takagi M, Ohta A (2002) *YlALK1* encoding the cytochrome P450ALK1 in *Yarrowia lipolytica* is transcriptionally induced by *n*-alkane through two distinct cis-elements on its promoter. Biochem Biophys Res Commun 294:1071–1078

Thevenieau F, Le Dall MT, Nthangeni B, Mauersberger S, Marchal R, Nicaud JM (2006) Characterization of *Yarrowia lipolytica* mutants affected in hydrophobic substrate utilization. Fungal Genet Biol 44:531–542

Yadav JS, Loper JC (1999) Multiple p450alk (cytochrome P450 alkane hydroxylase) genes from the halotolerant yeast *Debaryomyces hansenii*. Gene 226:139–146

Yamagami S, Morioka D, Fukuda R, Ohta A (2004) A basic helix-loop-helix transcription factor essential for cytochrome p450 induction in response to alkanes in yeast *Yarrowia lipolytica*. J Biol Chem 279:22183–22189

Zou Z, Tong F, Færgeman NJ, Børsting C, Black PN, DiRusso CC (2003) Vectorial acylation in *Saccharomyces cerevisiae*. Fat1p and fatty acyl-CoA synthetase are interacting components of a fatty acid import complex. J Biol Chem 278:16414–16422

Ambient pH Signalling in *Yarrowia lipolytica*

Sylvie Blanchin-Roland

Abstract Many yeast and fungi grow over a wide pH range, and their gene expression is tailored to the pH of their environment. A conserved ambient pH signal transduction pathway has been evidenced in both ascomycetous yeasts and filamentous fungi, called Rim or Pal, respectively. The transcriptional factor Rim101p/PacC is activated by a proteolytic C-terminal processing in response to alkaline ambient pH. In *A. nidulans* and probably in *Y. lipolytica*, this processing occurs in two steps, the first one being pH signal dependent and the second one proteasomal. The Rim101p/PacC truncated form is able to activate alkaline pH-responsive genes and to repress acid-induced genes. The Rim/Pal pathway involves both a plasma membrane complex including the 7-TMD protein and putative pH sensor Rim21p/PalH, the 3-TMD protein and putative assistant of Rim21p/PalH localization Rim9p/PalI, and the arrestin-like protein Rim8p/palF, and an ESCRT-associated Rim101 processing machinery which comprises, besides Rim101p/PacC, three interactors of the ESCRT-III-subunit Snf7p/Vps32p: the calpain-like signalling protease Rim13p/PalB, a scaffold Rim20p/PalA, and YlRim23p/PalC. According to the current model, in response to alkaline ambient pH, the interaction of Rim21/PalH with Rim8p/PalF bound to the ESCRT-I subunit Vps23p would promote the ESCRT machinery recruitment to the plasma membrane, like in retroviral budding. And the interaction of the ESCRT-III-subunit Snf7 with both Rim13p/PalB and Rim20p/PalA bound to Rim101p/PacC would trigger the pH-signal-dependent proteolytic processing of Rim101/PacC. Evolutionary conserved mechanisms would control the recruitment of the ESCRT machinery to Rim/Pal proteins in fungi and to retroviral Gag proteins in animal cells.

S. Blanchin-Roland (✉)
MICALIS, UMR1319, CBAI, INRA, 78850 Thiverval-Grignon, France
e-mail: syrondel@orange.fr

G. Barth (ed.), *Yarrowia lipolytica*, Microbiology Monographs 24,
DOI 10.1007/978-3-642-38320-5_6, © Springer-Verlag Berlin Heidelberg 2013

Contents

1 Introduction 122
2 Identification of the Factors Required for Ambient pH Signalling 123
2.1 The First PAL/RIM Mutant Isolation: Identification of a RIM101/pacC Homolog .. 123
2.2 The Rim Factors 123
2.3 The ESCRT Machinery 124
3 Characterization of the YlRim and YlVps Factors 126
3.1 The Transcriptional Factor YlRim101p/PacC 126
3.2 The Plasma Membrane Signalling Machinery 127
3.3 The ESCRT-Associated Rim101 Processing Machinery 129
3.4 Effects of YlRim and YlVps Mutations on Ambient pH Response 132
4 Other Functions of the Rim Pathway in *Y. lipolytica* and Other Ascomycetes 134
5 The Current Model of the pH Signalling Pathway in Yeasts and Filamentous Fungi 136
6 Conclusion 138
References 139

1 Introduction

Extracellular pH regulation plays major roles in the control of many biological processes including the survival of fungi and bacteria and the virulence of some of these microorganisms. Many organisms, especially those that are able to grow over a wide pH range, tailor gene expression to the pH of their growth environment. Ambient pH-responsive genes include those encoding permeases, plasma membrane transporters, secreted proteins, and enzymes involved in the synthesis of exported metabolites. In the yeast *Yarrowia lipolytica*, transcription of the *XPR2* and *AXP1* genes, encoding alkaline and acidic extracellular proteases, respectively, is tightly controlled by a combination of environmental stimuli including ambient pH and nutrient availability (Ogrydziak et al. 1977; Ogrydziak 1993). Both protease genes are similarly induced at the end of the exponential phase on complex media containing proteins. At acidic pH, induction of *AXP1* leads to secretion of the acid protease Axp, while at neutral/alkaline pH, the alkaline protease Aep is secreted as *XPR2* becomes induced (Glover et al. 1997).

These genes are regulated by a conserved pH signalling pathway called Pal in filamentous fungi and Rim in yeasts. This pathway, which has been intensively investigated in *Aspergillus nidulans* (Penalva and Arst 2002, 2004; Penalva et al. 2008; Calcagno-Pizarelli et al. 2011), has been studied in several other ascomycetes including *Y. lipolytica* (Lambert et al. 1997; Tréton et al. 2000; Gonzalez-Lopez et al. 2002; Blanchin-Roland et al. 2005, 2008; Blanchin-Roland 2011), *Saccharomyces cerevisiae* (Li and Mitchell 1997; Xu and Mitchell 2001; Xu et al. 2004; Boysen and Mitchell 2006; Herrador et al. 2010), and *Candida albicans* (Ramon and Fonzi 2003; Li et al. 2004; Kullas et al. 2004; Cornet et al. 2005; Barwell et al. 2005; Baek et al. 2006), in the basidiomycete *Ustilago maydis* (Arechiga-Carvajal and Ruiz-Herrera 2005), and in other fungi (Penalva and Arst 2002, 2004; Penalva et al. 2008). At neutral/alkaline pH, the proteins encoded by six *Pal*/*RIM* genes transmit a yet unidentified pH signal to the zinc finger transcriptional factor

Rim101p/PacC that is activated through a complex proteolytic processing of its C-terminal inhibitory domain. The Rim101p/PacC truncated form is able to activate alkaline-induced genes and to repress acid-induced genes (for reviews see Penalva and Arst 2002, 2004 and Penalva et al. 2008).

2 Identification of the Factors Required for Ambient pH Signalling

2.1 The First PAL/RIM Mutant Isolation: Identification of a RIM101/pacC Homolog

We isolated recessive mutations preventing *XPR2* derepression under conditions of carbon and nitrogen limitation at neutral pH (Lambert et al. 1997). They identified four unlinked genetic loci named *PAL1* to *PAL4*. Only two genes, *PAL2/YlRIM21* and *PAL3/YlRIM8* (see below and Tréton et al. 2000), could be identified by complementing these Pal mutants, as well as a dominant suppressor of all four Pal mutants (Lambert et al. 1997). This suppressor encoded a C-terminally truncated version of a transcriptional factor of the Cys_2His_2 zinc finger family, which we called YlRim101p. Interestingly, a point mutation (previously called *RPH2*-5) truncating YlRim101p was identified in an independent screen when looking for mutations expressing *XPR2* independently of the pH (Otero and Gaillardin 1996). YlRim101p is homologous to the transcriptional activators Rim101p of *S. cerevisiae*—initially isolated as required for entry into meiosis and called Rim1p (Su and Mitchell 1993), later identified as mediating ambient pH regulation (Li and Mitchell 1997)—and to PacC factors which mediate response to ambient pH in filamentous fungi (Tilburn et al. 1995; MacCabe et al. 1996).

It might seem surprising that screening for mutations affecting *XPR2* expression under conditions of carbon and nitrogen limitations would lead to the identification of mutations affecting the pH response. Actually, according to Ogrydziak et al. (1977), either carbon or nitrogen starvation alone is sufficient to derepress *XPR2*. So single mutations affecting any one of these two regulatory pathways might not block induction by the other pathway. Therefore, under the conditions of this screening, the only way to block *XPR2* expression was to affect its activation by the Rim pathway.

2.2 The Rim Factors

To identify other partners of this ambient pH response pathway, as well as other potential regulators of proteases, we screened an insertional library (Neuveglise et al. 1998) for altered expression of either acidic or alkaline protease genes, or both (Gonzalez-Lopez et al. 2002). Approximately 190,000 transformants were

screened, corresponding to about ten insertions per kb of genomic DNA, assuming random mutagenesis and a genome size of 20 Mb for *Y. lipolytica* (Dujon et al. 2004). Of 128 mutants obtained, most of which (94) affected expression of both Aep at pH 7.0 and Axp at pH 4.0. Identification of the disrupted loci by reverse PCR was rather efficient since 86 % of the mutants tested could be efficiently processed. In addition to the previously characterized genes *YlRIM101/pacC* and *YlRIM8/palF*, a subset of these loci identified three new genes, *YlRIM13/palB* (which complements *pal1* mutants), *YlRIM20/palA* (which complements *pal4* mutants), and *YlRIM9/palI*.

Altogether these approaches led to the identification of six genes in the Rim pathway of *Y. lipolytica* versus seven in the Pal pathway of *A. nidulans* (Negrete-Urtasun et al. 1999). The missing gene encoding a homolog of *PalC* (Tilburn et al. 2005) was identified from the *Y. lipolytica* genome (Dujon et al. 2004). The predicted 441-amino acid protein was named YlRim23p (Blanchin-Roland et al. 2008). A likely PalC ortholog, encoded by YGR122w, has been identified in *S. cerevisiae* (Galindo et al. 2007).

2.3 The ESCRT Machinery

Among the insertional mutants affecting non-Rim components of protease regulation, one and two mutants affected genes encoding homologs of *S. cerevisiae* Vps23p and Vps28p (vacuolar protein sorting), respectively (Gonzalez-Lopez et al. 2002; Blanchin-Roland et al. 2005). ScVps23p and ScVps28p are two class E components of the ESCRT-I complex (endosomal sorting complex required for transport) (Katzmann et al. 2001). This complex with the three other ESCRT complexes -0, -II, and -III acts sequentially in the sorting of endocytic ubiquitinated membrane proteins to the multivesicular body (MVB) vesicles that are ultimately delivered to the lumen of the vacuole for degradation (Hurley and Emr 2006). The MVB sorting pathway is conserved from yeast to higher eukaryotes and is required, besides receptor and transporter downregulation, for a growing list of cellular functions (Slagsvold et al. 2006) that include signalling (Sadowski et al. 2009) and even egress of enveloped ARN viruses (Calistri et al. 2009).

Like *YlRIM* genes, *YlVPS23 and YlVPS28* were shown to be absolutely required for cellular responses to neutral/alkaline pH mediated by Rim101 (Fig. 1b, c and Blanchin-Roland et al. 2005). Other data have completed and confirmed that some Vps factors are required for ambient pH response in yeasts and fungi. Vps23p was shown to bind Rim8p/PalF in *S. cerevisiae* (Herrador et al. 2010) and *Y. lipolytica* (Blanchin-Roland 2011), and this interaction is essential for Rim signalling. An interaction between Snf7p and Rim20p/PalA, initially detected in *S. cerevisiae* by a whole-genome screen of two-hybrid interactants (Ito et al. 2001), was confirmed in *S. cerevisiae*, *Y. lipolytica*, *A. nidulans*, and human (Vincent et al. 2003; Xu et al. 2004; Blanchin-Roland et al. 2008). Snf7p is a component of ESCRT-III (Babst et al. 2002) acting at a late stage in the sorting of endosomal cargo into the MVB

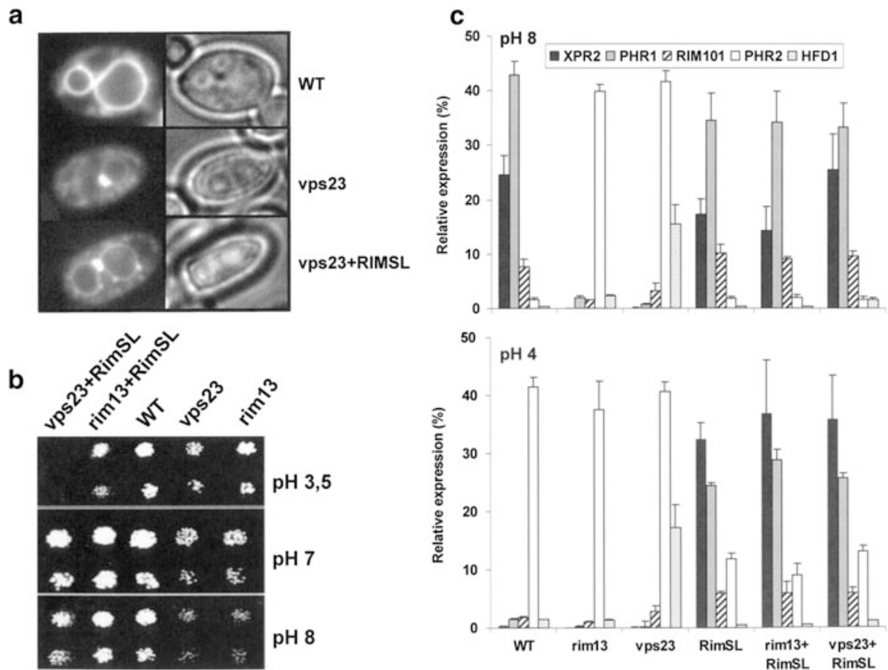


Fig. 1 Tools to study ambient pH response in *Y. lipolytica*. (**a**) The endocytic defect of *Ylvps23* null mutant is not suppressed by the constitutively active truncated allele of *RIM101*. This defect was visualized with the fluorescent dye FM4-64. The *left side* of each panel shows FM4-64 fluorescence, while the *right side* shows differential interference contrast (DIC) images. FM4-64 can be seen on the vacuole-limiting membrane in the wild-type (WT) cells. In the *vps23* null mutant cells as well as in this mutant carrying the dominant, constitutively active truncated allele of *RIM101* (RIMSL, SL for short length), the bulk of FM4-64 remains in small compartments adjacent to the vacuole. (**b**) Effects of the different mutations on *Y. lipolytica* sensitivity to ambient pH. Droplets of serial dilutions of late-exponential culture cells were deposited on solid YPD medium buffered at the indicated pH. The strains were the *rim13* and *vps23* null mutants, the wild type (WT), and each of these null mutants carrying RimSL. (**c**) Expression patterns of alkaline-induced genes (*XPR2*, *PHR1* and *RIM101*) and of acid-induced genes (*PHR2* and *HFD1*) in *Y. lipolytica* at pH 8.0 and 4. 0. The strains were the wild type (WT), the *rim13* and *vps23* null mutants, the strain expressing RimSL, and each of the null mutant carrying the RimSL allele. The levels of mRNA were measured by quantitative RT-PCR and expressed as a percentage of the constitutive expression level of the actin gene. Results are the mean +/− SD of four quantifications

sorting pathway. Another interaction between Snf7p and Rim13p/PalB, initially detected in *S. cerevisiae* (Ito et al. 2001), was confirmed in *Y. lipolytica* (Blanchin-Roland et al. 2008) and in human (Yorikawa et al. 2008). In *Y. lipolytica* and *A. nidulans*, YlRim23p/PalC was also shown to interact with Snf7 (Galindo et al. 2007; Blanchin-Roland et al. 2008).

Finally, a systematic phenotypic analysis of a nearly complete collection of gene-deletion mutants of *S. cerevisiae* indicated that several Vps mutants exhibited an impaired growth at alkaline pH or in the presence of high concentrations of

monovalent cations (Giaever et al. 2002), two phenotypes which are characteristic of Rim mutants in this organism (Lamb et al. 2001). In a separate study, the ESCRT-I and -II complexes, together with the Snf7p and Vps20p subunits of ESCRT-III, were shown to be required for Rim101p processing in *S. cerevisiae* (Xu et al. 2004). Interestingly, the whole endocytic pathway is not involved in pH signalling since not all ESCRT encoding genes and none of the non-ESCRT *VPS* genes are required for ambient pH signalling. This functional link between the Rim and Vps pathways is conserved in *C. albicans* (Xu et al. 2004; Kullas et al. 2004; Cornet et al. 2005) and in *A. nidulans* (Calcagno-Pizarelli et al. 2011).

A puzzling observation is that only two ESCRT-I subunit encoding genes were identified in our mutagenesis in *Y. lipolytica*. This may seem surprising in view of the different *VPS* genes reported in *S. cerevisiae*, from the screening of the haploid deletion strain library, as required for growth at mild alkaline pH (Serrano et al. 2004), or functioning in the Rim pathway (Barwell et al. 2005). Although our screen was not exhaustive, it identified five out of six *RIM* genes, but only two out of eight *VPS* genes so far identified in *S. cerevisiae* as required for Rim101p processing (Xu et al. 2004). Moreover, no *vps* mutant at all was identified in *A. nidulans*, in exhaustive screens for defects in pH signalling (Penalva and Arst 2004). The fact that *YlSNF7* is essential and that *Ylvps23* or *Ylvps28* mutations mildly affect growth at all pH tested (Fig. 1b and Blanchin-Roland et al. 2005) suggested that most if not all *VPS* genes are essential or lead to major growth defects in *A. nidulans* and to a lesser extent in *Y. lipolytica*. Indeed, in *A. nidulans* components of ESCRT-0, ESCRT-I, ESCRT-II, and ESCRT-III are nearly essential for growth (Calcagno-Pizarelli et al. 2011). pH signalling is not an essential function in fungi under most circumstances, but endocytosis is required for cellular growth.

3 Characterization of the YlRim and YlVps Factors

3.1 The Transcriptional Factor YlRim101p/PacC

YlRIM101 encodes a 585-residue polypeptide with 33 % identity to *Sclerotinia sclerotiorum* PacC (Rollins and Dickman 2001), 32 % to *A. nidulans* PacC (Tilburn et al. 1995), 30 % to *C. albicans* Rim101p (Davis et al. 2000; El Barkani et al. 2000), and 28 % identity to *S. cerevisiae* Rim101p (Su and Mitchell 1993). Like its orthologs, YlRim101p contains three canonical Cys_2His_2 zinc fingers, nearly N-terminally located, with very high amino acid sequence identity to other members of the Rim101p/PacC family. The degree of sequence identity between all members of this family is much lower outside the Zf region.

As the Zf regions of the Rim101p/PacC factors show great sequence conservation, they probably recognize very similar DNA sequences. The core consensus of

PacC binding site in *A. nidulans* (Espeso et al. 1997) and *P. chrysogenum* (Suarez and Penalva 1996) is 5′-GCCARG. In *Y. lipolytica*, in vivo genomic footprints clearly demonstrated that within the proximal UAS (UAS2) of the *XPR2* promoter, the directly repeated decameric sequence CGCCAAGACG, which includes a double direct repeat of the core PacC binding site, interacts permanently with protein(s) in vivo (Blanchin-Roland et al. 1994). Moreover, this decameric repeat is implicated in the activation of UAS2-driven expression at neutral/alkaline pH (Madzak et al. 1999). This regulation is abolished in a null *Ylrim101* mutant and is active in a pH-independent manner in a strain expressing a dominant, constitutively active C-terminally truncated form of YlRim101p (Lambert et al. 1997; Madzak et al. 1999). YlRim101p binding to the decameric repeat may therefore activate *XPR2* directly in response to alkaline ambient pH. In the presence of preferred carbon and nitrogen sources, the UAS2-driven expression cannot be turned on by the C-terminally truncated form of YlRim101p. This suggests that other protein(s) may compete with Rim101p binding under such conditions to shut off *XPR2* transcription whatever the ambient pH (Madzak et al. 1999). This agrees with the footprinting data mentioned above which indicate constant occupation of the PacC-like sites.

In *A. nidulans*, activation of PacC requires two sequential proteolytic steps (Diez et al. 2002 and references therein). The "closed" 674 residue translation product is converted to an accessible intermediate by a pH-regulated proteolytic cleavage between residues 493 and 500 by the cysteine protease PalB (Penas et al. 2007). This processing step is required for the second processing cleavage by the proteasome within residues 252–254 (Hervas-Aguilar et al. 2007). Sequence comparison of YlRim101p and PacC suggests that the putative second processing cleavage of YlRim101p may occur around amino acid 330, about 100 residues downstream from the Zf region as in PacC. A mutant with a C-terminal truncation at this position was completely pH unresponsive and mimicked alkalinity, whereas less extensive truncations led to increased pH susceptibility (Lambert et al. 1997). Thus, YlRim101p processing, like PacC processing, may require two steps, whereas in *S. cerevisiae*, Rim101p proteolytic activation requires one Rim pathway-dependent proteolytic step (Li and Mitchell 1997; Xu and Mitchell 2001).

3.2 The Plasma Membrane Signalling Machinery

Response to the alkaline ambient pH requires two plasma membrane proteins, the seven-transmembrane (7TM) protein Rim21p/PalH and the 3TM protein PalI, and the arrestin-related protein Rim8p/PalF.

3.2.1 The Likely pH Sensor YlRim21p/PalH

YlRIM21, formerly *PAL2*, encodes a 632-residue polypeptide (Tréton et al. 2000) with 30 % identity to *A. nidulans* PalH (Negrete-Urtasun et al. 1999), 24 % to *C. albicans* Rim21p (Davis 2003), and 23 % to *S. cerevisiae* Rim21p (formerly YNL294c) (Maurer et al. 1995). These proteins carry seven putative TM domains in their N-terminal moiety followed by a hydrophilic C-terminal tail which is predicted to be cytosolic.

PalH has two homologs in each *S. cerevisiae* and *C. albicans*, Rim21p and Dfg16p (Barwell et al. 2005), but only one in *Y. lipolytica*, Rim21p. We found two putative paralogs of YlRim21p (YALI0B22814g and YALI0B06710g) in the *Y. lipolytica* genome. Deletion of each of these ORF had no effect on ambient pH response. This suggests that *Y. lipolytica* like *A. nidulans* carries only one functional PalH/YlRim21p and no equivalent of ScDfg16p (Blanchin-Roland et al. 2008).

PalH/Rim21p is the likely ambient alkaline pH sensor (see below and Herranz et al. 2005; Calcagno-Pizarelli et al. 2007).

3.2.2 YlRim9p/PalI, the Putative Assistant of YlRim21p/PalH

YlRIM9 encodes a 742 amino acid polypeptide (Gonzalez-Lopez et al. 2006) with 37 % identity to *A. nidulans* PalI (Denison et al. 1998). Alignment of Rim9p/PalI sequences showed that YlRim9p and its orthologs in filamentous fungi are much longer than their orthologs in other yeasts. The highly hydrophilic C-terminal tails that lie outside of the region of homology between all the Rim9p/PalI sequences represent about 66 % of PalI, 72 % of *Neurospora crassa* PalI, 75 % of YlRim9p, and 82 % of *Magnaporthe grisea* PalI (Blanchin-Roland et al. 2008). Rim9/PalI carries a putative peptide signal and three predictive hydrophobic TM segments in their common N-terminal moiety. Strong sequence conservation is restricted to this region. Indeed, in alignment with the 239 residue Rim9p of *S. cerevisiae* and with the 346 residue Rim9p of *C. albicans*, YlRim9p shows 43 % identity over 243 residues and 36 % identity over 329 residues, respectively, including all the hydrophobic segments.

In *Y. lipolytica* and possibly in filamentous fungi, the long C-terminal tail of YlRim9p/PalI does not seem to be involved in pH response (Blanchin-Roland et al. 2008).

As mentioned below, both AnPalI and YlRim9p contribute to, but are not completely essential for, pH signalling (Denison et al. 1998; Herranz et al. 2005; Calcagno-Pizarelli et al. 2007; Blanchin-Roland et al. 2008). In *A. nidulans*, PalI was shown to localize to the plasma membrane. PalH requires the coexpression of PalI at similar levels to predominate in the plasma membrane. This suggests that one (but possibly not the only) role of PalI is to promote the plasma membrane localization of PalH (Calcagno-Pizarelli et al. 2007).

3.2.3 The Arrestin-Related Protein YlRim8p/PalF

YlRIM8, formerly *PAL3*, encodes an 881-residue polypeptide (Tréton et al. 2000) with 32 % identity to *C. albicans* Rim8p (Porta et al. 1999) and 31 % to *A. nidulans* PalF (Maccheroni et al. 1997). We observed that both YlRim8p and PalF share sequence similarity with two adjacent ORFs of *S. cerevisiae*, separated by 147 bp on chromosome VII (Feuermann et al. 1997) and transcribed in the same orientation. We checked the sequence of this genomic region. The revised sequence showed a unique ORF encoding a 542-residue polypeptide with 29 % identity to *A. nidulans* PalF (Maccheroni et al. 1997) and 25 % to YlRim8p (Tréton et al. 2000).

Like its orthologs in yeasts and like PalF, YlRim8p contains arrestin N-terminal and C-terminal domains, located within a large N-terminal moiety. Mammalian arrestins bind to 7-TMD receptors to regulate their function. The arrestin-related protein PalF binds to the 7-TMD protein PalH. This strong interaction is absolutely required for pH signalling in vivo (Herranz et al. 2005). This data strongly implicate Rim21p/PalH as an ambient pH sensor. In *S. cerevisiae*, the binding of Rim8p/PalF to the C-terminal cytoplasmic tail of Rim21p/PalH was confirmed, and moreover, Rim8p/PalF was shown to bind to the ESCRT-I subunit Vps23p whose ubiquitin E2 variant (UEV) domain seemed sufficient for this interaction that is essential for pH signalling (Herrador et al. 2010). In *Y. lipolytica*, the conserved Vps23p UEV domain is absolutely required for ambient pH response and for the direct binding of Vps23p to Rim8p (Blanchin-Roland 2011). In *S. cerevisiae*, a SXP motif within the Rim8p C-terminus, called box 1, is a Vps23p-binding site. A PSXP motif in a related box2 sequence was proposed for CaRim8p, whereas both boxes 1 and 2 were suggested to be present in AnPalF and YlRim8p (Herrador et al. 2010). A closer inspection of an alignment including other yeasts suggests that only one box may be present, the *S. cerevisiae* SXP motif and its environment being highly conserved in these organisms (Blanchin-Roland 2011). The entirety of these data strongly suggests that in response to alkaline ambient pH, Rim8p/PalF binding to the probable pH sensor Rim21/PalF mediates the specific recruitment of the Rim8-bound ESCRT machinery to the plasma membrane (Herranz et al. 2005; Herrador et al. 2010).

3.3 The ESCRT-Associated Rim101 Processing Machinery

Compelling evidences show that ambient pH response in yeasts and fungi also requires Rim20/PalA, Rim13/PalB, and Rim23/PalC together with ESCRT components.

3.3.1 The Scaffold YlRim20p/PalA

YlRIM20 encodes a 773-residue polypeptide (Gonzalez-Lopez et al. 2002) which shares 36, 32, 31, and 27 % identity with *A. nidulans* PalA (Negrete-Urtasun et al. 1997), *S. cerevisiae* Rim20p (Xu and Mitchell 2001), *C. albicans* Rim20p (Davis et al. 2000), and *Homo sapiens* Alix/AIP1 (Vito et al. 1999), respectively.

Rim20p/PalA belongs to the family of Bro1-domain-containing proteins which includes *S. cerevisiae* Bro1p/Vps31p, its homolog in *A. nidulans* BroA, and the mammalian protein Alix. One conserved feature of Bro1-domain-containing proteins is their ability to bind to Snf7p and its mammalian homolog CHMP4b (Kim et al. 2005). The interaction between Rim20p/PalA/Alix and Snf7p/CHMP4b is conserved across evolution from *S. cerevisiae* and *Y. lipolytica* Rim20p to *A. nidulans* PalA and to human Alix and requires alkaline growth conditions (Ito et al. 2001; Vincent et al. 2003; Katoh et al. 2003; Xu et al. 2004; Boysen and Mitchell 2006; Blanchin-Roland et al. 2008).

Moreover, the binding of Rim20p/PalA to Rim101p/PacC, firstly demonstrated in *S. cerevisiae* and *C. albicans* (Xu and Mitchell 2001), was later confirmed in *A. nidulans* where PalA was shown to bind to two YPXL/I motifs flanking the signalling protease cleavage site in PacC (Vincent et al. 2003). These motifs are completely conserved in all members of the PacC/Rim101p family and are required for the first signalling cleavage of Rim101p/PacC. Altogether these data strongly suggest that in response to alkaline ambient pH, Rim20p mediates the recruitment of Rim101p to the ESCRT-III subunit Snf7p, Rim101 being cleaved by the signalling protease Rim13p bound to Snf7p (Xu and Mitchell 2001; Vincent et al. 2003; Xu et al. 2004).

3.3.2 The Cysteine Protease YlRim13p/PalB

YlRIM13 encodes a 795-residue polypeptide (Gonzalez-Lopez et al. 2002) with 30 % identity to *A. nidulans* PalB (Denison et al. 1995), 28 % to *C. albicans* Rim13p (Li et al. 2004), 26 % to *Homo sapiens* PalBH (Futai et al. 2001), and 24 % to *S. cerevisiae* Rim13p (Futai et al. 1999). In *A. nidulans*, PalB was shown to be the "signalling" protease that catalyzes the first, pH-regulated cleavage of PacC (Penas et al. 2007), whereas the second processing step involves the proteasome (Hervas-Aguilar et al. 2007). Rim13p/PalB and its orthologs contain a calpain-like thiol protease catalytic domain. The Cys and His residues of this domain are strictly conserved in 16 Rim13p/PalB family members, the third residue (Asn) being replaced by an Asp residue in five of them, including ScRim13p and YlRim13p (Penas et al. 2007).

Moreover, YlRim13p and AnPalB both carry a calpain III domain toward their C-terminus, which seems absent in ScRim13p. AnPalB also carries an N-terminal microtubule interacting and trafficking (MIT) domain. This domain is necessary and sufficient for PalB to interact specifically and directly with the ESCRT-III

subunit Vps24p (Rodriguez-Galan et al. 2009). The MIT domain deletion impaired but did not prevent the association of PalB to Vps24-containing membranes and its function in PacC proteolytic processing. This indicates that the MIT-dependent association to Vps24p is probably not the sole mechanism by which PalB is recruited to the ESCRT machinery in vivo. Moreover, ScRim13p presumably does not interact with Vps24p because ScRim13p does not contain a MIT domain and *Scvps24*Δ does not prevent Rim101p processing (Xu et al. 2004). YlRim13p does not seem to contain a MIT domain either. Another interactor probably reinforces PalB recruitment to the ESCRT machinery, perhaps Snf7p. Indeed, in *S. cerevisiae*, Rim13p/PalB, like Rim20p/PalA, interacts with Snf7p as initially revealed by a genome-wide two-hybrid screen (Ito et al. 2001). This interaction was confirmed in *Y. lipolytica* (Blanchin-Roland et al. 2008) and *Homo sapiens* where the human calpain 7/PalBH was shown to interact with CHMP4c, another isoform of the homolog of Snf7p (and with the homolog of Vps46/Did2, CHMP1b) (Yorikawa et al. 2008).

3.3.3 YlRim23p/PalC

As mentioned before, the *Y. lipolytica* ortholog of PalC was identified from the genome and named YlRim23p. *YlRIM23* encodes a 441-residue polypeptide (Blanchin-Roland et al. 2008) with 26 % identity with *A. nidulans* PalC (Tilburn et al. 2005). *YlRIM23*, like other *YlRIM* genes, is absolutely required for transcriptional activation of alkaline-induced genes and for repression of acid-induced genes at alkaline pH. The *Ylrim23* null mutant induced the characteristic growth defect of rim mutants, affecting growth at alkaline pH but not at neutral or acidic pH (see below). Like other *YlRIM* genes, *YlRIM23* is therefore required for cellular responses to alkaline ambient pH (Blanchin-Roland et al. 2008).

PalC was shown to contain a functional Bro1 domain (Tilburn et al. 2005) albeit rather distant from those in Bro1p/BroA and Rim20p/PalA (Galindo et al. 2007). The YlRim23p/PalC unique interactant disclosed by a genome-wide two-hybrid screen is YlSnf7p (Blanchin-Roland et al. 2008). At least the N-terminal region of YlSnf7p is required for this interaction. This additional link between the Rim and MVB pathways is conserved in *A. nidulans* where the binding of PalC to Snf7 was shown to involve the PalC Bro1 domain. Moreover, the probable ortholog of YlRim23p/PalC in *S. cerevisiae*, Ygr122wp (Galindo et al. 2007), was previously shown to bind Snf7/Vps32p (Ito et al. 2001; Uetz et al. 2000) and to be required for activation of Rim101p (Rothfels et al. 2005; Barwell et al. 2005). PalC was shown to be recruited to plasma membrane-associated structures in an alkaline pH- and PalH-dependent manner but in a PalA-independent manner (Galindo et al. 2007). Altogether these data are in favour of a recruitment of the ESCRT-associated Rim101p/PacC processing machinery to plasma membrane pH signalling complexes, in response to alkaline ambient pH.

3.3.4 The ESCRT Components

Y. lipolytica was the sole organism where Vps mutants were isolated which drastically affected ambient pH signalling (Blanchin-Roland et al. 2005). As described above, most of the components of the ESCRT machinery are absolutely required for the Rim/Pal signalling pathway, but we identified only two ESCRT-I components in our mutagenesis, YlVps23p and YlVps28p. *YlVPS23* encodes a predicted protein of 476 amino acids displaying 35 % identity with *Homo sapiens* TSG101 (Babst et al. 2000), 34 % with *S. cerevisiae* Vps23p/Stp22p, and 29 % with *C. albicans* Vps23p (Xu et al. 2004). Comparison of the *Y*lVps23p amino acid sequence with that of its orthologs evidenced conservation of the domain structure predicted for ScVps23p and TSG101 (Katzmann et al. 2001): an N-terminal UEV domain (see above) homologous to ubiquitin-conjugating enzymes but unlikely to catalyze ubiquitination because of the substitution for the C residue which binds ubiquitin, a central Pro-rich region, a coiled-coil region, and a conserved C-terminal steadiness box. A functional analysis of YlVps23p was carried out by in-frame deletions of each of these conserved domains to test whether Vps23 functions in the Rim and ESCRT pathways could be separated. The UEV domain, the coiled-coil region, and the C-terminal steadiness box are required for both pathways, whereas the Pro-rich region seems required for neither of them. Therefore, identical domains of YlVps23p are required for both pathways, but the UEV domain was shown to bind Rim8p/PalF in the Rim pathway (see above), whereas it binds Vps27 in the ESCRT pathway (Blanchin-Roland 2011).

YlVPS28 encodes a predicted protein of 246 amino acids that matches putative homologous gene products in other species with following identity scores: 41 % with *C. albicans* Vps28p (Xu et al. 2004), 40 % with *S. cerevisiae* Vps28p (Rieder et al. 1996), and 39 % with *Homo sapiens* Vps28 (Bishop and Woodman 2001). Some regions are well conserved between YlVps28p and its orthologs (Gonzalez-Lopez et al. 2002).

In *Y. lipolytica*, contribution of *YlSNF7* to Rim101 activation could not be assessed due to its essential nature (Blanchin-Roland et al. 2005).

3.4 *Effects of YlRim and YlVps Mutations on Ambient pH Response*

3.4.1 On the Transcription of Alkaline- and Acid-Induced Genes

The alkaline-induced genes tested were *XPR2* encoding the alkaline protease *RIM101*, the expression of which is autoregulated (Lambert et al. 1997), and *YlPHR1* homologous to *C. albicans PHR1* that encodes a glycosylphosphatidylinositol-anchored cell surface protein, similar to β-glycosidase (Saporito-Irwin et al. 1995). At alkaline pH, *YlRim* mutants such as *rim13* (Fig. 1c)

abolished *XPR2* transcription and drastically reduced *RIM101* and *PHR1* expression (Lambert et al. 1997; Gonzalez-Lopez et al. 2002; Blanchin-Roland et al. 2005, 2008). The acid-induced genes tested were *AXP1* encoding the acidic protease (see below), the low-expressed gene *YlHFD1* encoding a putative aldehyde dehydrogenase (Zahedi et al. 2006) the homologs of which in *C. albicans* (D. Onesime, A. Lépingle and C. Gaillardin, unpublished results) and *S. cerevisiae* (Causton et al. 2001) are induced by acidic pH, and *YlPHR2*, a *CaPHR2* homolog, encoding a parietal protein similar to Phr1p and expressed at acidic pH (Muhlschlegel and Fonzi 1997). At alkaline pH, in *YlRim* mutants such as *rim13* (Fig. 1c), expression of *HFD1* or *PHR2* is derepressed compared to the wild-type level at the same pH, reaching or exceeding its level at acidic pH in the reference strain. At acidic pH (Fig. 1c) where the Rim pathway is inactive, *YlRim* mutations have no effect on expression of ambient pH-regulated genes (Gonzalez-Lopez et al. 2002; Blanchin-Roland et al. 2005, 2008).

Particularly *Ylrim9* null mutant has a less extreme phenotype than other *rim* mutants (Gonzalez-Lopez et al. 2002; Blanchin-Roland et al. 2008), like *A. nidulans palI* null mutant (Denison et al. 1998), whereas *rim9* null mutants in *S. cerevisiae* (Su and Mitchell 1993) and in *C. albicans* (Cornet et al. 2009) exhibit a tight Rim phenotype.

As described in *A. nidulans* and *C. albicans*, *rim* mutations mimic acidity by expressing acid-induced genes independently of the ambient pH. The artificial expression of a dominant, constitutively active truncated Rim101p/PacC form (RimSL and rim13 + RimSL in Fig. 1c) is able to activate alkaline-induced genes and to repress acid-induced genes independently of the ambient pH and thus mimics alkalinity, as described in several reports in *C. albicans* (El Barkani et al. 2000), *A. nidulans* (Espeso et al. 2000) and *Y. lipolytica* (Lambert et al. 1997).

Like *YlRIM* genes, at alkaline pH, *YlVPS23* (Fig. 1c) as well as *YlVPS28* are absolutely required for transcriptional activation of alkaline-induced genes and for repression of acid-induced genes. The dominant, constitutively active truncated form of *YlRIM101* which suppresses pH signalling defects of *YlRim* mutations also suppressed *Ylvps* defects in pH response, but not in endocytosis (Fig. 1a, c and Blanchin-Roland et al. 2005).

Unexpectedly, and contrary to the acid-mimicking phenotype of the *rim* mutations, expression of the acidic gene *AXP1* was significantly reduced at acidic pH in the *YlRim* mutants (Lambert et al. 1997; Gonzalez-Lopez et al. 2002). In *A. nidulans*, a residual level of PacC processing occurs irrespective of ambient pH and of pH signal transduction (reviewed in Penalva and Arst 2004). We cannot exclude such a residual processing for YlRim101p, but since *AXP1* expression is significantly reduced in *rim* mutants compared to the wild-type level, the entire Rim pathway would be partially active at acidic pH. In *C. albicans*, Rim101p undergoes a processing event at acidic pH, dependent on the cysteine protease Rim13p, which governs alkaline pH-independent changes in gene expression (Li et al. 2004). The resulting residual level of YlRim101 processed would not be able to repress *AXP1* because the *AXP1* expression is abolished in the null allele *rim101Δ* (Gonzalez-Lopez et al. 2002). All this suggests that low levels of activated Rim101p,

insufficient to turn *XPR2* transcription on, are required for *AXP1* induction at acidic pH. The nature of the signal transmitted by the Rim pathway in these conditions is unclear. It is probable that induction of *AXP1* at acidic pH would be dependent on the Rim pathway not for pH sensing but for interpreting other signals, possibly nutrient limitations. Indeed, effects of *rim* mutations on *AXP1* expression are less pronounced on the protease-inducing medium YPDm than on derepressing Y medium at the same pH (unpublished data).

AXP1 expression remains repressed at neutral/alkaline pH (Lambert et al. 1997); (Gonzalez-Lopez et al. 2002). This lack of *AXP1* expression may reflect a lack of induction. If Axp activity is required to generate specific inducing peptides as reported in *Penicillium roqueforti* for the aspartyl protease ASPA (Gente et al. 2001), no inducer would be generated at neutral/alkaline pH where the enzyme is inactive. Then, interruption of the Rim pathway would not be sufficient to activate *AXP1* transcription. *AXP1* expression may even be totally independent of pH signalling through the Rim pathway, which may nevertheless be required for responding to environmental signals other than pH, as proposed above for its induction at acidic pH.

3.4.2 On the Sensitivity of *Y. lipolytica* to Ambient pH

YlRim mutants strongly affect growth at alkaline but not at neutral or acidic pH (Fig. 1b). However, *Ylrim9* mutants had a less pronounced growth defect (Gonzalez-Lopez et al. 2006; Blanchin-Roland et al. 2008). In a similar way, a *palI* null mutation in *A. nidulans* allows partial growth at pH 8.0, whereas all other *pal* mutations completely prevent growth at this pH (Arst 1994; Denison et al. 1998). This is in stark contrast to the situation reported in *S. cerevisiae* and *C. albicans* where *rim9* mutants exhibit a tight phenotype (Su and Mitchell 1993; Cornet et al. 2009). These data are in good agreement with those reported above concerning the leaky phenotypes of *Ylrim9/AnpalI* mutants for transcription of alkaline- and acid-induced genes, compared to the tight phenotypes of Scrim9/Carim9 mutants.

Growth of the YlVps mutants was slightly impaired at acidic pH and drastically affected at neutral/alkaline pH, at least as much as the YlRim mutants one (Fig. 1c and Blanchin-Roland et al. 2005).

4 Other Functions of the Rim Pathway in *Y. lipolytica* and Other Ascomycetes

Besides its major role in the ambient neutral/alkaline pH adaptation which is conserved in all ascomycetes, the other roles of PacC/Rim101p appear more species-specific and may reflect acquisition of new targets. Whereas the Rim/Pal

pathway appears essential for sporulation in *Y. lipolytica* and *S. cerevisiae* (Li and Mitchell 1997), but not in *A. nidulans*, it is required for mating in *Y. lipolytica* only (Lambert et al. 1997). It controls sporulation in *S. cerevisiae* by activating expression of the meiotic activator *IME1* (Li and Mitchell 1997) and as a corepressor in concert with Nrg1 of expression of *DIT1*, a sporulation-specific gene in mitotic cells (Rothfels et al. 2005).

The Rim/Pal pathway also regulates the cell differentiation during haploid invasive growth and the diploid pseudohyphal differentiation in *S. cerevisiae* (Lamb and Mitchell 2003). It is required for hyphal formation at alkaline pH in *C. albicans*, in particular by repressing expression of the negative regulator *NRG1* in a pH-dependent way (El Barkani et al. 2000; Lotz et al. 2004). In contrast to *S. cerevisiae*, *Y. lipolytica rim* mutants exhibit no change in cell or colony morphology, and their ability to grow invasively is not impaired (Tréton et al. 2000). They do not impair hyphal formation (although hyphal branching frequency is increased), confirming that the Rim pathway is not essential for this transition which is not pH-dependent in *Y. lipolytica* (Tréton et al. 2000; Gonzalez-Lopez et al. 2002), in contrast to *C. albicans* (Fonzi 2002). In *Ustilago maydis*, Rim101p/pacC does not control dimorphism either (Arechiga-Carvajal and Ruiz-Herrera 2005).

Moreover, the resistance of cells to high levels of Ca^+, Na^+, and Li^+ requires activity of the Rim/Pal pathway in *S. cerevisiae* (Lamb and Mitchell 2003), in *C. albicans* (Kullas et al. 2007), in *Fusarium oxysporum* (Caracuel et al. 2003), and in *U. maydis* (Arechiga-Carvajal and Ruiz-Herrera 2005). In particular, activation of *ENA1*, encoding the Na^+-ATPase cation pump, is one way for *S. cerevisiae* adaptation to alkaline pH and/or elevated Li^+ and Na^+ concentrations. *ENA1* is controlled by several signalling pathways, including Rim101 repression of *NRG1* (Ruiz and Arino 2007).

In addition, the Rim101 pathway participates directly to the cell wall assembly and functions in parallel with the protein kinase C (PKC) pathway in *S. cerevisiae* (Castrejon et al. 2006). In *C. albicans*, the Rim pathway represses the acid-induced cell wall gene *RBR1*, encoding a GPI protein required for filamentation, directly or indirectly by downregulation of *NRG1* which is required for *RBR1* expression (Lotz et al. 2004). It regulates the alkaline-induced gene *PHR1* and the acid-induced gene *PHR2*, encoding two functional homologs of the cell wall glycosidases important for adaptation to ambient pH (Muhlschlegel and Fonzi 1997). Like *C. albicans*, *Y. lipolytica* expresses two isoforms of the surface proteins Phr1p/Phr2p with inverted pH regulation patterns, both YlRim101p-dependent (Blanchin-Roland et al. 2005). Moreover, the putative consensus binding site for YlRim101p, GCCARG (Blanchin-Roland et al. 1994; Lambert et al. 1997), was found in the promoters of *YlPHR1* and *YlPHR2*.

It is widely accepted that appropriate responses to ambient pH govern fungal virulence. The alkaline-induced gene *PHR1* and the acid-induced gene *PHR2* are essential to *C. albicans* virulence upon systemic and vaginal routes of murine

infection, respectively (Davis 2003). Additional evidences of the Rim101p/PacC-governed fungal virulence have been obtained in plant pathogenesis by *Colletotrichum acutatum* (You et al. 2007) and *S. sclerotiorum* (Rollins 2003) and in mammal pathogenesis by *C. albicans* (Davis 2003; Cornet et al. 2005; Mitchell et al. 2007; Villar et al. 2007) and by *A. nidulans* (Bignell et al. 2005). In the particular case of *Fusarium oxysporum*, PacC is required for full virulence on mice but not on tomato plants (Ortoneda et al. 2004). Rim101p/pacC is not involved in the virulence of the basidiomycete *U. maydis* on maize (Arechiga-Carvajal and Ruiz-Herrera 2005). The Rim/Pal pathway is thus crucial and seems almost universally integral to fungal pathogenic strategy. It may be a target for preventing fungal pathogenesis of animals (particularly humans) and/or plants.

5 The Current Model of the pH Signalling Pathway in Yeasts and Filamentous Fungi

PalI was shown to be a resident plasma membrane protein (Calcagno-Pizarelli et al. 2007). YlRim9p and AnPalI play an important but not essential role in pH signalling (Denison et al. 1998; Blanchin-Roland et al. 2008; Calcagno-Pizarelli et al. 2007), and overexpressed PalI cannot functionally replace PalH. These data indicate that PalI by itself is not a plasma membrane pH sensor (Calcagno-Pizarelli et al. 2007).

Several data strongly implicate the 7-TMD protein Rim21p/PalH as an ambient pH sensor; (1) while overexpressed PalH-GFP can localize to the plasma membrane, it predominates in cytosolic punctate structures under acidic conditions. PalH requires the coexpression of PalI at similar levels to predominate in the plasma membrane. Thus, PalI would play an accessory role upstream or in concert with Rim21p/PalH (Calcagno-Pizarelli et al. 2007); (2) In *A. nidulans* and *S. cerevisiae*, highly conserved arrestin domains within the arrestin-related protein Rim8p/PalF strongly interact with the C-terminal cytoplasmic tail of Rim21p/PalH (Herranz et al. 2005; Herrador et al. 2010).

In *S. cerevisiae* and in *Y. lipolytica*, Rim8p also appears to bind directly to the ESCRT-I subunit Vps23p, in an ambient pH-independent manner (Herrador et al. 2010; Blanchin-Roland 2011). Moreover, in response to ambient alkaline pH, Rim20p/PalA bound to Rim101p/PacC associates with the ESCRT-III subunit Snf7p (Xu and Mitchell 2001; Vincent et al. 2003; Boysen and Mitchell 2006).

Altogether these data support the following model (Fig. 2). In response to alkaline ambient pH, the interaction of Rim21/PalH with Rim8p/PalF bound to the ESCRT-I subunit Vps23p would promote the ESCRT machinery recruitment to plasma membrane pH signalling sites. And the interaction of the ESCRT-III subunit Snf7 with both Rim20p/PalA bound to Rim101p/PacC and the signalling protease Rim13p/PalB would trigger the pH signal-dependent proteolytic processing of

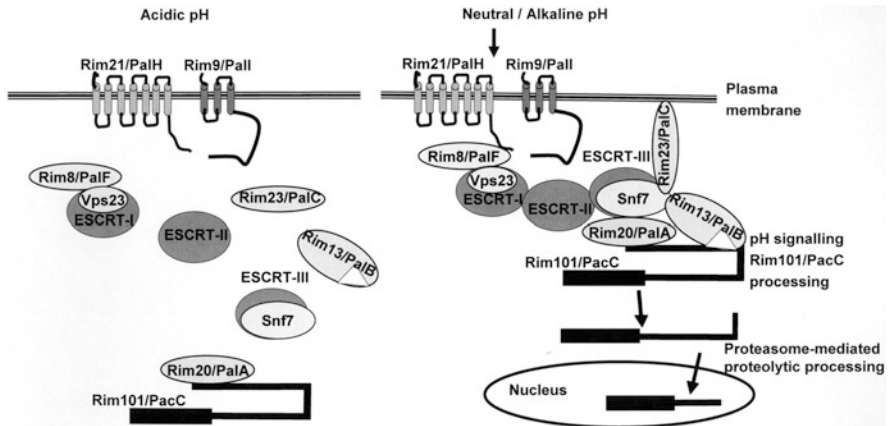


Fig. 2 The current model for pH signalling in *Y. lipolytica* and other ascomycetous yeasts and filamentous fungi. The predicted 7-TMD Rim21/PalH would be the ambient pH sensor that requires the 3-TMD protein Rim9/PalI to predominate at the plasma membrane. In the absence of neutral/alkaline pH signal transduction, Rim8/PalF is bound to the ESCRT-I subunit Vps23 and Rim20/PalA is bound to the C-terminus of Rim101/PacC. In response to the pH signal, Rim8/PalF binds to the cytoplasmic tail of Rim21/PalH, and this would promote the recruitment of ESCRT-I and downstream ESCRTs to plasma membrane. The Rim20/palA binding to Snf7 would mediate the recruitment of Rim101/PacC to the ESCRT machinery which interacts with Rim13/PalB, triggering the pH-dependent signalling proteolysis of Rim101/PacC by this protease. PalC would constitute another link between the ESCRT machinery and plasma membrane. The truncated Rim101/PacC would be activated by a second processing involving the proteasome in *A. nidulans* and possibly in *Y. lipolytica*

Rim101/PacC. This recruitment of the ESCRT-associated Rim101 processing machinery to the plasma membrane pH signalling sites would be reinforced by YlRim23p/PalC that was shown to bind Snf7p (Galindo et al. 2007; Blanchin-Roland et al. 2008) and to localize to cortical punctate structures at the plasma membrane in an ambient pH- and Rim21p/PalH-dependent manner (Galindo et al. 2007).

Such localization of ESCRTs recalls the retroviral budding which involves these protein complexes at the plasma membrane in mammalian cells. The same or related motifs mediate interactions of ESCRT proteins with Rim/Pal proteins in fungi and retroviral Gag proteins in animal cells (Herrador et al. 2010). As retroviruses hijack MVB components to facilitate the final stages of virus budding (Calistri et al. 2009), the pH signalling pathway may hijack MVB components for its Rim101p/PacC processing purposes (Xu et al. 2004; Blanchin-Roland et al. 2005; Boysen and Mitchell 2006; Herrador et al. 2010).

6 Conclusion

Although this pathway has been most intensively investigated in *A. nidulans*, studies in the yeasts *C. albicans*, *S. cerevisiae*, and *Y. lipolytica* have also made important contributions, as described above, in particular about the requirement and the potential role of the ESCRT machinery. A genomic survey confirmed that *Y. lipolytica* is very distantly related to *S. cerevisiae* and *C. albicans* (Dujon et al. 2004). Indeed, this study of the Rim/Pal pathway shows that *Y. lipolytica* is more related to filamentous fungi than to other yeasts: (1) *YlSNF7* is an essential gene (Blanchin-Roland et al. 2005); similarly Vps genes in *A. nidulans* are nearly essential for growth (Calcagno-Pizarelli et al. 2011). In contrast, *S. cerevisiae* and *C. albicans vps* mutants were reported to be viable although many of them present severe growth defects on several types of media (Giaever et al. 2002; Xu et al. 2004); (2) YlRim101p activation probably occurs in two sequential steps (Lambert et al. 1997), like PacC activation in *A. nidulans* (Diez et al. 2002), whereas in *S. cerevisiae*, Rim101p proteolytic activation requires one Rim pathway-dependent proteolytic step (Li and Mitchell 1997; Xu and Mitchell 2001); (3) *pacC*, *YlRIM101*, and *CaRIM101* transcriptions are autoregulated in a pH-dependent manner (Tilburn et al. 1995; Lambert et al. 1997; Bensen et al. 2004), whereas *ScRIM101* does not seem to be alkaline expressed (Lamb et al. 2001); (4) At neutral/alkaline pH, YlRim101p, CaRim101p, and PacC act directly as transcriptional activators of alkaline-induced genes and as transcriptional repressors of acid-induced genes (Gonzalez-Lopez et al. 2002; Blanchin-Roland et al. 2008; Penalva et al. 2008). On the contrary, ScRim101 seems to function only as a repressor; indeed, ScRim101p acts positively in an indirect manner by repressing expression of genes encoding transcriptional repressors, notably *NRG1* and *SMP1* (Lamb and Mitchell 2003); (5) *S. cerevisiae* and *C. albicans* have two homologs of PalH, Rim21p and Dfg16p (Rothfels et al. 2005; Barwell et al. 2005), whereas *Y. lipolytica* like *A. nidulans* carries only one functional YlRim21p/PalH and no equivalent of ScDfg16p (Blanchin-Roland et al. 2008); (6) Whereas *Anpal* and *YlRim* genes are constitutively expressed with respect to ambient pH (Denison et al. 1998; Negrete-Urtasun et al. 1997, 1999; Tréton et al. 2000), *ScRIM8* and *CaRIM8* are acid-expressed genes, *ScRIM8* being directly repressed by the processed form of ScRim101p (Porta et al. 1999; Lamb and Mitchell 2003); (7) Both An*PalI* and *YlRIM9* have an important, yet non essential, role in pH signalling (Denison et al. 1998; Gonzalez-Lopez et al. 2002; Blanchin-Roland et al. 2008), whereas Sc*RIM9* (Su and Mitchell 1993) and *CaRIM9* (Cornet et al. 2009) are absolutely required for Rim-dependent signalling; (8) In *Y. lipolytica* and in filamentous fungi, the Rim9p/PalI factor has a much longer C-terminal tail than in other yeasts. However, this YlRim9p/PalI long C-terminal tail does not seem to be involved in pH response (Blanchin-Roland et al. 2008).

References

Arechiga-Carvajal ET, Ruiz-Herrera J (2005) The RIM101/pacC homologue from the basidiomycete Ustilago maydis is functional in multiple pH-sensitive phenomena. Eukaryot Cell 4:999–1008

Arst HN Jr (1994) Regulation of gene expression by oxygen, phosphorus and pH. Prog Ind Microbiol 29:369–380

Babst M, Odorizzi G, Estepa EJ, Emr SD (2000) Mammalian tumor susceptibility gene 101 (TSG101) and the yeast homologue, Vps23p, both function in late endosomal trafficking. Traffic 1:248–258

Babst M, Katzmann DJ, Estepa-Sabal EJ, Meerloo T, Emr SD (2002) Escrt-III: an endosome-associated heterooligomeric protein complex required for mvb sorting. Dev Cell 3:271–282

Baek YU, Martin SJ, Davis DA (2006) Evidence for novel pH-dependent regulation of Candida albicans Rim101, a direct transcriptional repressor of the cell wall beta-glycosidase Phr2. Eukaryot Cell 5:1550–1559

Barwell KJ, Boysen JH, Xu W, Mitchell AP (2005) Relationship of DFG16 to the Rim101p pH Response Pathway in Saccharomyces cerevisiae and Candida albicans. Eukaryot Cell 4:890–899

Bensen ES, Martin SJ, Li M, Berman J, Davis DA (2004) Transcriptional profiling in Candida albicans reveals new adaptive responses to extracellular pH and functions for Rim101p. Mol Microbiol 54:1335–1351

Bignell E, Negrete-Urtasun S, Calcagno AM, Haynes K, Arst HN Jr, Rogers T (2005) The Aspergillus pH-responsive transcription factor PacC regulates virulence. Mol Microbiol 55:1072–1084

Bishop N, Woodman P (2001) TSG101/mammalian VPS23 and mammalian VPS28 interact directly and are recruited to VPS4-induced endosomes. J Biol Chem 276:11735–11742

Blanchin-Roland S (2011) Identical domains of Yarrowia lipolytica Vps23 are required for both ESCRT and Rim pathways, but the latter needs an interaction between the Vps23 UEV domain and Rim8/PalF. FEMS Yeast Res 11:473–486

Blanchin-Roland S, Cordero Otero RR, Gaillardin C (1994) Two upstream activation sequences control the expression of the XPR2 gene in the yeast Yarrowia lipolytica. Mol Cell Biol 14:327–338

Blanchin-Roland S, Da Costa G, Gaillardin C (2005) ESCRT-I components of the endocytic machinery are required for Rim101-dependent ambient pH regulation in the yeast Yarrowia lipolytica. Microbiology 151:3627–3637

Blanchin-Roland S, Da Costa G, Gaillardin C (2008) Ambient pH signalling in the yeast Yarrowia lipolytica involves YlRim23p/PalC, which interacts with Snf7p/Vps32p, but does not require the long C terminus of YlRim9p/PalI. Microbiology 154:1668–1676

Boysen JH, Mitchell AP (2006) Control of Bro1-domain protein Rim20 localization by external pH, ESCRT machinery, and the Saccharomyces cerevisiae Rim101 pathway. Mol Biol Cell 17:1344–1353

Calcagno-Pizarelli AM, Negrete-Urtasun S, Denison SH, Rudnicka JD, Bussink HJ, Munera-Huertas T, Stanton L, Hervas-Aguilar A, Espeso EA, Tilburn J, Arst HN Jr, Penalva MA (2007) Establishment of the ambient pH signaling complex in Aspergillus nidulans: PalI assists plasma membrane localization of PalH. Eukaryot Cell 6:2365–2375

Calcagno-Pizarelli AM, Hervas-Aguilar A, Galindo A, Abenza JF, Penalva MA, Arst HN Jr (2011) Rescue of Aspergillus nidulans severely debilitating null mutations in ESCRT-0, I, II and III genes by inactivation of salt-tolerance pathway allows examination of ESCRT gene roles in pH signalling. J Cell Sci 124:4064–4076

Calistri A, Salata C, Parolin C, Palu G (2009) Role of multivesicular bodies and their components in the egress of enveloped RNA viruses. Rev Med Virol 19:31–45

Caracuel Z, Casanova C, Roncero MI, Di Pietro A, Ramos J (2003) pH response transcription factor PacC controls salt stress tolerance and expression of the P-Type Na+ -ATPase Ena1 in Fusarium oxysporum. Eukaryot Cell 2:1246–1252

Castrejon F, Gomez A, Sanz M, Duran A, Roncero C (2006) The RIM101 pathway contributes to yeast cell wall assembly and its function becomes essential in the absence of mitogen-activated protein kinase Slt2p. Eukaryot Cell 5:507–517

Causton HC, Ren B, Koh SS, Harbison CT, Kanin E, Jennings EG, Lee TI, True HL, Lander ES, Young RA (2001) Remodeling of yeast genome expression in response to environmental changes. Mol Biol Cell 12:323–337

Cornet M, Bidard F, Schwarz P, Da Costa G, Blanchin-Roland S, Dromer F, Gaillardin C (2005) Deletions of endocytic components VPS28 and VPS32 affect growth at alkaline pH and virulence through both RIM101-dependent and RIM101-independent pathways in Candida albicans. Infect Immun 73:7977–7987

Cornet M, Richard ML, Gaillardin C (2009) The homologue of the Saccharomyces cerevisiae RIM9 gene is required for ambient pH signalling in Candida albicans. Res Microbiol 160:219–223

Davis D (2003) Adaptation to environmental pH in Candida albicans and its relation to pathogenesis. Curr Genet 44:1–7

Davis D, Wilson RB, Mitchell AP (2000) RIM101-dependent and-independent pathways govern pH responses in Candida albicans. Mol Cell Biol 20:971–978

Denison SH, Orejas M, Arst HN Jr (1995) Signaling of ambient pH in Aspergillus involves a cysteine protease. J Biol Chem 270:28519–28522

Denison SH, Negrete-Urtasun S, Mingot JM, Tilburn J, Mayer WA, Goel A, Espeso EA, Penalva MA, Arst HN Jr (1998) Putative membrane components of signal transduction pathways for ambient pH regulation in Aspergillus and meiosis in saccharomyces are homologous. Mol Microbiol 30:259–264

Diez E, Alvaro J, Espeso EA, Rainbow L, Suarez T, Tilburn J, Arst HN Jr, Penalva MA (2002) Activation of the Aspergillus PacC zinc finger transcription factor requires two proteolytic steps. EMBO J 21:1350–1359

Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuveglise C, Talla E, Goffard N, Frangeul L, Aigle M, Anthouard V, Babour A, Barbe V, Barnay S, Blanchin S, Beckerich JM, Beyne E, Bleykasten C, Boisrame A, Boyer J, Cattolico L, Confanioleri F, De Daruvar A, Despons L, Fabre E, Fairhead C, Ferry-Dumazet H, Groppi A, Hantraye F, Hennequin C, Jauniaux N, Joyet P, Kachouri R, Kerrest A, Koszul R, Lemaire M, Lesur I, Ma L, Muller H, Nicaud JM, Nikolski M, Oztas S, Ozier-Kalogeropoulos O, Pellenz S, Potier S, Richard GF, Straub ML, Suleau A, Swennen D, Tekaia F, Wesolowski-Louvel M, Westhof E, Wirth B, Zeniou-Meyer M, Zivanovic I, Bolotin-Fukuhara M, Thierry A, Bouchier C, Caudron B, Scarpelli C, Gaillardin C, Weissenbach J, Wincker P, Souciet JL (2004) Genome evolution in yeasts. Nature 430:35–44

El Barkani A, Kurzai O, Fonzi WA, Ramon A, Porta A, Frosch M, Muhlschlegel FA (2000) Dominant active alleles of RIM101 (PRR2) bypass the pH restriction on filamentation of Candida albicans. Mol Cell Biol 20:4635–4647

Espeso EA, Tilburn J, Sanchez-Pulido L, Brown CV, Valencia A, Arst HN Jr, Penalva MA (1997) Specific DNA recognition by the Aspergillus nidulans three zinc finger transcription factor PacC. J Mol Biol 274:466–480

Espeso EA, Roncal T, Diez E, Rainbow L, Bignell E, Alvaro J, Suarez T, Denison SH, Tilburn J, Arst HN Jr, Penalva MA (2000) On how a transcription factor can avoid its proteolytic activation in the absence of signal transduction. EMBO J 19:719–728

Feuermann M, de Montigny J, Potier S, Souciet JL (1997) The characterization of two new clusters of duplicated genes suggests a 'Lego' organization of the yeast Saccharomyces cerevisiae chromosomes. Yeast 13:861–869

Fonzi WA (2002) Role of pH response in Candida albicans virulence. Mycoses 45:16–21

Futai E, Maeda T, Sorimachi H, Kitamoto K, Ishiura S, Suzuki K (1999) The protease activity of a calpain-like cysteine protease in Saccharomyces cerevisiae is required for alkaline adaptation and sporulation. Mol Gen Genet 260:559–568

Futai E, Kubo T, Sorimachi H, Suzuki K, Maeda T (2001) Molecular cloning of PalBH, a mammalian homologue of the Aspergillus atypical calpain PalB. Biochim Biophys Acta 1517:316–319

Galindo A, Hervas-Aguilar A, Rodriguez-Galan O, Vincent O, Arst HN Jr, Tilburn J, Penalva MA (2007) PalC, one of two Bro1 domain proteins in the fungal pH signalling pathway, localizes to cortical structures and binds Vps32. Traffic 8:1346–1363

Gente S, Billon-Grand G, Poussereau N, Fevre M (2001) Ambient alkaline pH prevents maturation but not synthesis of ASPA, the aspartyl protease from Penicillium roqueforti. Curr Genet 38:323–328

Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B, Arkin AP, Astromoff A, El-Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A, Entian KD, Flaherty P, Foury F, Garfinkel DJ, Gerstein M, Gotte D, Guldener U, Hegemann JH, Hempel S, Herman Z, Jaramillo DF, Kelly DE, Kelly SL, Kotter P, LaBonte D, Lamb DC, Lan N, Liang H, Liao H, Liu L, Luo C, Lussier M, Mao R, Menard P, Ooi SL, Revuelta JL, Roberts CJ, Rose M, Ross-Macdonald P, Scherens B, Schimmack G, Shafer B, Shoemaker DD, Sookhai-Mahadeo S, Storms RK, Strathern JN, Valle G, Voet M, Volckaert G, Wang CY, Ward TR, Wilhelmy J, Winzeler EA, Yang Y, Yen G, Youngman E, Yu K, Bussey H, Boeke JD, Snyder M, Philippsen P, Davis RW, Johnston M (2002) Functional profiling of the Saccharomyces cerevisiae genome. Nature 418:387–391

Glover DJ, McEwen RK, Thomas CR, Young TW (1997) pH-regulated expression of the acid and alkaline extracellular proteases of Yarrowia lipolytica. Microbiology 143:3045–3054

Gonzalez-Lopez CI, Szabo R, Blanchin-Roland S, Gaillardin C (2002) Genetic control of extracellular protease synthesis in the yeast Yarrowia lipolytica. Genetics 160:417–427

Gonzalez-Lopez CI, Ortiz-Castellanos L, Ruiz-Herrera J (2006) The ambient pH response rim pathway in Yarrowia lipolytica: identification of YlRIM9 and characterization of its role in dimorphism. Curr Microbiol 53:8–12

Herrador A, Herranz S, Lara D, Vincent O (2010) Recruitment of the ESCRT machinery to a putative seven-transmembrane-domain receptor is mediated by an arrestin-related protein. Mol Cell Biol 30:897–907

Herranz S, Rodriguez JM, Bussink HJ, Sanchez-Ferrero JC, Arst HN Jr, Penalva MA, Vincent O (2005) Arrestin-related proteins mediate pH signaling in fungi. Proc Natl Acad Sci USA 102:12141–12146

Hervas-Aguilar A, Rodriguez JM, Tilburn J, Arst HN Jr, Penalva MA (2007) Evidence for the direct involvement of the proteasome in the proteolytic processing of the Aspergillus nidulans zinc-finger transcription factor. PACC J Biol Chem 282:34735–34747

Hurley JH, Emr SD (2006) The ESCRT complexes: structure and mechanism of a membrane-trafficking network. Annu Rev Biophys Biomol Struct 35:277–298

Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci USA 98:4569–4574

Katoh K, Shibata H, Suzuki H, Nara A, Ishidoh K, Kominami E, Yoshimori T, Maki M (2003) The ALG-2-interacting protein Alix associates with CHMP4b, a human homologue of yeast Snf7 that is involved in multivesicular body sorting. J Biol Chem 278:39104–39113

Katzmann DJ, Babst M, Emr SD (2001) Ubiquitin-dependent sorting into the multivesicular body pathway requires the function of a conserved endosomal protein sorting complex, ESCRT-I. Cell 106:145–155

Kim J, Sitaraman S, Hierro A, Beach BM, Odorizzi G, Hurley JH (2005) Structural basis for endosomal targeting by the Bro1 domain. Dev Cell 8:937–947

Kullas AL, Li M, Davis DA (2004) Snf7p, a component of the ESCRT-III protein complex, is an upstream member of the RIM101 pathway in Candida albicans. Eukaryot Cell 3:1609–1618

Kullas AL, Martin SJ, Davis D (2007) Adaptation to environmental pH: integrating the Rim101 and calcineurin signal transduction pathways. Mol Microbiol 66:858–871

Lamb TM, Mitchell AP (2003) The transcription factor Rim101p governs ion tolerance and cell differentiation by direct repression of the regulatory genes NRG1 and SMP1 in Saccharomyces cerevisiae. Mol Cell Biol 23:677–686

Lamb TM, Xu W, Diamond A, Mitchell AP (2001) Alkaline response genes of Saccharomyces cerevisiae and their relationship to the RIM101 pathway. J Biol Chem 276:1850–1856

Lambert M, Blanchin-Roland S, Le Louedec F, Lepingle A, Gaillardin C (1997) Genetic analysis of regulatory mutants affecting synthesis of extracellular proteinases in the yeast Yarrowia lipolytica: identification of a RIM101/pacC homolog. Mol Cell Biol 17:3966–3976

Li W, Mitchell AP (1997) Proteolytic activation of Rim1p, a positive regulator of yeast sporulation and invasive growth. Genetics 145:63–73

Li M, Martin SJ, Bruno VM, Mitchell AP, Davis DA (2004) Candida albicans Rim13p, a protease required for Rim101p processing at acidic and alkaline pHs. Eukaryot Cell 3:741–751

Lotz H, Sohn K, Brunner H, Muhlschlegel FA, Rupp S (2004) RBR1, a novel pH-regulated cell wall gene of Candida albicans, is repressed by RIM101 and activated by NRG1. Eukaryot Cell 3:776–784

MacCabe AP, Van den Hombergh JP, Tilburn J, Arst HN Jr, Visser J (1996) Identification, cloning and analysis of the Aspergillus niger gene pacC, a wide domain regulatory gene responsive to ambient pH. Mol Gen Genet 250:367–374

Maccheroni W Jr, May GS, Martinez-Rossi NM, Rossi A (1997) The sequence of palF, an environmental pH response gene in Aspergillus nidulans. Gene 194:163–167

Madzak C, Blanchin-Roland S, Cordero Otero RR, Gaillardin C (1999) Functional analysis of upstream regulating regions from the Yarrowia lipolytica XPR2 promoter. Microbiology 145:75–87

Maurer KC, Urbanus JH, Planta RJ (1995) Sequence analysis of a 30 kb DNA segment from yeast chromosome XIV carrying a ribosomal protein gene cluster, the genes encoding a plasma membrane protein and a subunit of replication factor C, and a novel putative serine/threonine protein kinase gene. Yeast 11:1303–1310

Mitchell BM, Wu TG, Jackson BE, Wilhelmus KR (2007) Candida albicans strain-dependent virulence and Rim13p-mediated filamentation in experimental keratomycosis. Invest Ophthalmol Vis Sci 48:774–780

Muhlschlegel FA, Fonzi WA (1997) PHR2 of Candida albicans encodes a functional homolog of the pH-regulated gene PHR1 with an inverted pattern of pH-dependent expression. Mol Cell Biol 17:5960–5967

Negrete-Urtasun S, Denison SH, Arst HN Jr (1997) Characterization of the pH signal transduction pathway gene palA of Aspergillus nidulans and identification of possible homologs. J Bacteriol 179:1832–1835

Negrete-Urtasun S, Reiter W, Diez E, Denison SH, Tilburn J, Espeso EA, Penalva MA, Arst HN Jr (1999) Ambient pH signal transduction in Aspergillus: completion of gene characterization. Mol Microbiol 33:994–1003

Neuveglise C, Nicaud JM, Ross-Macdonald P, Gaillardin C (1998) A shuttle mutagenesis system for tagging genes in the yeast Yarrowia lipolytica. Gene 213:37–46

Ogrydziak DM (1993) Yeast extracellular proteases. Crit Rev Biotechnol 13:1–55

Ogrydziak DM, Demain AL, Tannenbaum SR (1977) Regulation of extracellular protease production in Candida lipolytica. Biochim Biophys Acta 497:525–538

Ortoneda M, Guarro J, Madrid MP, Caracuel Z, Roncero MI, Mayayo E, Di Pietro A (2004) Fusarium oxysporum as a multihost model for the genetic dissection of fungal virulence in plants and mammals. Infect Immun 72:1760–1766

Otero RC, Gaillardin C (1996) Dominant mutations affecting expression of pH-regulated genes in Yarrowia lipolytica. Mol Gen Genet 252:311–319

Penalva MA, Arst HN Jr (2002) Regulation of gene expression by ambient pH in filamentous fungi and yeasts. Microbiol Mol Biol Rev 66:426–446

Penalva MA, Arst HN Jr (2004) Recent advances in the characterization of ambient pH regulation of gene expression in filamentous fungi and yeasts. Annu Rev Microbiol 58:425–451

Penalva MA, Tilburn J, Bignell E, Arst HN Jr (2008) Ambient pH gene regulation in fungi: making connections. Trends Microbiol 16:291–300

Penas MM, Hervas-Aguilar A, Munera-Huertas T, Reoyo E, Penalva MA, Arst HN Jr, Tilburn J (2007) Further characterization of the signaling proteolysis step in the Aspergillus nidulans pH signal transduction pathway. Eukaryot Cell 6:960–970

Porta A, Ramon AM, Fonzi WA (1999) PRR1, a homolog of Aspergillus nidulans palF, controls pH-dependent gene expression and filamentation in Candida albicans. J Bacteriol 181:7516–7523

Ramon AM, Fonzi WA (2003) Diverged binding specificity of Rim101p, the Candida albicans ortholog of PacC. Eukaryot Cell 2:718–728

Rieder SE, Banta LM, Kohrer K, McCaffery JM, Emr SD (1996) Multilamellar endosome-like compartment accumulates in the yeast vps28 vacuolar protein sorting mutant. Mol Biol Cell 7:985–999

Rodriguez-Galan O, Galindo A, Hervas-Aguilar A, Arst HN Jr, Penalva MA (2009) Physiological involvement in pH signaling of Vps24-mediated recruitment of Aspergillus PalB cysteine protease to ESCRT-III. J Biol Chem 284:4404–4412

Rollins JA (2003) The Sclerotinia sclerotiorum pac1 gene is required for sclerotial development and virulence. Mol Plant Microbe Interact 16:785–795

Rollins JA, Dickman MB (2001) pH signaling in Sclerotinia sclerotiorum: identification of a pacC/RIM1 homolog. Appl Environ Microbiol 67:75–81

Rothfels K, Tanny JC, Molnar E, Friesen H, Commisso C, Segall J (2005) Components of the ESCRT pathway, DFG16, and YGR122w are required for Rim101 to act as a corepressor with Nrg1 at the negative regulatory element of the DIT1 gene of Saccharomyces cerevisiae. Mol Cell Biol 25:6772–6788

Ruiz A, Arino J (2007) Function and regulation of the Saccharomyces cerevisiae ENA sodium ATPase system. Eukaryot Cell 6:2175–2183

Sadowski L, Pilecka I, Miaczynska M (2009) Signaling from endosomes: location makes a difference. Exp Cell Res 315:1601–1609

Saporito-Irwin SM, Birse CE, Sypherd PS, Fonzi WA (1995) PHR1, a pH-regulated gene of Candida albicans, is required for morphogenesis. Mol Cell Biol 15:601–613

Serrano R, Bernal D, Simon E, Arino J (2004) Copper and iron are the limiting factors for growth of the yeast Saccharomyces cerevisiae in an alkaline environment. J Biol Chem 279:19698–19704

Slagsvold T, Pattni K, Malerod L, Stenmark H (2006) Endosomal and non-endosomal functions of ESCRT proteins. Trends Cell Biol 16:317–326

Su SS, Mitchell AP (1993) Identification of functionally related genes that stimulate early meiotic gene expression in yeast. Genetics 133:67–77

Suarez T, Penalva MA (1996) Characterization of a Penicillium chrysogenum gene encoding a PacC transcription factor and its binding sites in the divergent pcbAB-pcbC promoter of the penicillin biosynthetic cluster. Mol Microbiol 20:529–540

Tilburn J, Sarkar S, Widdick DA, Espeso EA, Orejas M, Mungroo J, Penalva MA, Arst HN Jr (1995) The Aspergillus PacC zinc finger transcription factor mediates regulation of both acid- and alkaline-expressed genes by ambient pH. EMBO J 14:779–790

Tilburn J, Sanchez-Ferrero JC, Reoyo E, Arst HN Jr, Penalva MA (2005) Mutational analysis of the pH signal transduction component PalC of Aspergillus nidulans supports distant similarity to BRO1 domain family members. Genetics 171:393–401

Tréton B, Blanchin-Roland S, Lambert M, Lepingle A, Gaillardin C (2000) Ambient pH signalling in ascomycetous yeasts involves homologues of the Aspergillus nidulans genes palF and palH. Mol Gen Genet 263:505–513

Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamoda G, Yang M, Johnston M, Fields S, Rothberg JM (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature 403:623–627

Villar CC, Kashleva H, Nobile CJ, Mitchell AP, Dongari-Bagtzoglou A (2007) Mucosal tissue invasion by Candida albicans is associated with E-cadherin degradation, mediated by transcription factor Rim101p and protease Sap5p. Infect Immun 75:2126–2135

Vincent O, Rainbow L, Tilburn J, Arst HN Jr, Penalva MA (2003) YPXL/I is a protein interaction motif recognized by aspergillus PalA and its human homologue, AIP1/Alix. Mol Cell Biol 23:1647–1655

Vito P, Pellegrini L, Guiet C, D'Adamio L (1999) Cloning of AIP, a novel protein that associates with the apoptosis-linked gene ALG-2 in a Ca^{2+}-dependent reaction. J Biol Chem 274:1533–1540

Xu W, Mitchell AP (2001) Yeast PalA/AIP1/Alix homolog Rim20p associates with a PEST-like region and is required for its proteolytic cleavage. J Bacteriol 183:6917–6923

Xu W, Smith FJ Jr, Subaran R, Mitchell AP (2004) Multivesicular body-ESCRT components function in pH response regulation in Saccharomyces cerevisiae and Candida albicans. Mol Biol Cell 15:5528–5537

Yorikawa C, Takaya E, Osako Y, Tanaka R, Terasawa Y, Hamakubo T, Mochizuki Y, Iwanari H, Kodama T, Maeda T, Hitomi K, Shibata H, Maki M (2008) Human calpain 7/PalBH associates with a subset of ESCRT-III-related proteins in its N-terminal region and partly localizes to endocytic membrane compartments. J Biochem 143:731–745

You BJ, Choquer M, Chunk KR (2007) The Colletotrichum acutatum gene encoding a putative pH-responsive transcription regulator is a key virulence determinant during fungal pathogenesis on citrus. Mol Plant Microbe Interact 20:1149–1160

Zahedi RP, Sickmann A, Boehm AM, Winkler C, Zufall N, Schonfisch B, Guiard B, Pfanner N, Meisinger C (2006) Proteomic analysis of the yeast mitochondrial outer membrane reveals accumulation of a subclass of preproteins. Mol Biol Cell 17:1436–1450

The Gpr1/Fun34/YaaH Protein Family in the Nonconventional Yeast *Yarrowia lipolytica* and the Conventional Yeast *Saccharomyces cerevisiae*

Falk Matthäus and Gerold Barth

Abstract The sequencing of the genomes of several organisms was the first step in understanding genome organisation, gene function and of course life itself. Until now very much information could be gained by searching for similarities between genes or structures and finding homologies or direct orthologous genes. But with the sequencing data, more and more questions arise, which cannot be answered by simply looking the data. Although the yeast *Saccharomyces cerevisiae* was the first eukaryotic organism, the genome of which was sequenced, almost 14 years later there are still 900 uncharacterised and 800 dubious genes (26 % at all). The unconventional yeast *Yarrowia lipolytica* also harbours a high number of genes with unknown function—ca. 2,300 (35 %). One hundred seventy out of 220 *Yarrowia lipolytica* specific genes with no homology to other yeasts are also uncharacterised, which should carry the "differences" between the yeast species. So far only 50 industrial useful genes were characterised (proteases, lipases, esterases, etc.).

Genes or gene products that do not have any known function have to be analysed in a more classical, genetical way. Especially the functions of membrane proteins which are resistant to a lot of investigation steps are only rarely elucidated. The Gpr1 protein of *Yarrowia lipolytica* is such an example. The function cannot be estimated by finding similarities to other proteins with known function. The more the information and phenotypic effects are available, the more complex the interacting network appears. This chapter will show the ongoing discovery of the function of the Gpr1/FUN34/YaaH protein family. This is especially shown for the Gpr1 protein from *Yarrowia lipolytica* and its orthologues in *Saccharomyces cerevisiae*. Here the different and also controversial facts are reviewed and discussed.

F. Matthäus • G. Barth (✉)
Institute of Microbiology, Dresden University of Technology, Zellescher Weg 20b, 01062 Dresden, Germany
e-mail: Gerold.Barth@TU-Dresden.de

G. Barth (ed.), *Yarrowia lipolytica*, Microbiology Monographs 24,
DOI 10.1007/978-3-642-38320-5_7, © Springer-Verlag Berlin Heidelberg 2013

Contents

1 Introduction 146
2 History of the Gpr1/FUN34/YaaH Family 147
3 The Gpr1/FUN34/YaaH Protein Family 148
4 Gpr1p and Its Homologues in *Yarrowia lipolytica* 150
 4.1 Expression of *GPR1* 152
 4.2 Putative Functions 152
5 Gpr1p Orthologues in *Saccharomyces cerevisiae* 155
 5.1 Localisation of the Three Gpr1p Orthologues 155
 5.2 Expression 156
 5.3 Postulated Functions 158
6 Discussion 160
References 161

1 Introduction

The cell membrane as communicating structure between cell inside and outside environment harbours several distinct functional regions regulating many intracellular as well as cell-to-cell communicating processes. These processes are based on complex interactions of single proteins and strongly organised protein complexes with other molecules ensuring a highly sophisticated cellular diversity in response to intracellular as well as to extracellular environmental changes. There are a lot of channels, transporters, receptors and other proteins of signal transduction cascades which are often localised to detergent-resistant membrane domains but may also be distributed all over the membrane. In a global analysis of complete genomes of five hemiascomycete yeasts, De Hertogh et al. (2006) have shown that on average about 18 % of all proteins encoded in the five genomes are predicted membrane proteins localised to different membranes. On the basis of sequence comparisons as well as functional and phylogenetic criteria, several families, subfamilies and singletons of membrane proteins have been identified. In this way 850 predicted membrane proteins of *Saccharomyces cerevisiae* have been ordered to the Transporter Classification (TC) system including all transporter proteins and the Membrane Classification (MC) system, which includes non-transporter membrane proteins (De Hertogh et al. 2002). Many membrane proteins are already well studied and their function could be clarified or suggested. However, there are some of these proteins of which the function is still unknown or results of functional studies are contradictory at least. The Gpr1/FUN34/YaaH is one of such a protein family, of which only a few members have been investigated and results of functional studies are divergent, in spite of the quickly growing number of detected members. Interestingly, members of this protein family could be detected in different species of *Bacteria* and *Archaea* but are restricted to lower developed species in the kingdom of *Eukarya*. However, although phenotypic effects caused by several mutations in the Gpr1p encoding gene of *Yarrowia lipolytica* were detected about two decades

ago, for the first time, no unambiguous theory on the function of these proteins or other members of this class of proteins exists. Extensive studies on intracellular localisation and function were only done in *Y. lipolytica* and *Saccharomyces cerevisiae*, however, resulting in contradictory hypotheses on putative functions of these proteins.

The aim of this review is to give an overview on the hitherto known data of cytological, genetical and functional studies of the Gpr1 protein and its five homologues in *Yarrowia lipolytica* and to compare these data with results of studies on members of these protein family in *S. cerevisiae* to get hints on the putative function/functions of these membrane proteins.

2 History of the Gpr1/FUN34/YaaH Family

In the late 1980s, different groups isolated and characterised acetate non-utilising mutants (Acu$^-$) of the yeast *Y. lipolytica* (Matsuoka et al. 1980; Ogrydziak et al. 1982; Barth and Weber 1985, 1987). Kujau et al. (1992) published results of an extensive genetical and physiological study including 550 of such mutants which were selected after chemical mutagenesis of strain B204-12C by *N*-methyl-*N′*-nitro-*N*-nitrosoguanidine. Based on the utilisation of acetic acid, ethanol and *n*-alkans, four phenotypic classes were defined, analysed for the activity of acetyl-CoA-synthetase (ACS), isocitrate lyase (ICL) and malate synthase (MAS) and further divided into different groups after complementation analysis.

Eight mutants out of these exhibited a dominant Acu$^-$ phenotype, which was generated by mutations in only one gene. Incubation on acetic acid-containing medium led to decreased activity of the enzymes of the glyoxylate shunt and of ACS. Therefore, the unknown gene was named *GPR1* (*g*lyoxylate *p*athway *r*egulation). Subsequent cloning and sequencing of the wild-type allele revealed an ORF of 813 base pairs encoding a protein of 270 amino acids. Four of these dominant mutant strains harbouring different alleles of *GPR1* were phenotypically characterised in more detail, and several pleiotropic effects were observed (Kujau et al. 1992; Barth and Scheuber 1993; Tzschoppe et al. 1999).

The first orthologue of *GPR1*, the gene *YCR010c*, formally *YCR10c*, was found during mapping of chromosome III of *Saccharomyces cerevisiae* (Yoshikawa and Isono 1990). First data about size of transcript length and expression of this gene were generated. Later Skala et al. (1992) postulated six transmembrane domains of the encoded protein and detected different regulation sites up- and downstream of the gene. *FUN34* (*YNR002c*), a homologous gene of *YCR010c* in *S. cerevisiae*, which is suggested to be emerged from interchromosomal duplication of *YCR010c* (Lalo et al. 1993), was firstly described in more detail by Lalo et al. (1994). The third *S. cerevisiae* homologue *YDR384c* was discovered with the yeast genome projects (Brandt et al. 1996; Goffeau et al. 1996; Jacq et al. 1997), but without any functional hints.

A further orthologous protein to Gpr1p represents YaaHp from *E. coli*. Although high homology is known for several years, the published data are limited to the results of sequencing approaches (Yura et al. 1992; James et al. 1993; Blattner et al. 1997; Hayashi et al. 2006) and global inner membrane proteome topology (Daley et al. 2005).

On the basis of highly conserved sequence features among these homologous proteins, the new Gpr1/FUN34/YaaH protein family was created in 1995. As presented below, the growing number of available genome sequences strongly increases the number of orthologous proteins belonging to this family. However, the data on function of any of these proteins are still controversial and will therefore be discussed in detail in the following section.

3 The Gpr1/FUN34/YaaH Protein Family

The genes of the Gpr1/FUN34/YaaH protein family encode highly conserved proteins, which were found in archaea and eubacteria as well as in lower eukaryotic organisms. Higher eukaryotic organisms do not contain any proteins containing the Gpr1/FUN34/YaaH protein family motif so far known hitherto, which makes this protein family interesting in concern of evolution and medical application.

So far, the database of the "National Center for Biotechnology Information" contains 968 entries of proteins belonging to the Gpr1/FUN34/YaaH protein family in 173 species (double entries removed). Unfortunately, this database includes redundancy and cross links to proteins that share only single domains.

The family motif is NP[AV]P[LF]GLX[GSA]F and the 223 proteins related to this family could be found within the Prosite database. One hundred forty-one are bacterial proteins (66 after duplicate removal), 18 from archaea and 64 from eukaryota (63 fungi). Interestingly, all bacteria and archaea harbour a single Gpr1 orthologue harbouring the family motif, whereas 16 of the 34 fungi harbour at least two orthologous proteins.

The most conspicuous difference between all of these sequences is that almost all prokaryotic Gpr1p orthologues do not have a hydrophilic N-terminus. So far only *Desulfuromonas acetoxidans* harbours a family member with an extended N-terminus (ca. 84 aa longer than other bacterial orthologues). But it should be noticed that the data are preliminary data from whole-genome shotgun of this organism.

The biggest group of Gpr1p orthologous protein containing species is the group of prokaryotes, led by the bacteria. The homology of the proteins is very high, and in contrast to the fungi, the prokaryotic orthologous proteins do not absolutely cluster to its phylogenetic classes in alignment studies (data not shown). Although YaaH from *E. coli* contributes to the family name, nothing is known about its function.

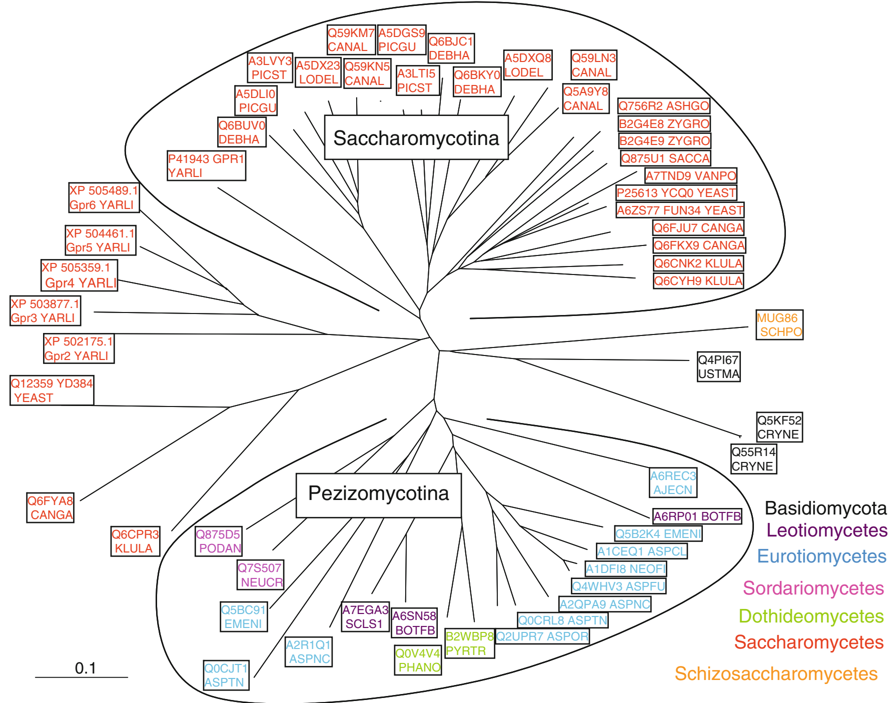


Fig. 1 Phylogeny tree of the Gpr1/FUN34/YaaH family motif containing fungi from UniProt protein database without fragment entries and reduced redundancy. Sequences were aligned and tree was drawn by ClustalX 1.83 (Thompson et al. 1997) and Treeview (Page 1996)

Interestingly, fungi often harbour more than one homologue. In some species, a higher number of homologous genes can be found (*Candida albicans* 4, *Debaryomyces hansenii* 3). Also in *Y. lipolytica* 5 genes homologous to *GPR1* could be detected in the entire genome of strain CLIB 122. Besides 37 different ascomycetous fungi, also the basidiomycetous *Cryptococcus neoformans* and *Ustilago maydis* contain proteins belonging to this protein family.

In Fig. 1 a phylogeny tree is shown which only includes the fungal group of the Gpr1/FUN34/YaaH protein family and the homologues of Gpr1p from *Y. lipolytica*. As visible the proteins cluster mainly to its phylogenetic classes. The proteins of Ascomycota (Pezizomycotina and Saccharomycotina) and Basidiomycota build individual branches, except the orthologous protein of ascomycetous yeast *Schizosaccharomyces pombe* which clusters to Basidiomycota.

Interestingly, the members of the Saccharomycotina show higher diversity, than the other groups. In this group five sub-branches could be defined, wherein all five Gpr1p homologues from *Y. lipolytica* could be found in one sub-branch together with each an orthologue of *S. cerevisiae*, *K. lactis* and *C. glabrata*. However, Gpr1p itself builds up an individual sub-branch. The proteins of the Sordariomycetes group form a sub-branch in the Eurotiomycetes group. The two Dothideomycetes

proteins Q0V4V4p of *Phaeosphaeria nodorum* and B2WBP8 of *Pyrenophora tritici-repentis* can also be found within the Eurotiomycetes. The great distance of Ydr384cp to its homologous proteins from *S. cerevisiae* may be a hint for different functions. Here also the five homologues from *Y. lipolytica* could be found, which show highest homology to each homologue. All of the five homologues of Gpr1p in *Y. lipolytica* exhibit higher homologies to Ydr384cp of *S. cerevisiae* than to Ycr010cp and Ynr002cp. These differences in homologies and resulting clustering of these proteins may suggest different functions for Ydr384cp and the homologues of Gpr1p on the one hand and Gpr1p as well as Ycr010cp and Ynr002cp on the other hand.

4 Gpr1p and Its Homologues in *Yarrowia lipolytica*

The protein Gpr1p consists of 270 amino acids. It is an intra-membrane protein consisting of five to six putative transmembrane domains. Localisation approaches by GFP fusion under the control of the native Gpr1 promoter showed plasma membrane localisation after 2 h and increasing vesicular and vacuolar localisation, which may be connected with recycling and/or degradation processes of the protein, after 4 h of incubation in 30 mM acetic acid-containing medium (Tzschoppe 1998; Augstein 2001; Gentsch 2005). Furthermore, the mutant forms of Gpr1p are located in the plasma membrane (Tzschoppe 1998). Also cell fractionation by differentially centrifugation confirmed plasma membrane localisation (Gentsch 2005). Moreover, lipid raft association could be shown by cold Triton-X 100 extraction and sucrose gradient centrifugation (Buchweitz 2007). Plasma membrane localisation and detergent-resistant membrane association were shown in a similar way for all three *S. cerevisiae* orthologues, where Ycr010cp was pH dependent and Ydr384cp continuously localised in plasma membrane patches and Ynr002cp was evenly distributed in the plasma membrane (Ricicova et al. 2007).

Phytochelatins fused to the first and third transmembrane domain were able to bind extracellular silver, whereas fusion to the second domain did not (Gentsch 2005). This observation supports intracellular termini localisation. Furthermore, Gentsch and Barth (2005) have detected a phosphorylation site at Ser_{37} and a further on Ser_{17} or Ser_{19}, shown by mass spectrometry data (Matthäus, unpublished), of Gpr1p. Ser_{37} is located before the first transmembrane domain and is phosphorylated/dephosphorylated dependent on available carbon source and growth phase. Presupposing that phosphorylation/dephosphorylation of this protein takes place intracellularly, these data support the results of Gentsch (2005) and suggest an intracellular localisation of the N-terminus of Gpr1p (Fig. 2).

G248D and L65Q (Gpr1-1 and Gpr1-2, respectively) amino acid substitution mutagenesis leads to a dominant acetic acid sensitivity, which was shown by

```
1          11         21         31         41         51
MNTEIPDLEK QQIDHNSGSD DPQPIHDDMA PVSRIRSSGP NHEYIHIADQ KFHRDDFYRA
                                  Pho.-Site
61         71         81         91         101        111
FGGTLNPGGA PQPSRKFGNP APLGLSAFAL TTLVFSLCTV QARGVPNPSI AVGLALFYGG
                   family-motif
121        131        141        151        161        171
VCQFAAGMWE FVQENTFGAA ALTSYGGFWM SWAAIEMNAF GIKDSYNDPI EVQNAVGIYL

181        191        201        211        221        231
FGWFIFTLML TLCTLKSTVA FFGLFFMLMM TFLVLACANV TQHHGTAIGG GWLGIITAFF

241        251        261
GFYNAYAGLA NPGNSYIVPV PLDMPFVKKD
 Dim.-Site
```

Fig. 2 Primary amino acid sequence of Gpr1p from *Y. lipolytica*, *grey background* shows hypothetical transmembrane domains (http://www.enzim.hu/hmmtop/), *red mark* is the carbon source-dependent phosphorylation site, *black background* shows trans-dominant acetic acid sensitivity mediating sites in *Y. lipolytica*, *bold underline* is the Gpr1/FUN34/YaaH family motif, and *bold* YNAYA motif shows the putative dimerisation sequence

characterisation of the acetic acid sensitive mutants B204-12C-156 and B204-12C-112 (Augstein 2001). Due to further mutagenesis experiments, mutations of F61L, G62S, G63D, N66S, S197 and T198 were found to exhibit acetic acid hypersensitivity, too (Kuschel 2006).

Although N-terminal sequences often determine final protein localisation, no localisation signals could be identified in this part of Gpr1p. Also no N-terminal processing (cleavage) was shown, because mass spectrometry data revealed acetylation of the start methionine of native Gpr1p (unpublished data). Furthermore, deletion of the N-terminus up to A60 did not show any effect on Gpr1p function. In contrast, complete deletion including F61 or even the single deletion or mutation F61E mediates acetic acid sensitivity. An exchange of F61 by the structurally similar tyrosine has no influence on growth and acetic acid sensitivity (Augstein 2001; Augstein et al. 2003). The primary sequence of Gpr1p contains a possible ER retention sequence (F-V-K-K) (Augstein 2001). The lack of N-terminal signal sequence (SignalP and TargetP program, Emanuelsson et al. 2000; Bendtsen et. al. 2004b) and the membrane associated location let suppose a non-classical secretion mechanism of this protein. This was further confirmed by SecretomeP (Bendtsen et al. 2004a) that gave a NN-score of 0.654. Moreover, all homologues of Gpr1p in *Y. lipolytica* show values above 0.5, the lower limit for non-classical secretion. The proteins YALI0C04257p, YALI0C12034p and YALI0E05973p, homologous to the putative non-classical transport proteins Nce1p, Nce2p and Ste6p of *S. cerevisiae*, seem not to be involved in the membrane localisation of this protein. Deletions of the genes did not result in an abolished mutant phenotype that would be expected when delocalisation occurs (Werner 2009).

4.1 *Expression of* GPR1

The regulation of the native *GPR1* promoter was investigated using *lacZ* as reporter gene. The transcription is increased by ethanol and acetate about two- to threefold and repressed by oleic acid, glucose and most strongly by glycerol. Nevertheless, Gpr1p is not essential for growth on these carbon sources, because deletion of *GPR1* shows no effect even at pH 4. The cultivation on minimal medium with lactate or pyruvate as sole carbon source causes derepression of the gene. On media with glucose and other carbon sources, low activities were observed. Increased promoter activity is detectable when glucose starvation is reached. Such an increased activity was also observed in media without any carbon source. These results suggest that the *GPR1* promoter may be more regulated by derepression than by induction (Augstein 2001; Augstein et al. 2003). The promoter studies were confirmed by Western blot analysis. *GPR1* expression strongly increases after transfer to minimal media with or without carbon source. Induction by acetate was delayed about 4 h after addition of acetate to the culture. Complex media shows always very low *GPR1* expression (Gentsch 2005).

Deletion of promoter regions, gel shift experiments and β-galactosidase activity assay of Gpr1p-*lacZ* fusion protein have shown that a FACB-binding motif (acetate regulatory DNA binding protein) similar to *Emericella nidulans* including a potential CSRE motif (carbon source responsive element, Cat8p/Sip4p-binding domain) similar to *S. cerevisiae* is present in the promoter and has great influence on promoter activity (Augstein 2001). Despite the fact that *Y. lipolytica* possesses only a hypothetical protein (YALI0C19151gp) with weak similarity to Cat8p or FacBp, similar transcription factors could be responsible for the regulation of the expression of *GPR1* in *Y. lipolytica*. In the promoter region of the five homologues of *GPR1* from *Y. lipolytica*, which were presented below in more detail, no FACB binding but CSRE motifs are predicted. This is also the case for the promoter regions of the *YCR010c* homologues of *S. cerevisiae* (http://www.genomatix.de/cgi-bin/matinspector/matinspector.pl).

4.2 *Putative Functions*

Within 10 min after the addition of acetate or acetic acid, Gpr1p was completely phosphorylated at Ser_{37} independently on the pH value of the cultivation medium. Gpr1p undergoes fast dephosphorylation of Ser_{37} after depletion of a carbon source. In the presence of glucose as carbon source in minimal medium, Gpr1p was only partially phosphorylated in contrast to acetic acid as carbon source. In almost the same manner, glycerol, ethanol or oleic acid led to partial phosphorylation of Gpr1p as in glucose-containing medium. This phosphorylation has no pH, osmolarity or nitrogen source-dependent behaviour (Gentsch and Barth 2005).

Several mutations mediate dominant acetic acid sensitivity (GPR^d phenotype) at pH values below 6 even in the presence of glucose (Tzschoppe et al. 1999). This effect is also present in cells with strongly reduced expression of the mutant alleles as well as in the presence of a 100-fold overexpressed wild-type Gpr1p (Augstein et al. 2003). Augstein et al. (2003) showed that the GPR^d mutant strain *Y. lipolytica* B204-12C-112 carries a single base mutation in the gene *GPR1* that led to an amino acid substitution (G248D). Also in the three further GPR^d mutants B204-12C-38, -124 and -156, the polar amino acids L65Q, G62S and G63D were exchanged, respectively. Surprisingly, deletion and overexpression of *GPR1* in *Y. lipolytica* strains do not induce acetic acid sensitivity and have no obvious phenotypic effects (Augstein 2001) except a prolonged lag phase of growth after transfer into fresh media (Gentsch et al. 2007).

The deletion of the family motif NPAPLGL in the *Gpr1-2* mutant allele abolishes acetic acid sensitivity and results in the occurrence of only the dephosphorylated form of Gpr1p (Gentsch et al. 2007). This is the first hint for a functional significance of this highly conserved family motif. In spite of that, the mode of action remains to be clarified.

Three of the original isolated and further studied four *GPR1*-mutants alleles have acetic acid sensitivity mediating amino acid changes in the N-terminal region which is not present in the prokaryotic members, whereas the mutation of the fourth mutant is located in the C-terminal region. Furthermore, when mutations are introduced into the F-G-G-[TE]-[LF]-[NQ]-motif, which is highly conserved in the fungal group, acetic acid sensitivity is exhibited in *Y. lipolytica* (Augstein et al. 2003) as well as in *S. cerevisiae* (Gentsch et al. 2007).

C-terminal deletions in Gpr1-2p up to Y_{256} or A_{247} expressed extrachromosomal did not change acetic acid sensitivity in *Y. lipolytica* PO1d but recovered acetic acid tolerance in the *gpr1* deleted strain *Y. lipolytica* PO1d*Δgpr1*. Deletion from the C-terminus up to F_{242} or deletion of the region Y_{243}-A_{247} (YNAYA) in Gpr1-2p leads to recovery of acetic acid tolerance both in *Y. lipolytica* PO1d and in *Y. lipolytica* PO1d*Δgpr1*. This suggests that an interaction of the C-terminal-truncated Gpr1-2p and the chromosomal-encoded wild-type form of Gpr1p is necessary to express the dominant phenotype. Furthermore, this implicates that Gpr1p forms oligomers and that the YNAYA motif is needed for the interaction of monomers.

A further interaction could take place with the FGGTLN-motif of neighboured Gpr1p. Mutations in this region (e.g. Grp1-2p) could interfere with the interaction (may be by conformational changes) and thus activate Gpr1p what could further inhibit adaptation to acetic acid stress (Gentsch 2005). Studies by Blue Native and 2D gel electrophoresis have shown that Gpr1p is involved in a membrane-located protein complex of about 440 kDa in size (Schlegel 2005; Buchweitz 2007). This native protein complex could not be found, when the YNAYA motif was deleted in the Gpr1 protein (Fuchs 2010).

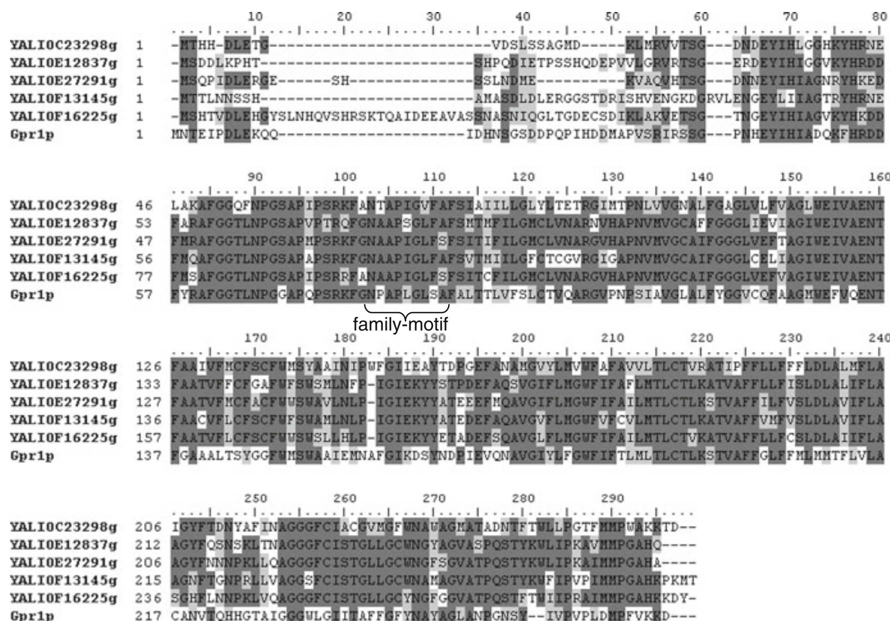


Fig. 3 Sequence alignment of Gpr1p and its homologues. *Dark grey* and *light grey* are the identical and similar amino acids, respectively. Alignment was performed using ClustalW (Thompson et al. 1997) and Bioedit (Hall 1999)

As mentioned above, five further genes encoding proteins homologous to Gpr1p are present in the genome of strain *Y. lipolytica* CLIB 122. Every two of the homologous genes are located on the same chromosome. The sequence alignment is shown in Fig. 3.

The amino acid identity of the homologous proteins of *Y. lipolytica* to Gpr1p ranges from 39 to 44 %. Although the homologies are high, family motif is found modified in the amino acid sequences of the homologous proteins. Four of the five homologues show amino acid exchange on position 2 and 5. YALI0C23298gp shows a further exchange at position 7. Higher homology was found in between the homologous proteins (56–79 %) than to Gpr1p. Higher diversity was found in the N-terminal parts of these proteins. These may suggest different localisation within the cell and/or different function of the proteins. Expression levels of the homologous proteins, investigated by *lacZ* fusion, were significantly lower than that of Gpr1p. Gpr2p (Yali0e27291p) showed the highest expression levels of the homologues on nearly all tested carbon sources and conditions (32 % and 37 % of Gpr1p expression after 4 h of incubation on 1 % ethanol and 30 mM acetate, respectively). Gpr3p (Yali0f13145p) and Gpr6p (Yali0c23298p) were lowly expressed, whereas Gpr4p (Yali0e12837p) and Gpr5p (Yali0f16225p) exhibited no expression, at all, under several tested conditions (Angermann 2009).

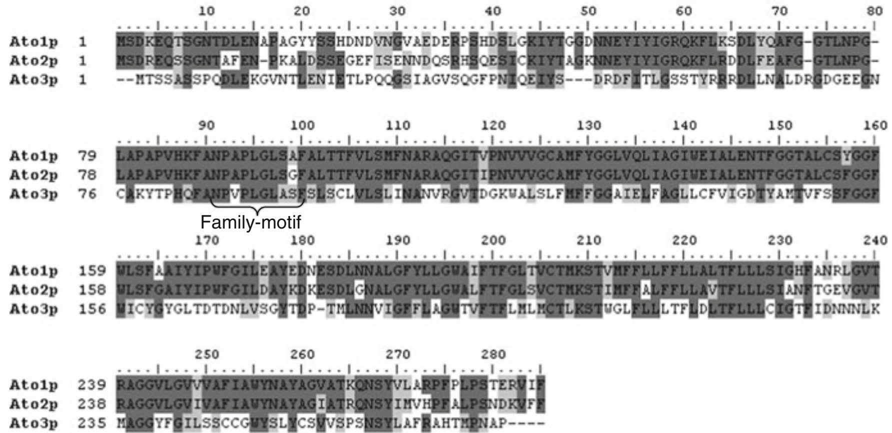


Fig. 4 Sequence alignment of Gpr1p orthologues in *S. cerevisiae*. *Dark grey* and *light grey* are the identical and similar amino acids, respectively. Alignment was performed using ClustalW (Thompson et al. 1997) and Bioedit (Hall 1999)

5 Gpr1p Orthologues in *Saccharomyces cerevisiae*

As already mentioned above, *S. cerevisiae* harbours two Gpr1p highly orthologous proteins: Ycr010cp (Ato1p, Ady2) and Ynr002cp (Ato2p, Fun34p). A further one, Ydr384cp (Ato3p), is also homologous to Gpr1p but has a slightly divergent family motif. The gene *YNR002c* could be an interchromosomal duplication of *YCR010c* (Lalo et al. 1993). Langkjaer et al. (2003) showed that this duplication was followed by asynchronous differentiation. These proteins are the only ones out of the GPR1/FUN34/YaaH family for which functions as transporter are suggested by different studies. Therefore, the results of these studies will be presented here in comparison with the data on the Gpr proteins of *Y. lipolytica*. As shown in Fig. 4, the homology between Ycr010cp and Ynr002cp is higher than to Ydr384cp. Differences occur mainly between the 20th and 40th amino acid.

5.1 *Localisation of the Three Gpr1p Orthologues*

The data on localisation of these proteins inside the cells of *S. cerevisiae* are contradictory. Although all three homologous proteins Ycr010cp, Ynr002cp and Ydr384cp were isolated from mitochondrial membrane fractions (Sickmann et al. 2003; Reinders et al. 2006, 2007), no mitochondrial localisation could be shown so far. Furthermore, *YNR002c* is induced more than twofold by overexpression of the *HAP4* respiratory transcription factor, what also could suggest a mitochondrial localisation (Lascaris et al. 2003).

All three proteins were localised to the plasma membrane as shown by GFP fusions and are associated with detergent-resistant membranes (Guaragnella and Butow 2003; Kuschel 2006; Ricicova et al. 2007) as it was also shown for Gpr1p in *Y. lipolytica* (see above). Interestingly, Ycr010cp forms patches above a pH of 6 (Ricicova et al. 2007) which disappear at pH 5 or lower.

5.2 Expression

Different *lacZ* reporter gene analyses were performed for the three homologues in *S. cerevisiae* (Augstein 2001; Gentsch 2005). All three homologues in *S. cerevisiae* are repressed by glucose and induced by glycerol, oleic acid, ethanol and carbon source depletion. The depletion of glucose in media with additional ethanol or acetate increased promoter activity of the genes *YCR010c* and *YNR002c*, but only weakly of *YDR384c* (Kuschel 2006).

In contrast to the *GPR1* gene of *Y. lipolytica*, *YCR010c*, *YNR002c* and *YDR384c* are only weakly induced by acetate and lactate. In contrast to this observation, Paiva et al. (2004) showed strong induction of *YCR010c* by shifts from glucose to acetic, lactic and pyruvic acid-containing medium. The homologous genes were independently regulated as proven by reporter gene analyses in the respective knockout strain (Dillschneider 2003).

As shown for *GPR1* of *Y. lipolytica*, *YCR010c*, *YNR002c* and *YDR384c* are induced by carbon depletion in *S. cerevisiae*, whereas no induction for *YCR010c* and low basal expression for *YNR002c* and *YDR384c* were shown for nitrogen, phosphorus and sulphur limitation (Boer et al. 2003). Northern blot analyses revealed high expression of the homologous genes in *S. cerevisiae* after transfer of cells into minimal medium with glucose, but already 30 min later, a strong decrease was observed. *YCR010c* showed no and *YNR002c* very low expression during the ongoing cultivation up to 12 h. In contrast, Lalo et al. (1994) showed no repression of *YNR002c* by glucose. On the contrary, *YDR384c* was well expressed up to the 8th hour and the expression declined after reaching the log phase. After 24 h of cultivation, the expression was completely different between the homologues (*YDR384c* < *YCR010c* < *YNR002c*) (Kuschel 2006). In contrast, Paiva et al. (2004) showed glucose catabolite repression of all three homologues.

Cat8p is essential for the consumption of non-fermentable carbon sources. With the help of DNA microarrays, the expression of Cat8p-dependent genes was investigated during diauxic growth on glucose (Haurie et al. 2001). The expression of *YCR010c* was diminished in *Δcat8* strain. *YNR002c* was the only one of the three genes that was higher expressed in that mutant. The expression of *YDR384c* was not effected by deletion of *CAT8*. Only the promoter of *YCR010c* contains a CSRE motif. Therefore, *YCR010c* could be positively regulated by Cat8p after glucose depletion and *YNR002c* negatively.

The influence of the pH value on the expression of the homologues is negligible in media with glucose, ethanol or without any carbon source. Lower pH values resulted in reduced expression of all three homologues with acetate as sole carbon source (Kuschel 2006).

As found per Northern blot analysis, *YCR010c* is induced by lactic acid, pyruvic acid and glycerol (Paiva et al. 2004). Furthermore, *YCR010c* and *YNR002c* are strongly induced, but *YDR384c* is repressed by ethanol. In contrast to *GPR1* in *Y. lipolytica*, all three homologues were only weakly induced by acetate. The expression of the homologous *YCR010c* and *YDR384c*, but not *YNR002c*, is induced by general stress (Kuschel 2006). The gene *YNR002c* is positively regulated by War1p (Schuller et al. 2004). This transcription factor binds to the consensus sequence 5′-CGGc/tTg/ct/gTAa/t-3′ and is necessary for the adaptation to weak organic acids like sorbic, benzoic and propionic acid, but not acetic acid (Kren et al. 2003; Schuller et al. 2004).

All three homologous proteins are subjected to regulation by Sok2p. The deletion of *SOK2* leads to lower expression of *YCR010c*, *YNR002c* and *YDR384c*. Furthermore, the *Δsok2* strain produces lower ammonium in the alkalised growth phase. The transcription factor Sok2p is involved in different cellular processes like growth, differentiation and adaptation to environmental changes, especially availability of nutrients (Vachova et al. 2004). These different regulatory mechanisms may be hints for different functions of the *GPR1* orthologues in *S. cerevisiae*.

Heterologous expression of Ycr010cp in *E. coli* leads to SOS response (Perkins et al. 1999). Furthermore, the null mutant of *YCR010c* only forms dyads (Rabitsch et al. 2001). Nevertheless, single, double or even the knockout of three homologues in *S. cerevisiae* have no phenotypic effect on minimal media with glucose, glycerol, ethanol and acetate as sole carbon source and in combination thereof (Kuschel 2006) like the deletion of *GRP1* in *Y. lipolytica*.

In ρ_0 cells the expression of *YDR384c* (*ATO3*) is dramatically increased (Guaragnella and Butow 2003). Furthermore, it was found that the transcriptional activator Gcn4p and the Ssy1-Ptr3-Ssy5 complex play a role in the expression of *YDR384c*. The authors supposed that *ATO3* is induced in ρ_0 cells for reducing ammonium excess because of potential defect in ammonium assimilation.

Like Gpr1p all three orthologues are subjects to phosphorylation. Ycr010cp is phosphorylated at Thr_{47}, which is equally located to Ser_{37} in Gpr1p, whereas Ynr002cp is phosphorylated at $Ser_{2,6,7,21,22,28,40}$ and Ydr348cp at $Ser_{3,4,6,7}$ (Reinders et al. 2007).

Heterologous expression of the *GPR1* mutant alleles under control of the *MET25* promoter was functional in *S. cerevisiae*, and acetic acid sensitivity could be observed also in the presence of glucose as in *Y. lipolytica* (Kuschel 2006). L74Q, L75Q and G259D amino acid substitution in Ycr010cp and Ynr002cp, respectively, that were expressed under control of the *MET25* promoter showed acetic acid sensitivity in the presence of glucose. Same mutant alleles showed only acetic acid sensitivity without glucose presence when expressed under their native promoters (Kuschel 2006). The glucose catabolite repression seems in general more strictly regulated in *S. cerevisiae* than in *Y. lipolytica*. Therefore, it is not surprising

that *GPR1* expression is only moderately repressed by glucose in contrast to *ADY2*. In Ycr010cp, mutagenesis of E35G, A70V, F71V, T74I, A88V, E144G, N145D, T209A, M211K, F212S and H230R also leads to acetic acid hypersensitivity. In Ynr002cp, L65P, N66D, A87T, G147D, S265L and M268I are further amino acids that promote acetic acid sensitivity if mutagenised. In Ydr384cp, no amino acid exchanges were found that induce acetic acid sensitivity (Kuschel 2006).

The YNAYA motif is also present in the C-termini of the orthologues Ycr010cp and Ynr002cp, but not in Ydr384cp. Like in Gpr1p this motif is functional important as demonstrated by C-terminal truncation of the proteins as well as by singular deletion of the YNAYA motif (Kuschel 2006).

Two-hybrid analysis revealed an interaction of Ydr384c with Ymr009wp (Ito et al. 2001). *YMR009w* codes for an acireductone dioxygenase (ARD). The enzyme takes up two oxygen atoms and is involved in the production of L-methionine (Hirano et al. 2005).

The knockout of *YCR010c* leads to a significant defect in the expression of some hexose transport-encoding genes (Paiva et al. 2004).

5.3 Postulated Functions

Interestingly, the hints for different functions of Gpr1p from *Y. lipolytica* and its orthologues in *S. cerevisiae* are manifold. They reach from DNA metabolism (Perkins et al. 1999) and cell development (Rabitsch et al. 2001) to nitrogen metabolism, especially amino acid synthesis (Ito et al. 2001; Hirano et al. 2005), nutrient sensor (Vachova et al. 2004) and ammonium exporter (Palkova et al. 2002; Ricicova et al. 2007) (Gori et al. 2007); to carbon metabolism, especially acetate uptake (Paiva et al. 2004), and organic acid adaptation (Kren et al. 2003; Schuller et al. 2004); and last but not least to acetic acid adaptation (Gentsch et al. 2007).

The most prominently discussed functions of the homologous proteins are the ammonium transport, the acetate uptake and the adaptation to acetic acid.

Palkova et al. (2002) found out that colonies of *S. cerevisiae* change the pH value of the surrounding solid medium from acidic to alkali and vice versa. With the help of cytogenetic and genetic investigations as well as microarray analysis, the transition from acidic to alkali was divided into six phases. This transition is associated with the production of volatile non-protonated ammoniac, which represents a signal for starvation between cells of a colony. The bases of the ammoniac production are amino acids (Zikanova et al. 2002). The first ammoniac pulse is generated after inoculation and proceeds undirected, followed by a fast acidification of the medium. The second pulse seems to be directed and influences the neighboured cells. The induced ammoniac production synchronises the acidic/alkali pulse in neighboured cells independently of the specific stage of development (Palkova et al. 1997; Palkova and Forstova 2000). Ammoniac signalises nutrient deficiency to neighbouring cells and leads to temporary interruption of growth. This also prevents the unification of neighbouring colonies. Growth only takes place in the

direction where no colonies are located (Palkova et al. 1997). After the alkali phase, the ammoniac production is reduced and the colonies can proceed with growth (Palkova et al. 2002).

During the initial phase of the second ammoniac pulse, a strong induction of the *GPR1* orthologous genes *YCR010c*, *YDR384c* and *YNR002c* in *S. cerevisiae* takes place. In this phase strains disrupted in single *GPR1* orthologous genes produce a lower amount of ammoniac compared to the wild-type strain BY4742. Gori et al. (2007) showed different ammonium secretions in *Debaryomyces hansenii*, *S. cerevisiae*, *Geotrichum candidum* and *Y. lipolytica*. In contrast to the assumption of general signalling phenomenon in yeasts, *Y. lipolytica* (CBS 2075) and *G. candidum* (CBS 615.84) showed no ammonium secretion on glycerol-containing agar. They propose that ammonia is not involved in signalling in these yeast species.

Palkova et al. (2002) assume that the proteins Ycr010cp, Ydr384cp and Ynr002cp function as ammonium–proton antiporter. Therefore, they were named *ATO1-3* (*a*mmonia *t*ransport *o*utward). The homology of the three homologues to a potential ammonium transporter from *C. elegans* (Palkova et al. 2002) is very weak, and ammonium transporter signature (ATS; DXAGXXXVHLXGGXXXX-XAXXXXXPR) is different from the consensus sequence of the homologues and the ammonium transporter (XFXGGXXXLXXGXXXXXXXXXXXXX) (Kuschel 2006). Dillschneider (2003), in contrast to Palkova et al., assumes that Ato1p, Ato2p and Ato3p do not work as ammonium transporter but stimulate a hypothetical ammonium transporter in *S. cerevisiae*. This hypothesis is based on the observation that single mutants showed reduced ammonium secretion ($\Delta ycr010c > \Delta ydr384c > \Delta ynr002c$), whereas double mutants showed only a small difference to single mutants and triple mutants no difference to double mutants. However, triple mutants were still able to secrete ammonium and the ammonium secretion is not influenced by the acetic acid sensitivity mediating mutation in the homologous proteins in *S. cerevisiae* (Kuschel 2006).

The assumption that *GPR1* from *Y. lipolytica* and *YCR010c* from *S. cerevisiae* might have different function because of different glucose repression (Paiva et al. 2004) can be refused, because mutations at acetic acid sensitivity mediating amino acid sequences are functional in Gpr1p from *Y. lipolytica* as well as Ycr010cp from *S. cerevisiae* (Kuschel 2006). Random mutagenesis of Ycr010cp and Ynr002cp revealed further sequence positions that mediate acetic acid sensitivity, whereas no mutants could be isolated for Ydr384cp (Kuschel 2006).

Newer results give reason to allocate YCR010cp an essential function in the acetate uptake in *S. cerevisiae* (Paiva et al. 2004). *JEN1* from *S. cerevisiae* encodes for a lactate–pyruvate–propionate transporter that is induced in lactic or pyruvic acid-grown cells (Casal et al. 1999). Because the deletion of *JEN1* still permits growth on acetate, at least one further acetate transporter has to exist (Casal et al. 1996; Makuc et al. 2001). Although microarray analyses showed that all three *GPR1* orthologous genes are significantly activated during the transition from glucose to acetate-containing medium, the deletion of only *YCR010c* strongly

reduced the acetate uptake. Because the protein showed no influence on the transcriptional regulation in the presence of acetate, Ycr010cp was postulated to be directly involved in the acetate transport (Paiva et al. 2004).

Knockout of *YCR010c* had strong effect on acetate uptake in *S. cerevisiae*, but *ΔYNR002c* and *ΔYDR384c* did not. Furthermore, no difference in acetate adaptation could be found for wild type and *ΔYCR010c* at pH 6 (Paiva et al. 2004). Nevertheless, acetic acid sensitivity was observed at pH below 6 and mutated *YCR010c* (Gentsch et al. 2007).

6 Discussion

As known from *K. lactis*, high-sequence homology might presume *iso*-function of the genes, e.g. *JEN1* encodes a lactate–pyruvate transporter, whereas *JEN2* encodes a succinate–malate–acetate transporter. A similar functional relationship could be observed for Gpr1p and its homologues. Also here different functions could be assumed, because intact homologues could not complement acetic acid sensitivity when a *GPR1*^{d} allele is present.

On the basis of the results concerning acetic acid sensitivity in *Y. lipolytica* and *S. cerevisiae*, a model was proposed (Augstein 2001; Gentsch 2005), suggesting Gpr1p and its orthologues are active in the absence of acetic acid and inhibit acetic acid adaptation processes. At low pH values acetic acid is mainly present in a non-dissociated protonated form. In this case acetic acid can enter the cell by simple diffusion (Casal et al. 1996). At low pH values also the permeability of the cell membrane for protons increases (Foster 2000). Inside the cell dissociation of acetic acid occurs to acetate anion and proton because of higher internal pH value. This interferes with different metabolic processes (Krebs et al. 1983). It is supposed that the intracellular acetic acid concentration causes the Gpr1p deactivation via recognition through the FGGTLN-motif and thus the inhibition of a hypothetical acetate exporter and a proton pump is abolished (Gentsch 2005). Pma1p (plasma membrane ATPase) is activated by low intracellular pH value (Carmelo et al. 1997), and so Pma1p-orthologues in *Y. lipolytica* and *S. cerevisiae* could be activated. Phosphorylation of the proteins is performed by a hypothetical acetate-activated kinase. It is supposed that mutated Gpr1p cannot be deactivated and/or that acetate is not recognised by the mutated Gpr1p. The lack of an activation of the proton pump was shown for the *GPR1*^{d}-mutant B204-12C-156 by patch clamp studies (Shapoval, unpublished).

That a single nucleotide polymorphism in a gene, which is not essential and even has no obvious null mutant phenotype, leads to such a dramatic mutant phenotype is impressing. Although many investigations were performed in the last years, the final understanding of this process remains questionable.

References

Angermann S (2009) Charakterisierung der *GPR1*-Homologen in der Hefe *Yarrowia lipolytica* und Deletion des *GPR1*-Homologen *YALI0E27291g* (*GPR2*). Diploma thesis, Institute of Microbiology, Dresden University of Technology, Dresden

Augstein A (2001) Molekularbiologische Charakterisierung und funktionelle Analyse des GPR1-Genproduktes in der Hefe *Yarrowia lipolytica*. PhD thesis, Institute of Microbiology, Dresden University of Technology, Dresden

Augstein A, Barth K et al (2003) Characterization, localization and functional analysis of Gpr1p, a protein affecting sensitivity to acetic acid in the yeast Yarrowia lipolytica. Microbiology 149 (Pt 3):589–600

Barth G, Scheuber T (1993) Cloning of the isocitrate lyase gene (ICL1) from Yarrowia lipolytica and characterization of the deduced protein. Mol Gen Genet 241(3–4):422–430

Barth G, Weber H (1985) Improvement of sporulation in the yeast Yarrowia lipolytica. Antonie Van Leeuwenhoek 51(2):167–177

Barth G, Weber H (1987) Genetic analysis of the gene ICL1 of the yeast Yarrowia lipolytica. Yeast 3(4):255–262

Bendtsen JD, Jensen LJ et al (2004a) Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng Des Sel 17(4):349–356

Bendtsen JD, Nielsen H et al (2004b) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340(4):783–795

Blattner FR, Plunkett G 3rd et al (1997) The complete genome sequence of Escherichia coli K-12. Science 277(5331):1453–1474

Boer VM, de Winde JH et al (2003) The genome-wide transcriptional responses of Saccharomyces cerevisiae grown on glucose in aerobic chemostat cultures limited for carbon, nitrogen, phosphorus, or sulfur. J Biol Chem 278(5):3265–3274

Brandt P, Ramlow S et al (1996) Nucleotide sequence analysis of a 32,500 bp region of the right arm of Saccharomyces cerevisiae chromosome IV. Yeast 12(1):85–90

Buchweitz J (2007) Funktionelle Analyse des Membranproteins Gpr1p in der Hefe *Yarrowia lipolytica*. Diploma thesis, Institute of Microbiology, Dresden University of Technology, Dresden

Carmelo V, Santos H et al (1997) Effect of extracellular acidification on the activity of plasma membrane ATPase and on the cytosolic and vacuolar pH of Saccharomyces cerevisiae. Biochim Biophys Acta 1325(1):63–70

Casal M, Cardoso H et al (1996) Mechanisms regulating the transport of acetic acid in Saccharomyces cerevisiae. Microbiology 142(Pt 6):1385–1390

Casal M, Paiva S et al (1999) The lactate-proton symport of Saccharomyces cerevisiae is encoded by JEN1. J Bacteriol 181(8):2620–2623

Daley DO, Rapp M et al (2005) Global topology analysis of the Escherichia coli inner membrane proteome. Science 308(5726):1321–1323

De Hertogh B, Carvajal E et al (2002) Phylogenetic classification of transporters and other membrane proteins from Saccharomyces cerevisiae. Funct Integr Genomics 2(4–5):154–170

De Hertogh B, Hancy F et al (2006) Emergence of species-specific transporters during evolution of the hemiascomycete phylum. Genetics 172(2):771–781

Dillschneider A (2003) Expressionsstudien der dem Gpr1 Protein aus *Yarrowia lipolytica* homologen Proteine in der Hefe *Saccharomyces cerevisiae*. Diploma thesis, Institute of Microbiology, Dresden University of Technology, Dresden

Emanuelsson O, Nielsen H et al (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300(4):1005–1016

Foster JW (2000) Microbial responses to acid stress. ASM, Washington, DC

Fuchs J (2010) Isolation des Gpr1p-enthaltenen Komplexes aus der Hefe *Yarrowia lipolytica* bzw. des Ycr010cp-enthaltenen Komplexes aus *Saccharomyces cerevisiae* zur Charakterisierung der Proteininteraktionspartner. Institute of Microbiology, Dresden University of Technology, Dresden

Gentsch M (2005) Funktionelle analyse und Charakterisierung des Gpr1-Proteins in der Hefe *Yarrowia lipolytica*. PhD thesis, Institute of Microbiology, Dresden University of Technology, Dresden

Gentsch M, Barth G (2005) Carbon source dependent phosphorylation of the Gpr1 protein in the yeast Yarrowia lipolytica. FEMS Yeast Res 5(10):909–917

Gentsch M, Kuschel M et al (2007) Mutations at different sites in members of the Gpr1/Fun34/YaaH protein family cause hypersensitivity to acetic acid in Saccharomyces cerevisiae as well as in Yarrowia lipolytica. FEMS Yeast Res 7(3):380–390

Goffeau A, Barrell BG et al (1996) Life with 6000 genes. Science 274(5287):546, 563–567

Gori K, Mortensen HD et al (2007) Ammonia production and its possible role as a mediator of communication for Debaryomyces hansenii and other cheese-relevant yeast species. J Dairy Sci 90(11):5032–5041

Guaragnella N, Butow RA (2003) ATO3 encoding a putative outward ammonium transporter is an RTG-independent retrograde responsive gene regulated by GCN4 and the Ssy1-Ptr3-Ssy5 amino acid sensor system. J Biol Chem 278(46):45882–45887

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Haurie V, Perrot M et al (2001) The transcriptional activator Cat8p provides a major contribution to the reprogramming of carbon metabolism during the diauxic shift in Saccharomyces cerevisiae. J Biol Chem 276(1):76–85

Hayashi K, Morooka N et al (2006) Highly accurate genome sequences of Escherichia coli K-12 strains MG1655 and W3110. Mol Syst Biol 2:2006.0007

Hirano W, Gotoh I et al (2005) Membrane-type 1 matrix metalloproteinase cytoplasmic tail binding protein-1 (MTCBP-1) acts as an eukaryotic aci-reductone dioxygenase (ARD) in the methionine salvage pathway. Genes Cells 10(6):565–574

Ito T, Chiba T et al (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci USA 98(8):4569–4574

Jacq C, Alt-Morbe J et al (1997) The nucleotide sequence of Saccharomyces cerevisiae chromosome IV. Nature 387(6632 Suppl):75–78

James R, Dean DO et al (1993) Five open reading frames upstream of the dnaK gene of E. coli. DNA Seq 3(5):327–332

Krebs HA, Wiggins D et al (1983) Studies on the mechanism of the antifungal action of benzoate. Biochem J 214(3):657–663

Kren A, Mamnun YM et al (2003) War1p, a novel transcription factor controlling weak acid stress response in yeast. Mol Cell Biol 23(5):1775–1785

Kujau M, Weber H et al (1992) Characterization of mutants of the yeast Yarrowia lipolytica defective in acetyl-coenzyme A synthetase. Yeast 8(3):193–203

Kuschel M (2006) Funktionelle analyse von Proteinen der Gpr1/Fun34/yaaH Proteinfamilie in den Hefen *Yarrowia lipolytica* und *Saccharomyces cerevisiae*. PhD thesis, Institute of Microbiology, Dresden University of Technology, Dresden

Lalo D, Stettler S et al (1993) Two yeast chromosomes are related by a fossil duplication of their centromeric regions. C R Acad Sci III 316(4):367–373

Lalo D, Stettler S et al (1994) Organization of the centromeric region of chromosome XIV in Saccharomyces cerevisiae. Yeast 10(4):523–533

Langkjaer RB, Cliften PF et al (2003) Yeast genome duplication was followed by asynchronous differentiation of duplicated genes. Nature 421(6925):848–852

Lascaris R, Bussemaker HJ et al (2003) Hap4p overexpression in glucose-grown Saccharomyces cerevisiae induces cells to enter a novel metabolic state. Genome Biol 4(1):R3

Makuc J, Paiva S et al (2001) The putative monocarboxylate permeases of the yeast Saccharomyces cerevisiae do not transport monocarboxylic acids across the plasma membrane. Yeast 18 (12):1131–1143

Matsuoka M, Ueda Y et al (1980) Role and control of isocitrate lyase in Candida lipolytica. J Bacteriol 144(2):692–697

Ogrydziak D, Bassel J et al (1982) Development of the genetic map of the yeast Saccharomycopsis lipolytica. Mol Gen Genet 188(2):179–183

Page RD (1996) TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12(4):357–358

Paiva S, Devaux F et al (2004) Ady2p is essential for the acetate permease activity in the yeast Saccharomyces cerevisiae. Yeast 21(3):201–210

Palkova Z, Forstova J (2000) Yeast colonies synchronise their growth and development. J Cell Sci 113(Pt 11):1923–1928

Palkova Z, Janderova B et al (1997) Ammonia mediates communication between yeast colonies. Nature 390(6659):532–536

Palkova Z, Devaux F et al (2002) Ammonia pulses and metabolic oscillations guide yeast colony development. Mol Biol Cell 13(11):3901–3914

Perkins EL, Sterling JF et al (1999) Yeast and human genes that affect the Escherichia coli SOS response. Proc Natl Acad Sci USA 96(5):2204–2209

Rabitsch KP, Toth A et al (2001) A screen for genes required for meiosis and spore formation based on whole-genome expression. Curr Biol 11(13):1001–1009

Reinders J, Zahedi RP et al (2006) Toward the complete yeast mitochondrial proteome: multidimensional separation techniques for mitochondrial proteomics. J Proteome Res 5(7):1543–1554

Reinders J, Wagner K et al (2007) Profiling phosphoproteins of yeast mitochondria reveals a role of phosphorylation in assembly of the ATP synthase. Mol Cell Proteomics 6(11):1896–1906

Ricicova M, Kucerova H et al (2007) Association of putative ammonium exporters Ato with detergent-resistant compartments of plasma membrane during yeast colony development: pH affects Ato1p localisation in patches. Biochim Biophys Acta 1768(5):1170–1178

Schlegel S (2005) Identifizierung von Interaktionspartnern des Gpr1-Proteins in der Hefe *Yarrowia lipolytica*. Institute of Microbiology, Dresden University of Technology, Dresden

Schuller C, Mamnun YM et al (2004) Global phenotypic analysis and transcriptional profiling defines the weak acid stress response regulon in Saccharomyces cerevisiae. Mol Biol Cell 15 (2):706–720

Sickmann A, Reinders J et al (2003) The proteome of Saccharomyces cerevisiae mitochondria. Proc Natl Acad Sci USA 100(23):13207–13212

Skala J, Purnelle B et al (1992) The complete sequence of a 10.8 kb segment distal of SUF2 on the right arm of chromosome III from Saccharomyces cerevisiae reveals seven open reading frames including the RVS161, ADP1 and PGK genes. Yeast 8(5):409–417

Thompson JD, Gibson TJ et al (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25 (24):4876–4882

Tzschoppe K (1998) Lokalisierung und Untersuchungen zur Funktionsaufklärung des Gpr1p ("Glyoxylate pathway regulator") Proteins in der Hefe *Yarrowia lipolytica* unter Anwendung des "grün fluoreszierenden Proteins" (Gfp). Institute of Microbiology, Dresden University of Technology, Dresden

Tzschoppe K, Augstein A et al (1999) trans-dominant mutations in the GPR1 gene cause high sensitivity to acetic acid and ethanol in the yeast Yarrowia lipolytica. Yeast 15(15):1645–1656

Vachova L, Devaux F et al (2004) Sok2p transcription factor is involved in adaptive program relevant for long term survival of Saccharomyces cerevisiae colonies. J Biol Chem 279 (36):37973–37981

Werner K (2009) Studies on secretion of Gpr1 protein in *Yarrowia lipolytica* and of the Gpr1p orthologues in *Saccharomyces cerevisiae*. Dresden University of Technology, Institute of Microbiology, Dresden

Yoshikawa A, Isono K (1990) Chromosome III of Saccharomyces cerevisiae: an ordered clone bank, a detailed restriction map and analysis of transcripts suggest the presence of 160 genes. Yeast 6(5):383–401

Yura T, Mori H et al (1992) Systematic sequencing of the Escherichia coli genome: analysis of the 0–2.4 min region. Nucleic Acids Res 20(13):3305–3308

Zikanova B, Kuthan M et al (2002) Amino acids control ammonia pulses in yeast colonies. Biochem Biophys Res Commun 294(5):962–967

Sulphur Metabolism of the Cheese-Ripening Yeast *Yarrowia lipolytica*

Agnès Hébert, Jean Marie Beckerich, Sophie Landaud, and Pascal Bonnarme

Abstract Due to their importance in flavouring capacities (e.g. production of volatile sulphur compounds (VSCs)), their role in redox regulation (e.g. production of thiols) and their possible role in microbial interactions, sulphur compounds play a key role in traditional fermented products. A better knowledge and control of their microbial production is therefore of major interest not only to maintain the high quality of these foodstuffs but also to get generic knowledge on sulphur metabolism in the microbial world. *Yarrowia lipolytica* is a commonly found yeast in numerous cheeses, and its enzymatic feature gives this microorganism a competitive superiority over other yeast species found in cheese. Several studies have shown that this yeast could efficiently degrade sulphur amino acids, leading to VSC production.

The main knowledge on *Y. lipolytica* sulphur metabolism is presented in this chapter. A metabolic reconstruction of *Y. lipolytica* sulphur metabolism is first presented. The common occurrence of *Y. lipolytica* in cheeses is demonstrated together with the importance of VSCs in cheese flavouring. Then, L-methionine and L-cysteine catabolism's state of the art is presented. Likely *Y. lipolytica* L-methionine and L-cysteine degradation pathways are also presented.

A. Hébert
AgroParisTech-INRA, UMR782 Génie et Microbiologie des Procédés Alimentaires, Thiverval-Grignon, France

AgroParisTech-INRA, UMR1319 Micalis, 78850 Thiverval Grignon, France

J.M. Beckerich
AgroParisTech-INRA, UMR1319 Micalis, 78850 Thiverval Grignon, France

S. Landaud • P. Bonnarme (✉)
AgroParisTech-INRA, UMR782 Génie et Microbiologie des Procédés Alimentaires, Thiverval-Grignon, France
e-mail: bonnarme@grignon.inra.fr

G. Barth (ed.), *Yarrowia lipolytica*, Microbiology Monographs 24,
DOI 10.1007/978-3-642-38320-5_8, © Springer-Verlag Berlin Heidelberg 2013

Contents

1 Introduction 166
2 Metabolic Reconstruction of Sulphur Metabolism in *Yarrowia lipolytica* 167
3 *Yarrowia lipolytica* in Cheese Ecosystems 170
4 Importance of Volatile Sulphur Compounds on the Quality of Cheeses 171
5 L-Methionine and L-Cysteine Catabolisms and Volatile Sulphur Compounds Biosynthesis in *Y. lipolytica* 174
5.1 L-Cysteine Catabolism 174
5.2 Evidence of Transamination as a Key Step for L-Methionine Degradation to Volatile Sulphur Compounds 175
5.3 In Silico Analyses of Genes Encoding Branched-Chain Amino Acid Aminotransferases or Aromatic Acid Aminotransferases 176
5.4 *Y. lipolytica* Transcriptome Analysis in Response to Various Sulphur Sources 179
6 Future Perspectives 181
References 182

1 Introduction

Yarrowia lipolytica is one of the most extensively studied non-conventional yeasts for it is quite different from other intensively studied yeasts like *Saccharomyces cerevisiae*. Its ecological niche encompasses lipid-rich food like margarine, olive oil and cheese. *Y. lipolytica* is non-pathogenic to human and has been approved for several industrial processes. It is the only species recognized within the *Yarrowia* genus (Barnett et al. 2000). It is an obligate aerobe which secretes large amounts of various metabolites and enzymes. On rich media, it naturally secretes an alkaline protease at neutral pH or an acidic protease at low pH. Secreted RNase, phosphatase, lipase and esterase were also detected under diverse growth conditions (Barth and Gaillardin 1997). These characteristic features give this microorganism a competitive superiority over other yeast species found in various habitats including cheese (see Sect. 3). As a consequence, *Y. lipolytica* which develops on the cheese surface constitutes an adventitious microorganism from the cheese environment (e.g. brine, ripening shelves and personnel), and they rapidly outnumber the commercial cultures (Mounier et al. 2006). *Y. lipolytica* is a ubiquitous, naturally developing yeast in many cheeses (Corsetti et al. 2001; Cosentino et al. 2001) and constitutes a part of the cheese surface microflora. Its adventitious nature and its enzymatic activities make it a good candidate for ripening. Due to its efficient productivity, this yeast has been used for the preparation of cheese flavour compounds. Among flavour compounds, volatile sulphur compounds (VSCs) are estimated to represent about 10 % of the volatile components detected in food and beverages (Boelans and van Gemert 1993). Such compounds are commonly found in dairy products, especially ripened cheeses (Berger et al. 1999). Their low odour thresholds make them important contributors to the odour and aroma of cheeses and may participate to the organoleptic properties of cheeses (Boelans and van Gemert 1993). However, the metabolism leading to the synthesis of VSCs and related sulphur metabolism have been, until recently, the subject of limited investigation.

In cheese, VSCs essentially arise from L-methionine and L-cysteine—a 1:5 L-cysteine/L-methionine proportion was found in casein (Wood et al. 1985)—which are the major sulphur-bearing precursors found in casein.

L-Methionine being the main sulphur amino acid found in cheese curd, its catabolism has been extensively investigated in several cheese-ripening yeasts and bacteria (Bonnarme et al. 2000, 2001) with respect to VSC production. It is well established that in the cheese ecosystem, VSCs primarily arise from the degradation of L-methionine to methanethiol (MTL), the latter being subsequently converted to other sulphur-bearing compounds, including the MTL oxidation products dimethyl disulphide (DMDS) and dimethyl trisulphide (DMTS) and other VSCs such as thioesters and thioethers (Landaud et al. 2008). In yeasts, L-methionine to MTL conversion proceeds via a two-step degradation pathway which is initiated by an aminotransferase, leading to the formation of the transamination product 4-methylthio-2-oxobutyric acid (KMBA) which is subsequently converted to MTL. Tracing experiments enabled to identify volatile and non-volatile L-methionine degradation intermediates and end products (Arfi et al. 2006; Bonnarme et al. 2004). In contrast, owing to its low levels in cheese, the study of L-cysteine catabolism has been rather neglected until now in cheese-ripening microorganisms.

It has been established that *Y. lipolytica* produces many more VSCs than other commonly found cheese-ripening yeasts, such as *Debaryomyces hansenii* and *Kluyveromyces lactis* (Cernat-Bondar et al. 2005; Cholet et al. 2007). L-Methionine and *S*-methyl-methionine degradation have been studied in several cheese-ripening yeasts, including *Y. lipolytica*, *Geotrichum candidum*, *K. lactis* and *D. hansenii* (Spinnler et al. 2001). It was found that *Y. lipolytica* was an efficient VSC producer leading to (1) dimethyl sulphide (DMS) accumulation through *S*-methyl-methionine degradation and (2) MTL, DMDS, DMTS and methylthioacetate (MTA) through L-methionine degradation.

The recent publications of the genomes of several yeasts involved in cheese ripening, among others *Y. lipolytica* (Dujon et al. 2004), opened up new opportunities to investigate the VSC production abilities of this yeast of great technological interest at the molecular level.

2 Metabolic Reconstruction of Sulphur Metabolism in *Yarrowia lipolytica*

From in silico analysis of the sequenced *Y. lipolytica* genome, a metabolic reconstruction of the sulphur metabolism of this yeast has been done and is presented below (Table 1; Fig. 1). In order to efficiently assimilate sulphate from the extracellular medium, *Yarrowia lipolytica* possesses a high-affinity sulphate transporter (*YALI0B17930g*). Once in the cell, sulphate requires several steps before it can be utilized. In fact, sulphate is too stable to be directly incorporated into a carbon chain. Sulphate is primarily activated and then reduced (Fig. 1, black arrows),

Table 1 Genes involved in sulphur metabolism in *Yarrowia lipolytica*. The genes were identified by an in silico comparison with the yeast model *Saccharomyces cerevisiae*. The databases utilized are the *Saccharomyces* Genome Database (http://www.yeastgenome.org/) and Genolevures (http://www.genolevures.org/)

Sulphate transporter	
YALI0B17930g	High-affinity sulphate permease
Sulphate assimilation	
YALI0B08184g	ATP sulphurylase
YALI0E00418g	Adenosine-5-phosphosulphate kinase
YALI0B08140g	3′-Phosphoadenylsulphate reductase
YALI0D11176g	Sulphite reductase beta subunit
YALI0E16368g	Sulphite reductase alpha subunit
YALI0E00836g	L-Homoserine-*O*-acetyltransferase
YALI0D25168g	*O*-Acetyl homoserine-*O*-acetyl serine sulphhydrylase
Methyl cycle	
YALI0E12683g	Cobalamin-independent methionine synthase
YALI0B14509g	*S*-Adenosylmethionine synthetase
YALI0F11759g	*S*-Adenosyl-L-homocysteine hydrolase
Transsulphuration pathway	
YALI0E09108g	Cystathionine beta-synthase
YALI0F05874g	Cystathionine gamma-lyase
YALI0D17402g	Cystathionine gamma-synthase
YALI0D00605g	Cystathionine beta-lyase
Glutathione synthesis	
YALI0E30129g	Gamma glutamylcysteine synthetase
YALI0C17831g	Glutathione synthetase

leading to the production of sulphide. The sulphate activation is realized in two successive steps. First, the ATP sulphurylase (*YALI0B08184g*) catalyses the formation of adenylyl sulphate (APS) from ATP and sulphate. Secondly, the APS kinase (*YALI0E00418g*) realizes the phosphorylation of APS, producing PAPS. The sulphate activation is followed by the reduction of PAPS. This reduction requires the successive action of two enzymes, the PAPS reductase (*YALI0B08140g*) which produces sulphite and the sulphite reductase composed of two subunits (*YALI0E16368g* (alpha unit); *YALI0D11176g* (beta unit)), which finally leads to the production of sulphide. The sulphide produced can be incorporated into a carbon chain, the *O*-acetylhomoserine, to form homocysteine.

In yeasts, homocysteine plays a central role in sulphur metabolism. This molecule is at the crossroad of the transsulphuration pathway (grey arrows) and the methyl cycle (black arrows intermittent line), which lead respectively to the production of L-cysteine and L-methionine. The reverse transsulphuration pathway renders possible the conversion of homocysteine to L-cysteine via cystathionine. The enzymes involved in L-cysteine synthesis from homocysteine are the cystathionine β-synthase (*YALI0E09108g*) and the cystathionine γ-lyase (*YALI0F05874g*).

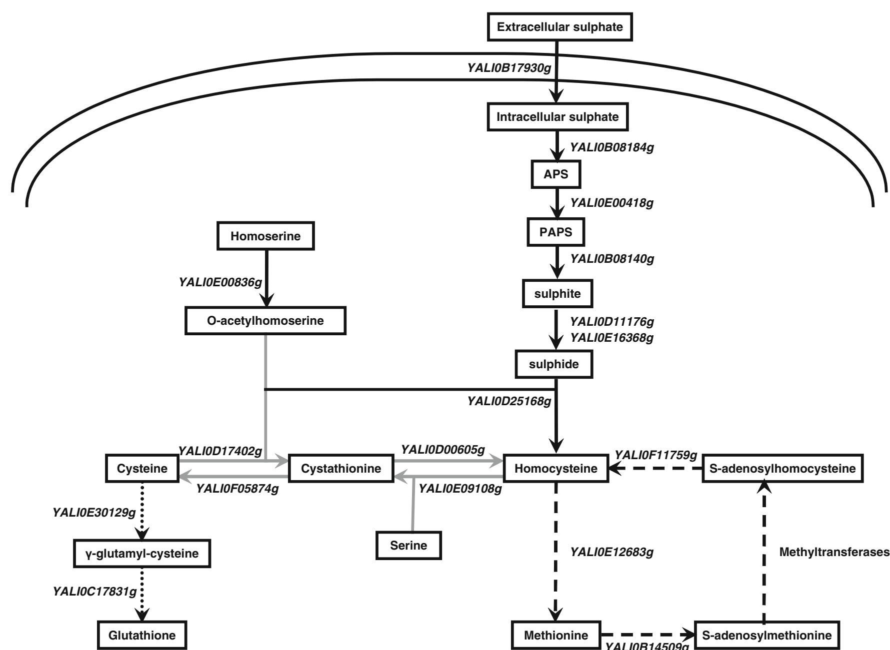


Fig. 1 Survey of sulphur metabolism in *Yarrowia lipolytica*. Sulphate assimilation: *black arrows*. Transsulphuration pathway: *grey arrows*. Methyl cycle: *black arrows intermittent line*. Glutathione synthesis: *black arrows dotted line*

The synthesis of homocysteine from L-cysteine, called transsulphuration pathway, involves the cystathionine γ-synthase (*YALI0D17402g*) and the cystathionine β-lyase (*YALI0D00605g*). Once produced, L-cysteine may lead to the production of glutathione (black arrows dotted line), via γ-glutamylcysteine. The two enzymes responsible for glutathione biosynthesis are the γ-glutamylcysteine synthase (*YALI0E30129g*) and the glutathione synthase (*YALI0C17831g*). Glutathione is an essential molecule involved in defence against oxidative stress.

Methionine is directly synthesized from homocysteine by the methionine synthase (*YALI0E12683g*). The methyl cycle leads to the production of an important molecule, *S*-adenosylmethionine (SAM) by the *S*-adenosylmethionine synthase (*YALI0B14509g*). SAM is the major donor of methyl in the cell and is involved in several metabolisms. Methyltransferase converts SAM to *S*-adenosylhomocysteine, this latter molecule being the substrate of *S*-adenosylhomocysteine hydrolase (*YALI0F11759g*), which leads to de novo homocysteine synthesis.

Interestingly, in silico analysis revealed a striking redundancy of putative L-methionine permease *MUP1* and/or *MUP3* genes in *Y. lipolytica* genome (*YALI0D16137g*, *YALI0F03498g*, *YALI0F25795g*, *YALI0D19646g* and *YALI0F07018g*), higher than in other cheese-ripening yeast genomes, like *Saccharomyces cerevisiae*, *Kluyveromyces lactis* and *Debaryomyces hansenii*. The same feature is also observed with GAP1, a general amino acid permease, for which seven orthologues *YALI0B16522g*,

YALI0B19492g, *YALI0B19800g*, *YALI0C09889g*, *YALI0C17237g*, *YALI0E10219g* and *YALI0F19866g* were found in *Y. lipolytica*.

This suggests that such amino acid transporters may give a competitive advantage to *Y. lipolytica*, allowing it to rapidly grow on L-methionine. Interestingly, MUP1, which is a high-affinity permease, could also be involved in cysteine uptake.

3 *Yarrowia lipolytica* in Cheese Ecosystems

Y. lipolytica occurs frequently in milk products, and owing to its high catabolic activities (e.g. proteolytic and lipolytic activities), this species can play an important role in the formation of aroma precursors such as amino acids, fatty acids and esters, as well as their subsequent conversion to aroma-active or other bioactive compounds (Guerzoni et al. 2001; van den Tempel and Jakobsen 2000). Strains of the yeasts *D. hansenii* and *Y. lipolytica*, isolated from blue mould cheeses, were examined for their technological characteristics and potential use as starter cultures in cheese ripening, while *Y. lipolytica* was more sensitive to NaCl and did not assimilate lactose, contrary to *D. hansenii*. Furthermore, *Y. lipolytica* strains were much more lipolytic and proteolytic than *D. hansenii* (van den Tempel and Jakobsen 2000).

Due to its catabolic activities, *Y. lipolytica* can therefore strongly influence the organoleptic properties of cheese through the production of precursors from the cheese curd, leading to aroma compound biosynthesis during the ripening process. For instance, chemico-physical parameters of major importance for cheese ripening (e.g. pH, NaCl, milk fat) markedly influenced the lipolytic activity of *Y. lipolytica* strains with substantial changes in the free fatty acid (FFA) production profiles. Such FFAs are known as potential aroma precursors (Guerzoni et al. 2001).

Y. lipolytica has been identified in many cheeses. For instance, in Camembert and Brie cheeses, although a broad spectrum of yeasts was isolated from both cheeses, *Y. lipolytica* and *D. hansenii* were the most abundant yeast species isolated (Viljoen et al. 2003). In the red-smear cheese, Livarot cheese, the microbial diversity of the cheese surface ecosystem was studied using culture-dependent and culture-independent approaches. It was found that *Y. lipolytica* accounted for 7.5 % of the total culturable yeast microbiota (Mounier et al. 2009). Microbial interactions occurring within a cheese microbial community has been investigated using the Lotka–Volterra model and yeasts omissions studies to evaluate species interactions (Mounier et al. 2008). It was shown that negative interactions occurred between yeasts. Although mechanisms involved in such interactions still remained to be elucidated, the authors found that *Y. lipolytica* inhibited mycelial expansion of the yeast *G. candidum*, a yeast of great importance in numerous soft cheeses such as Camembert or Livarot cheeses (Mounier et al. 2008).

The growth of *Y. lipolytica* can also be dramatically affected by the presence of other microorganisms of the ecosystem, especially the bacterium *Staphylococcus xylosus* (Mansour et al. 2009b). The presence of *S. xylosus* C2a resulted in a

100-fold decrease in the *Y. lipolytica* cell count compared to the pure culture. It was postulated that competition for amino acids between *Y. lipolytica* and *S. xylosus* may explain this phenomenon, since the amino acids were dramatically consumed in *Y. lipolytica*–*S. xylosus* coculture compared to *Y. lipolytica* or *S. xylosus* monoculture. As a result of the low amino acid concentration in the medium, the expression of *Y. lipolytica* genes involved in amino acid catabolism (*GDH2*, *BAT1*, *KAD*) was downregulated in the presence of *S. xylosus* compared to the yeast monoculture.

Gene expression and biochemical analyses were performed in a coculture of the yeasts *D. hansenii*, *Kluyveromyces marxianus* and *Y. lipolytica* (Cholet et al. 2007). The time-course expression of target genes possibly involved in lactose/lactate catabolism and the biosynthesis of sulphur-flavoured compounds were studied in pure cultures of each yeast, as well as in coculture, and compared to biochemical data. A high expression of the *LAC* genes was observed in *K. marxianus*, a yeast which degrades lactose. Several lactate dehydrogenase encoding genes were also expressed essentially in *D. hansenii* and *K. marxianus*, which are two efficient deacidifying yeasts in cheese ripening. Contrary to *D. hansenii* and *K. marxianus*, several genes involved in L-methionine catabolism were highly expressed in *Y. lipolytica*. Biochemical analyses revealed that this yeast efficiently assimilates L-methionine and also exhibited a high expression of the *S. cerevisiae* orthologues *BAT2* and *ARO8*, which are involved in the L-methionine degradation pathway (Cholet et al. 2007).

4 Importance of Volatile Sulphur Compounds on the Quality of Cheeses

Due to their high reactivity, thiols such as MTL or H_2S are common precursors for a variety of other VSCs. Both thiols arise from the degradation of the sulphur amino acids L-methionine and L-cysteine which are present in high amounts in the cheese matrix. VSCs are key compounds for the typicity and quality of ripened cheeses, giving various flavour notes to the product. However, owing to their reactivity, these thiols are not always easy to quantify in the cheese matrix. MTL is the first volatile degradation product of L-methionine (see Sect. 5) and can be found in high amounts in numerous cheeses among others camembert, vintage cheddar, parmesan, pecorino, grana padano and blue cheese. Furthermore, works have reported a good correlation between cheese flavour intensity and MTL concentration, suggesting that MTL alone could be a key contributor of the cheese flavour (Landaud et al. 2008).

Another possible VSC precursor is H_2S which is considered as the primary degradation product of L-cysteine. This thiol has the unpleasant odour of “rotten

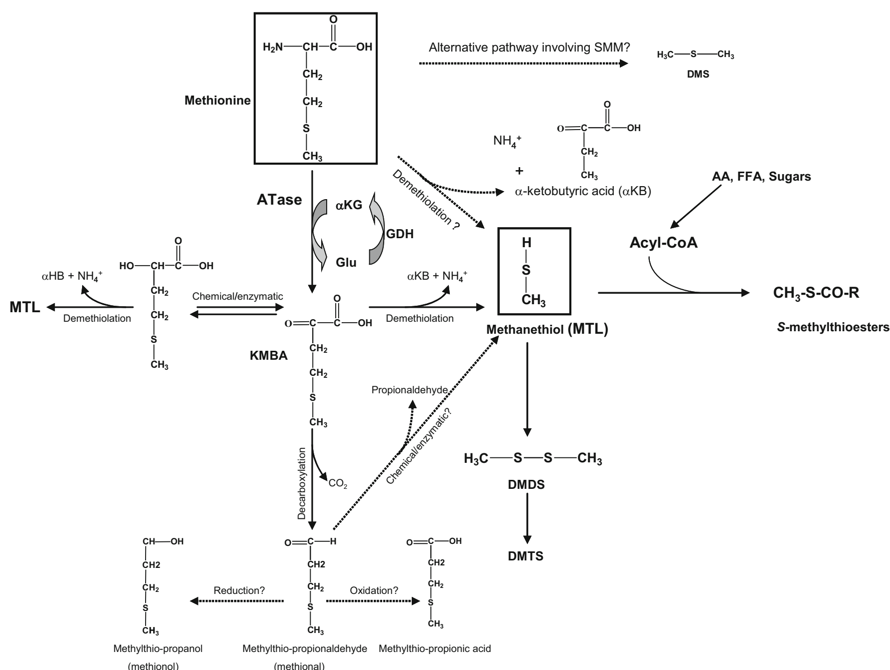


Fig. 2 Likely pathways of L-methionine catabolism to volatile sulphur compounds in *Y. lipolytica*. *ATase* aminotransferase, *α-KG* alpha ketoglutarate, *Glu* glutamate, *GDH* glutamate dehydrogenase, *SMM S*-methylmethionine, *MTL* methanethiol, *KMBA* α-keto γ-methylthiobutyrate, *AA* amino acid, *FFA* free fatty acid, *DMS* dimethyl sulphide, *DMDS* dimethyl disulphide, *DMTS* dimethyl trisulphide. *Dashed lines* are for hypothetical routes

eggs" and has been reported in several cheeses including limburger and cheddar (Landaud et al. 2008).

MTL (and possibly H_2S) are subsequently oxidized to form other VSCs such as sulphides and thioesters (Fig. 2) giving cheeses various flavour notes. For instance, DMS was reported in several types of cheeses such as camembert, cheddar or parmesan. Its aroma descriptor was found to be "boiled cabbage, sulphurous", and its odour threshold is quite low (~1 ppb), suggesting that this compound could be a contributor to the cheese aroma (Landaud et al. 2008). The most common sulphide reported in ripened cheese aroma is DMDS. DMDS has a low odour threshold (~20 ppb) and a typical "garlic" odour which is desired in the final aroma of numerous cheeses among others camembert, cheddar, parmesan, grana padano, maroilles, Livarot, pont-l'évêque, langres and époisses. Although present in very low amounts but owing to their much lower detection threshold than DMDS, other sulphides like DMTS and dimethyl tetrasulphide (DMQS) are probably more important than DMDS for the cheese aroma (Landaud et al. 2008). It has been

reported by Frank et al. (2004) that DMTS had a "strong" to "extremely strong" perceived intensity in vintage cheddar, parmesan, pecorino, grana padano and blue cheese. DMQS is less frequently detected in cheeses, but due to its much lower detection threshold (~1 ppb) compared to DMTS (~8 ppb) and DMDS (~23 ppb; Martin et al. 2004), it probably significantly influences cheese aroma, as reported in parmesan or grana padano (Frank et al. 2004).

Another family of VSCs, *S*-methylthioesters, has also been reported in several cheeses and extensively studied with respect to their detection thresholds—which ranged from 1 to 3 ppb—and flavour note descriptors such as "cabbage", "garlic" and "cheesy" (Martin et al. 2004; Landaud et al. 2008). For instance, *S*-methylthioacetate (MTA) which has been detected in vacherin, pont-l'évêque, langres and époisses has flavour descriptors such as "cabbage", "cheesy" and "crab".

Another VSC found in numerous cheeses including camembert, cheddar, blue cheese, parmesan, grana padano and pecorino is methional (methylthiopropionaldehyde). It arises from the Ehrlich degradation of L-methionine (see Sect. 5 and Fig. 2). It has a typical "boiled potato" note and could probably play an important role in the typical cheesy aroma in the cheeses where it is present (Frank et al. 2004).

Among the above-cited CSVs, MTL, H_2S, DMS, DMDS, DMTS, MTA and methional have been shown to be produced to some extent by *Yarrowia lipolytica* depending on culture media, precursors added and strain (Cholet et al. 2008; Spinnler et al. 2001; Arfi et al. 2002; Bonnarme et al. 2001). The production of MTL, DMS, DMDS, DMTS, MTA and methional was induced by L-methionine supplementation to the culture medium. The effect of low (LM) or high (HM) concentration of L-methionine was studied for *Y. lipolytica* cultivated in a cheese-like medium with respect to VSC production (Cholet et al. 2008). It was found that, in the HM medium, L-methionine consumption was accompanied by a transient accumulation of the transamination product of L-methionine, α-keto γ-methylthiobutyrate (KMBA), the latter being subsequently degraded to VSCs. MTL, DMS, DMDS and DMTS—methional being detected as traces—were detected only in the HM medium, which suggests a strong relationship between VSC production and KMBA disappearance. Other results have shown that using a medium supplemented with L-methionine, *Y. lipolytica* could also produce MTA and consistent amounts (350 ppb) of methional together with MTL, DMDS and DMTS (Arfi et al. 2002). Substituting precursor L-methionine for *S*-methylmethionine resulted in the production of only DMS by *Y. lipolytica* (Spinnler et al. 2001). In this case, it was speculated that enzymatic biosynthesis of DMS most probably proceeded through a demethiolation of *S*-methylmethionine. Nevertheless, the presence of *S*-methylmethionine in cheese still remains controversial. Another example of the importance of the precursor for VSC production profile is illustrated by L-cysteine. In *Y. lipolytica*, it was found that H_2S was the major VSCs produced from L-cysteine degradation, relative importance of H_2S production being essentially strain dependent (López del Castillo-Lozano et al. 2007a).

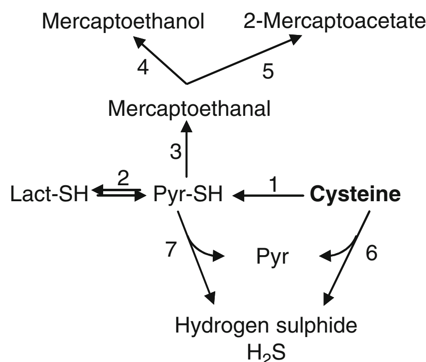


Fig. 3 Putative L-cysteine degradation pathways. *PyrSH* mercaptopyruvate, *Pyr* pyruvate, *Lact-SH* mercaptolactate. Enzymes: (*1*) transaminase, (*2*) reductase, (*3*) decarboxylase, (*4*) alcohol dehydrogenase, (*5*) oxidase, (*6*) desulphhydrase, (*7*) lyase

5 L-Methionine and L-Cysteine Catabolisms and Volatile Sulphur Compounds Biosynthesis in *Y. lipolytica*

5.1 L-Cysteine Catabolism

The catabolism of L-cysteine has been investigated in several yeasts and bacteria from cheese origin (López del Castillo-Lozano et al. 2007a). It was found that hydrogen sulphide (H_2S) production was dramatically enhanced in media supplemented with L-cysteine. Although H_2S production capabilities greatly varied depending on strain, it clearly appeared that yeast strains, especially *Y. lipolytica* YL200 (53.00 μg ml^{-1} ±1.37), were greater producers of H_2S than bacteria. Furthermore, a linear relation was found between L-cysteine consumption and H_2S production for all yeasts producing large amounts of H_2S (López del Castillo-Lozano et al. 2007a). However, in some cases including *Y. lipolytica* strains, L-cysteine was consumed without H_2S production, suggesting that γ-elimination is perhaps not the only catabolic pathway for L-cysteine. It has been hypothesized that a transamination step could be involved in the first step of L-cysteine catabolism leading to the formation mercaptopyruvate (Fig. 3). The transamination product mercapto-pyruvate (1) could be converted to H_2S and pyruvate by a chemical reaction or (2) could be reduced to the corresponding aldehyde (mercaptoethanol) as suggested in the yeast *S. cerevisiae* (Vermeulen et al. 2006) through the Ehrlich pathway (Fig. 3). However, the transsulphuration pathway and glutathione synthesis cannot be ruled out in *Y. lipolytica*, this pathway being quite active in the yeast *S. cerevisiae* as demonstrated by a combined metabolome and proteome approach (Lafaye et al. 2005).

In an attempt to study the possible effect of L-cysteine on aroma compounds production, VSC biosynthesis was studied in culture media supplemented with L-methionine or L-methionine/L-cysteine mixtures, using five cheese-ripening yeasts (López del Castillo-Lozano et al. 2007b). It was found that *Y. lipolytica* YL200 produced DMDS, and trace amounts of DMTS, 2-methyl-tetrahydroth-iophen-3-one and *S*-methylthioacetate were also produced to some extent by *Y. lipolytica*. However, VSC production diminished in a strain-dependent behaviour

when L- cysteine was supplemented, even at a low concentration (0.2 g l^{-1}). This effect was attributed to a significant decrease in L-methionine consumption in all the yeasts except YL200, for cultures supplemented with L-cysteine. Hydrogen sulphide produced through L-cysteine catabolism did not seem to contribute to VSC generation at the acid pH of yeast cultures.

5.2 Evidence of Transamination as a Key Step for L-Methionine Degradation to Volatile Sulphur Compounds

In several microorganisms isolated from cheese, the degradation of amino acids is initiated by an aminotransferase in which the amino group of an amino acid is transferred to an α-keto acid (e.g. α-ketoglutarate), resulting in the formation of the corresponding amino acid (e.g. glutamate) and keto acids, which are degraded to flavour compounds (Yvon and Rijnen 2001; Landaud et al. 2008). Evidence of a transamination as the initial degradation step of L-methionine to VSCs was first demonstrated in the yeast *G. candidum* (Bonnarme et al. 2001). In this yeast, L-methionine transamination leads to the transient accumulation of the transamination product KMBA which is subsequently converted to MTL and its direct oxidation products (e.g. DMDS, DMTS) as well as thioesters (Fig. 2). Several yeasts, among which are *Y. lipolytica*, *G. candidum*, *S. cerevisiae*, *K. lactis* and *D. hansenii*, were compared with respect to their ability to degrade L-methionine to VSCs (Arfi et al. 2002). It was found that all yeasts could produce VSCs to some extent while degrading L-methionine. Apart from *S. cerevisiae* cultures where L-methionine was poorly consumed (9–10 %), all strains had consumed almost all (≥85 %) of the L-methionine after 48 h. However, although KMBA was produced by all the yeasts, it was much more significantly accumulated by *Y. lipolytica* than did the other four yeasts; this strongly suggests that L-methionine transamination is of major importance in this yeast. Quite interestingly, while *Y. lipolytica* produced substantial amounts (350 ppb) of methional, this compound was not detected or poorly produced in the other yeasts. Methional has been detected in various types of cheeses, including cheddar and camembert (Dunn and Lindsay 1985; Kubickova and Grosch 1998). It was generally associated with a broth-like or potato odour. The production of methional by *Y. lipolytica* is likely to result from the enzymatic decarboxylation of the KMBA via the Ehrlich pathway. This suggests the occurrence of a decarboxylase activity in this yeast which converts KMBA to methional.

The availability of the genome sequence of *Y. lipolytica* enabled to carry out the functional analysis of genes putatively involved in L-methionine catabolism. However, due to the wide specificity of substrates of aminotransferases, the functional analysis of genes possibly involved in L-methionine transamination has been initiated. For instance, the enzymatic properties of the purified aromatic

aminotransferase Aro8 of *Y. lipolytica* have been investigated. Its Kms for several amino acid substrates ranged from 2.6 mM for L-phenylalanine to 12.1 mM for L- methionine and to 61.4 mM for L-cysteine. Attempt for heterologous production of other putative aminotransferases of *Y. lipolytica* has been carried out but without success until now (Nathalie Merault, personal communication). In this yeast, it has been found that L-methionine degradation via transamination, by a branched-chain aminotransferase (*BAT1*), could be involved in VSC formation. In *Y. lipolytica*, the functional analysis of a branched-chain aminotransferase gene (Yl*BCA1*) has shown that the corresponding enzyme was able to convert L-methionine to KMBA, this compound being degraded to VSCs (Cernat Bondar et al. 2005). The Yl*BCA1* gene was overexpressed in a *BCA1* transformant for which the ability to degrade L-methionine to VSCs was compared to the parental strain. A 62 % increase in KMBA biosynthesis, which is in agreement with an increase in aminotransferase activity, was obtained in the *BCA1* transformant as compared to the parental strain. This is consistent with a 55 % increase in VSC production in the modified strain. Furthermore, the thiol-producing activity was increased 2.5-fold on L-methionine in the transformant strain as compared to the parental one (Cernat Bondar et al. 2005).

Concerning the decarboxylation of KMBA to methional in *Y. lipolytica*, gene *YALI0D06930g* putatively assigned asYl*PDC6* (pyruvate decarboxylase) has been renamed Yl*ARO10* (phenylpyruvate decarboxylase) and could be a good candidate for this degradation step. In agreement with this, it was found that Yl*ARO10* was highly expressed at late stationary phase in a medium supplemented with a high concentration of L-methionine (Cholet et al. 2008). A combined proteome and transcriptome analysis of *Y. lipolytica* has been performed in response to amino acid supplementation (Mansour et al. 2009a). Following amino acid addition, yeast cells reorganize their metabolism towards amino acid catabolism. In this process, the expression of Yl*GAP1* gene encoding a membrane protein involved in amino acid transport was highly (72-fold) induced. Furthermore, the comparison of the proteome and the transcriptome data revealed a concordance of the observed effect for Yl*ARO10* and Yl*BAT2* which were both highly induced in response to amino acid supplementation. Production of DMDS following amino acid addition is an indicator of L-methionine catabolism.

5.3 *In Silico Analyses of Genes Encoding Branched-Chain Amino Acid Aminotransferases or Aromatic Acid Aminotransferases*

A search for genes encoding aminotransferases in the full genome of *Y. lipolytica* (Génolevures: Genomic Exploration of the Hemiascomycete Yeasts—http://www.genolevures.org/) has been done.

In yeasts, amino acid aminotransferases are able to catalyse the first step of the catabolism of most of the amino acids. Aminotransferases are pyridoxal phosphate (PLP)-dependent enzymes catalysing the transfer of the amino group of the amino acid to α-ketoglutarate generating the corresponding α-keto acid. The biochemical structure and functioning of several aminotransferases have been well studied and classified. BATs essentially catalyse the transamination of the branched-chain amino acids leucine, isoleucine and valine. Except for the *Escherichia coli* and *Salmonella* proteins, which are homohexamers arranged as a double trimer, BATs are homodimers. Structurally, the BATs belong to the fold type IV class of PLP enzymes. Catalysis is on the *re* face of the PLP cofactor, whereas in other classes, catalysis occurs from the *si* face of PLP. Crystal structures of the fold type IV proteins show that they are distinct from the fold type I aspartate aminotransferase family and represent a new protein fold. Because the fold type IV enzymes catalyse diverse reactions, it is not surprising that the greatest structural similarities involve residues that participate in PLP binding rather than residues involved in substrate binding (Hutson 2001). Since these enzymes have a wide specificity of substrates, several enzymes can perform the transamination of the same amino acid. This is the case for L-methionine which has no specific aminotransferase to be catabolized but relies on branched-chain amino acid aminotransferases (BATs) or aromatic acid aminotransferases (AROs).

BATs are widely distributed in the microbial kingdom, where they are involved in the synthesis/degradation of branched-chain amino acids. However, bacteria contain one single BAT, whereas in eukaryotes, there are generally two isozymes: one is mitochondrial and the other is cytosolic. In *Y. lipolytica*, two gene products belonging to the BAT protein family are found as already shown in the yeast *Saccharomyces cerevisiae*. *BAT1* encodes a protein with a mitochondrial targeting sequence, whereas *BAT2* encodes a cytoplasmic protein (Fig. 4). In *S. cerevisiae*, BAT1 and BAT2 have been reported as an ohnologue pair, which means that they would result from a whole genome duplication (Byrne and Wolfe 2005). The ohnologues that have undergone functional divergence are particularly interesting because they may indicate the adaptation of a species to a certain environment or ecological niche.

The mitochondrial location of BATs from *Y. lipolytica* was deduced from the presence of an N-terminal extension enriched in serine, threonine and polar amino acids. The amino acid sequence alignment of these genes with those of several other organisms has been performed using the ClustalX multiple sequence alignment programme. Residues identical or similar in all proteins of these sequences are dashed in black (Fig. 4a). Some species such as *S. cerevisiae* or *Y. lipolytica* have two BAT genes, one with a mitochondrial targeting signal and the second which is cytoplasmic. In other species, there is only one BAT gene and it has either cytoplasmic features as in *D. hansenii* or mitochondrial features as in *K. lactis*, suggesting that there is no preferential compartmentation for BAT. More precisely, it was found that both *Y. lipolytica* BAT genes exhibited in conserved positions, all the amino acid residues interacting with the pyridoxal phosphate attachment site deduced from structural studies carried out on crystallized enzymes such as human

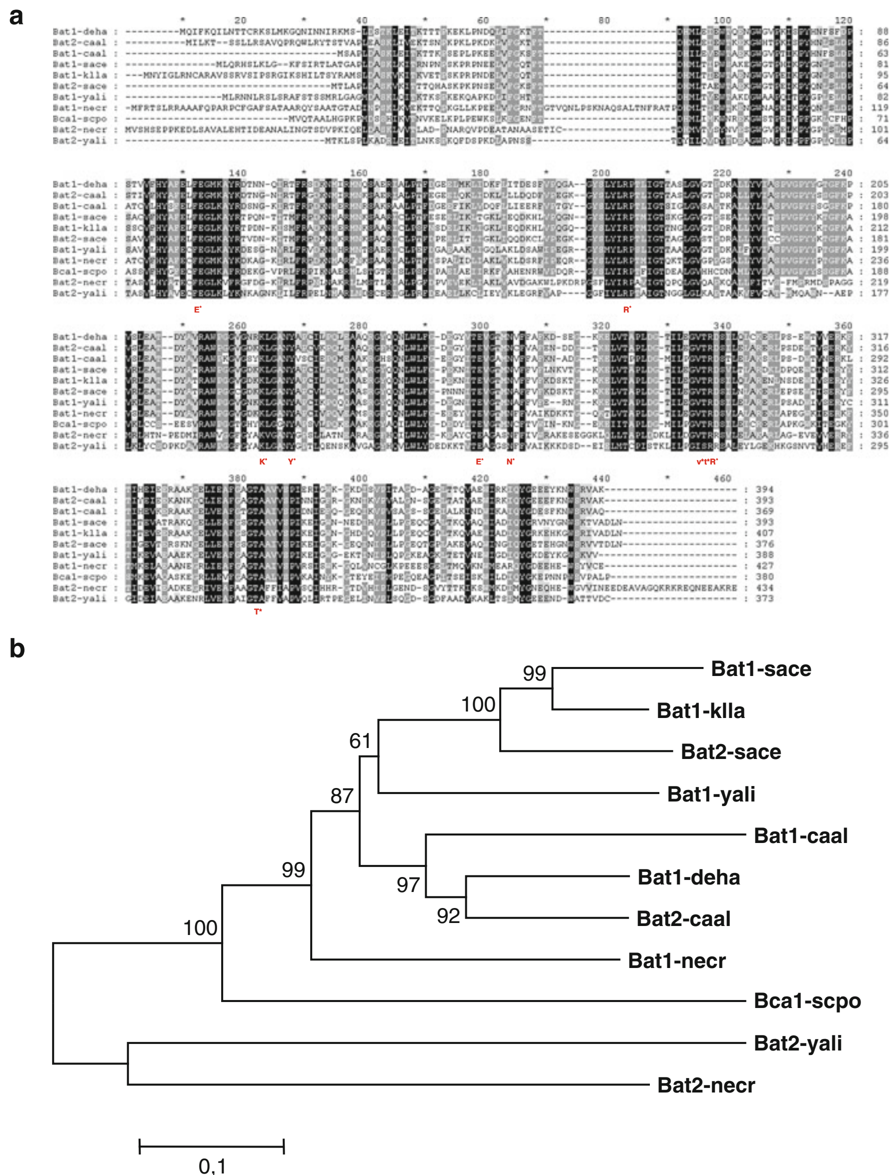


Fig. 4 (**a**) Alignments of branched-chain amino acid aminotransferases. The sequences are as follows: Bat1-Sace (*S. cerevisiae* YHR208W), Bat2-sace (*S. cerevisiae* YJR148W), Bat1-klla (KLLA0A10307g), Bat1-yali (YALI0D01265g), Bat2-yali (YALI0F19910g), Bat1-caal (CaO19.797), Bat2-caal (CaO19.6994), Bat1-deha (DEHA2D06952p), Bca1-scpo (SPBC428.02c), Bat1-necr (NCU04754) and Bat2-necr (NCU04292). The alignment created with ClustalX was formatted using Genedoc 2.6. The positions interacting with the pyridoxal phosphate are indicated underneath in *red*. (**b**) A tree of the selected BAT sequences built using the ClustalX v1.81 software and the MEGA 3.1 package. The test of phylogeny was a bootstrap of the neighbour-joining test using the default values

BCAT1 and BCAT2 (Yennawar et al. 2001). This was highlighted in the alignment. In ylBAT1, the residue Lys^{235} is involved in the formation of the Schiff base intermediate with pyridoxal phosphate (Kispal et al. 1996). Contrary to the aromatic amino acid aminotransferases (see below), there is no BAT1 and BAT2 aminotransferase subfamilies but one single family of branched amino acid aminotransferase. This was shown in the tree (Fig. 4b) where the BAT sequences are interspersed. For instance, several yeast species such as *K. lactis*, *D. hansenii*, *Saccharomyces kluyveri* or *Zygosaccharomyces rouxii* having only a mitochondrial homologue are not clustered, but species displaying this feature are dispersed along the clade of the hemiascomycetous yeasts (Fig. 4).

Concerning aromatic amino acid aminotransferases, there is one homologue of Sc*ARO8* and one homologue of Sc*ARO9* in *Y. lipolytica*. In the euascomycetes clade, both genes can be clustered in two subfamilies as shown on the tree (Fig. 5b). Neither ARO8 nor ARO9 displays any targeting signals. Their differences are scattered along the sequences (Fig. 5a). Moreover, ARO9 partners have a short N-terminal extension without identifiable function. In *S. cerevisiae*, both *ARO8* and *ARO9* control aromatic amino acid catabolism. In the case of *D. hansenii*, only an ARO8 orthologue can be identified (Fig. 5). According to Jensen and Wei (1996), they belong to the family I of the aminotransferase superfamily, and anchor residues involved mainly in pyridoxal phosphate binding can be easily identified and are indicated on Fig. 5.

5.4 Y. lipolytica *Transcriptome Analysis in Response to Various Sulphur Sources*

The regulation of genes related to sulphur metabolism has been carried out in *Y. lipolytica* in response to various sulphur sources, among others L-methionine which is the main sulphur source in casein. Expression levels of several genes predicted to be associated with L-methionine catabolism and pyruvate metabolism were simultaneously investigated at transcriptional level in *Y. lipolytica*. Gene expression profilings were analysed and compared when *Y. lipolytica* cells were grown in a cheese-like medium under high L-methionine (HM) or low L-methionine (LM) concentration (Cholet et al. 2008). In Cholet's experiments, gene expression was measured in late stationary phase when L-methionine was largely consumed. A rearrangement in the expression of some genes was observed when the L-methionine concentration was changed in the growth medium. Among them, the Yl*ARO8*, Yl*BAT1* and Yl*BAT2* genes (predicted to be involved in amino acid transamination pathway) were found to be modulated by L-methionine concentration, strongly suggesting their involvement in the L-methionine transamination step in *Y. lipolytica*. The Yl*ARO8* gene is the most strongly modulated in HM medium. Furthermore, Yl*ARO8* gene product has been overproduced in *Escherichia coli* and purified. It was found that YlAro8p had transaminase activity and was highly active on L-methionine.

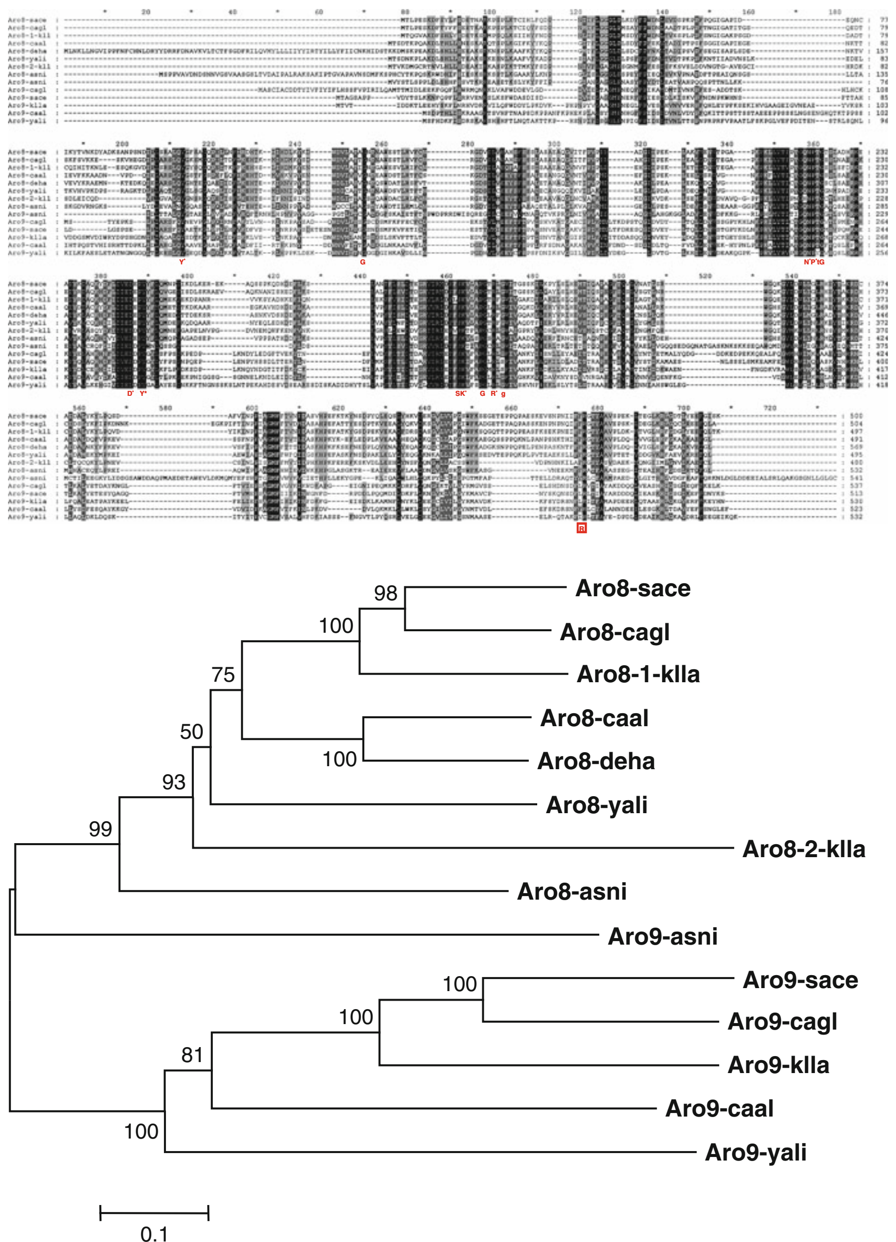


Fig. 5 (**a**) Alignments of the aromatic acid aminotransferases. The sequences are as follows: Aro8-sace (*S. cerevisiae* YGL202w), Aro9-sace (*S. cerevisiae* YHR137w), Aro8-cagl (CAGL0G01254g), Aro9-cagl (CAGL0G06028g), Aro8-1-klla (KLLA0F10021g), Aro8-2-klla (KLLA0A04906g), Aro9-klla (KLLA0D11110g), Aro8-deha (DEHA2A06886g), Aro8-caal (CaO19.9645), Aro9-caal (CaO19.1237), Aro8-yali (YALI0E20977g), Aro9-yali (YALI0C05258g), Aro8-asni (An02g05540) and Aro9-asni (An09g05080). The alignment created with ClustalX was formatted using Genedoc 2.6. The positions interacting with the pyridoxal

Cholet et al. (2008) also reported that the Yl*BAT1* gene was highly induced by L-methionine: it is in good agreement with results showing that the overexpression of the Yl*BAT1* gene significantly increased L-methionine transamination as well as VSC production (Cernat Bondar et al. 2005). In contrast, the expression levels of Yl*ARO9* gene were hardly modulated by L-methionine concentration. However, the observation that the L-methionine transamination step is highly active in *Y. lipolytica* was confirmed. A transient accumulation of the transamination product—KMBA—was measured in HM medium, which coincides with the maximum rate of L-methionine consumption and VSC production.

In other experiments in progress in our laboratory (A. Hébert, PhD work), the regulation of sulphur metabolism in response to various sulphur sources (e.g. L-methionine, cystine, sulphate) is being studied using an ORFeomic microarray containing the probes of all the identified ORFs from the *Y. lipolytica* genome. The growth conditions were different from Cholet's experiments for which cells were grown on a cheese-like medium supplemented with 6.7 mM (LM) or 40.3 mM (HM) L-methionine in late stationary phase. The work of A. Hébert was focused on the identification of the genes possibly induced by an excess of sulphur substrates. The cells were grown on a chemically defined medium mimicking a technological cheese medium and optimized to control its sulphur content. In this case, the high L-methionine concentration was fixed to 10 mM and the low L-methionine concentration was 10 μM. The same concentration ratios were used when comparing cystine and sulphate as sulphur source. These concentrations allowed the same growth rate and the cultures were kept to exponential phase for at least ten generations in order to get a steady state expression level. Under such conditions, a few genes have their expression modified under high and low sulphur concentration. However, ylBAT1 appeared the only aminotransferase to be induced on L-methionine as well as on cystine although the *BAT1* gene product should not be very active on cystine.

6 Future Perspectives

Owing to its metabolic features, the non-conventional yeast *Y. lipolytica* represents a unique model microorganism which is also of great technological interest. This review shows that the unique metabolism of Y. lipolytica—especially well adapted to milk-derived media such as cheese—is currently being investigated. This is now well established that this species is a very efficient VSC producer compared to other yeasts (Cholet et al. 2007, 2008).

Fig. 5 (continued) phosphate are indicated underneath in *red*. *R boxed* in *red* indicates a residue interacting with the substrate according Jensen et al. (**b**) A tree of the selected ARO sequences built using the ClustalX v1.81 software and the MEGA 3.1 package. The test of phylogeny was a bootstrap of the neighbour-joining test using the default values

A next step would be to study sulphur metabolism in an integrated way in pure culture by comparative genomics (e.g. conservation of pathways, regulation), functional genomics using the tools of molecular biology (e.g. gene expression), proteomics (e.g. synthesis of the corresponding enzyme) and metabolomics (e.g. identification of metabolic intermediates). A more in-depth knowledge of sulphur metabolism and its regulation is currently under investigation in this yeast. This will probably provide a better understanding of the metabolic machinery of this atypical yeast and could be of great interest for the understanding of sulphur metabolism in other microorganisms. Preliminary data strongly suggest that *Y. lipolytica* obeys quite distinct regulation mechanisms—compared to other yeasts such as *S. cerevisiae*, *K. lactis* or D. hansenii—when supplemented with various sulphur sources (e.g. L-methionine, L-cysteine, sulphate). It may also be questioned whether VSCs are only secondary products arising from sulphur amino acid catabolism produced under conditions of relative abundance of these amino acids or possess a still unidentified role in the cell life. Owing to the reductive effect of many sulphur compounds, especially thiols, a possible role of VSCs could be a regulative effect of the redox balance of the cell.

Another important step towards a better knowledge of the adaptative metabolism of *Y. lipolytica* is to study this yeast in the presence of many other species of the cheese ecosystem. It is therefore to be expected that its behaviour and metabolic adaptation may be dramatically influenced by other microbial species, including other yeasts and bacteria (Mansour et al. 2009b). So-called meta-approaches (meta-genome, meta-transcriptome, meta-metabolome) are emerging and surely will provide answers for the functional analysis of complex microbial ecosystems in which *Y. lipolytica* is involved.

References

Arfi K, Spinnler HE, Tâche R, Bonnarme P (2002) Production of volatile compounds by cheese-ripening yeasts: requirement for a methanethiol donor for *S*-methyl thioacetate synthesis by *Kluyveromyces lactis*. Appl Microbiol Biotechnol 58:503–510

Arfi K, Landaud S, Bonnarme P (2006) Evidence for distinct L-methionine catabolic pathways in the yeast *Geotrichum candidum* and the bacterium *Brevibacterium linens*. Appl Environ Microbiol 72:2155–2162

Barnett JA, Payne RW, Yarrow D (2000) *Yarrowia lipolytica*. In: Barnett JA, Payne RW, Yarrow D (eds) Yeasts: characteristics and identification, 3rd edn. Cambridge University Press, Cambridge, pp 785–786

Barth G, Gaillardin C (1997) Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica*. FEMS Microbiol Rev 19:219–237

Berger C, Martin N, Collin S, Gijs L, Khan JA, Piraprez G, Spinnler HE, Vulfson EN (1999) Combinatorial approach to flavor analysis: II. Olfactory investigation of a library of S-methylthioesters and sensory evaluation of selected components. J Agric Food Chem 47:3247–3279

Boelans MH, van Gemert LJ (1993) Sensory properties of optical isomers. Perfumer Flavorist 18:2–16

Bonnarme P, Psoni L, Spinnler HE (2000) Diversity of L-methionine catabolism pathways in cheese-ripening bacteria. Appl Environ Microbiol 66:5514–5517

Bonnarme P, Lapadatescu C, Yvon M, Spinnler HE (2001) L-methionine degradation potentialities of cheese-ripening microorganisms. J Dairy Res 68:663–674

Bonnarme P, Amarita F, Chambellon E, Semon E, Spinnler HE, Yvon M (2004) Methylthioacetaldehyde, a possible intermediate metabolite for the production of volatile sulphur compounds from L-methionine by *Lactococcus lactis*. FEMS Microbiol Lett 236:85–90

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

Cernat Bondar D, Beckerich JM, Bonnarme P (2005) Involvement of a branched-chain aminotransferase in production of volatile sulfur compounds in *Yarrowia lipolytica*. Appl Environ Microbiol 71:4585–4591

Cholet O, Henaut A, Casaregola S, Bonnarme P (2007) Gene expression and biochemical analysis of cheese-ripening yeasts: focus on catabolism of L-methionine, lactate, and lactose. Appl Environ Microbiol 73:2561–2570

Cholet O, Henaut A, Hebert A, Bonnarme P (2008) Transcriptional analysis of L-methionine catabolism in the cheese-ripening yeast *Yarrowia lipolytica* in relation to volatile sulfur compound biosynthesis. Appl Environ Microbiol 74:3356–3367

Corsetti A, Rossi J, Gobbetti M (2001) Interactions between yeasts and bacteria in the smear surface-ripened cheeses. Int J Food Microbiol 69:1–10

Cosentino S, Fadda ME, Deplano M, Mulargia AF, Palmas F (2001) Yeasts associated with Sardinian ewe's dairy products. Int J Food Microbiol 69:53–58

Dujon B, Sherman D, Fischer C, Durrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuveglise C, Talla E, Goffard N, Frangeul L, Aigle M, Anthouard V, Babour A, Barbe V, Barnay S, Blanchin S, Beckerich JM, Beyne E, Bleykasten C, Boisrame A, Boyer J, Cattolico L, Confanioleri F, De Daruvar A, Despons L, Fabre E, Fairhead C, Ferry-Dumazet H, Groppi A, Hantraye F, Hennequin C, Jauniaux N, Joyet P, Kachouri R, Kerrest A, Koszul R, Lemaire M, Lesur I, Ma L, Muller H, Nicaud JM, Nikolski M, Oztas S, Ozier-Kalogeropoulos O, Pellenz S, Potier S, Richard GF, Straub ML, Suleau A, Swennen D, Tekaia F, Wesolowski-Louvel M, Westhof E, Wirth B, Zeniou-Meyer M, Zivanovic I, Bolotin-Fukuhara M, Thierry A, Bouchier C, Caudron B, Scarpelli C, Gaillardin C, Weissenbach J, Wincker P, Souciet JL (2004) Genome evolution in yeasts. Nature 430:35–44

Dunn HC, Lindsay RC (1985) Evaluation of the role of microbial Strecker-derived aroma compounds in unclean-type flavors of cheddar cheese. J Dairy Sci 68:2859–2874

Frank DC, Owen CM, Patterson J (2004) Solid phase microextraction (SPME) combined with gas-chromatography and olfactory-mass spectrometry for characterization of cheese aroma compounds. Lebensm Wiss u Technol 37:139–154

Guerzoni ME, Lanciotti R, Vannini L, Galgano F, Favati F, Gardini F, Suzzi G (2001) Variability of the lipolytic activity in *Yarrowia lipolytica* and its dependence on environmental conditions. Int J Food Microbiol 69:79–89

Hutson S (2001) Structure and function of branched chain aminotransferases. Prog Nucleic Acid Res Mol Biol 70:175–206

Jensen RA, Wei G (1996) Evolutionary recruitment of biochemically specialized subdivisions of family I within the protein superfamily of aminotransferases. J Bacteriol 178:2161–2171

Kispal G, Steiner H, Courts DA, Rolinski B, Lill R (1996) Mitochondrial and cytosolic branched-chain amino acid transaminases from yeast, homologs of the *myc* oncogene-regulated Eca39 protein. J Biol Chem 271:24458–24464

Kubickova J, Grosch W (1998) Quantification of potent odorants in camembert cheese and calculation of their odour activity values. Int Dairy J 8:17–23

Lafaye A, Junot C, Pereira Y, Lagniel G, Tabet JC, Ezan E, Labarre J (2005) Combined proteome and metabolite-profiling analyses reveal surprising insights into yeast sulfur metabolism. J Biol Chem 280:24723–24730

Landaud S, Helinck S, Bonnarme P (2008) Formation of volatile sulfur compounds and metabolism of methionine and other sulfur compounds in fermented food. Appl Microbiol Biotechnol 77:1191–1205

López del Castillo-Lozano M, Tâche R, Bonnarme P, Landaud S (2007a) Evaluation of a quantitative screening method for hydrogen sulfide production by cheese-ripening microorganisms: the first step towards L-cysteine catabolism. J Microbiol Methods 69:70–77

López del Castillo-Lozano M, Delile A, Spinnler HE, Bonnarme P, Landaud S (2007b) Comparison of volatile sulphur compound production by cheese-ripening yeasts from methionine and methionine–cysteine mixtures. Appl Microbiol Biotechnol 75:1447–1454

Mansour S, Bailly J, Delettre J, Bonnarme P (2009a) A proteomic and transcriptomic view of amino acids catabolism in the yeast *Yarrowia lipolytica*. Proteomics 9:4714–4725

Mansour S, Bailly J, Landaud S, Monnet C, Sarthou AS, Cocaign-Bousquet M, Leroy S, Irlinger F, Bonnarme P (2009b) Investigation of associations of *Yarrowia lipolytica*, *Staphylococcus xylosus*, and *Lactococcus lactis* in culture as a first step in microbial interaction analysis. Appl Environ Microbiol 75:6422–6430

Martin N, Neelz V, Spinnler HE (2004) Suprathreshold intensity and odour quality of sulphides and thioesters. Food Qual Prefer 15:247–257

Mounier J, Goerges S, Gelsomino R, Vancanneyt M, Vandemeulebroecke K, Hoste B, Brennan NM, Scherer S, Swings J, Fitzgerald GF, Cogan TM (2006) Sources of the adventitious microflora of a smear-ripened cheese. J Appl Microbiol 101:668–681

Mounier J, Monnet C, Vallaeys T, Arditi R, Sarthou AS, Helias A, Irlinger F (2008) Microbial interactions within a cheese microbial community. Appl Environ Microbiol 74:172–181

Mounier J, Monnet C, Jacques N, Antoinette A, Irlinger F (2009) Assessment of the microbial diversity at the surface of Livarot cheese using culture-dependent and independent approaches. Int J Food Microbiol 133:31–37

Spinnler HE, Berger C, Lapadatescu C, Bonnarme P (2001) Production of sulfur compounds by several yeasts of technological interest for cheese ripening. Int Dairy J 11:245–252

van den Tempel T, Jakobsen M (2000) The technological characteristics of *Debaryomyces hansenii* and *Yarrowia lipolytica* and their potential as starter cultures for production of Danablu. Int Dairy J 10:263–270

Vermeulen C, Lejeune I, Tran TTH, Collin S (2006) Occurrence of polyfunctional thiols in fresh lager beers. J Agric Food Chem 54:5061–5068

Viljoen BC, Khoury AR, Hattingh A (2003) Seasonal diversity of yeasts associated with white-surface mould-ripened cheeses. Food Res Int 36:275–283

Wood AF, Aston JW, Douglas GK (1985) The determination of free amino acids in cheese by capillary column gas liquid chromatography. Aust J Dairy Technol 40:166–169

Yennawar NH, Dunbar J, Conway ME, Hutson SM, Farber GK (2001) The structure of human mitochondrial branched-chain aminotransferase. Acta Crystallogr D Biol Crystallogr 57:506–515. doi:10.1107/S0907444901001925

Yvon M, Rijnen L (2001) Cheese flavour formation by amino acid catabolism. Int Dairy J 11:185–201

Index

A
A and B boxes, 82, 85, 87, 100, 102–104
Acetate transporter, 159
Acid protease Axp, 122
  Axp, 124, 134
ACS I, 113
ACS II, 113
Acyl carrier proteins (mitochondrial), 48
Acyl-CoA synthetase, 112
α-keto g-methylthiobutyrate, 175
Aldehydes, 112
Alkaline protease Aep, 122, 124
Alkane responsive element 1, 116
Alternative NADH dehydrogenase, 33
Alternative splicing, 73
Alternative terminal oxidase (AOX), 33
Aminotransferases, 176, 177, 179
Ammonium-proton antiporter, 159
Ammonium secretion, 159
Annotation, 4
AOX. *See* Alternative terminal oxidase (AOX)
Aroma precursors, 170
Arrestin, 121, 127, 129, 136
AXP1, 122, 133, 134

B
Bro1 domain, 130, 131

C
Cell wall assembly, 135
Centromeres, 7
Cheese ripening, 170
Class II elements, 70
Colletotrichum acutatum, 136
Complex I, 35
  affinity purification, 41
  catalytic core, 39
  inhibitors, 39
  iron-sulfur clusters, 40
  mutagenesis, 39
  quinone binding pocket, 40
  reaction mechanism, 49–51
  specific assembly factors, 49
  subcomplexes, 45
Cytochrome P450ALK, 114

D
de novo homocysteine synthesis, 169
Dimorphism, 135
DMDS, 172
DNA metabolism, 20
Dominant acetic acid sensitivity, 150

E
ESCRT, 121, 126, 132, 137
  ESCRT-0, 126
  ESCRT-I, 121, 124, 126, 129, 132, 136, 137
  ESCRT-II, 126
  ESCRT-III, 124, 126, 130, 136
  machinery, 121, 124–126, 129, 132,
    136, 137
Evolutionary profiling, 49

F
Fatty acids, 112
Fatty alcohol dehydrogenase, 115
Fatty alcohol oxidase, 115
Fatty aldehyde dehydrogenase, 116
Fusarium oxysporum, 136

G. Barth (ed.), *Yarrowia lipolytica*, Microbiology Monographs 24,
DOI 10.1007/978-3-642-38320-5, © Springer-Verlag Berlin Heidelberg 2013

G
Gene
  acid-induced, 121, 123, 131, 133, 134, 138
  alkaline-induced, 123, 125, 131–133, 138
  counts, 12–13
  families, 17
Génolevures I programme, 2
Genome feature nomenclature, 5
Genome sizes, 5
Genomic rearrangements, 41
Glyoxylate pathway regulation, 147
Gpr1/FUN34/YaaH, 146
Gpr1p, 146

H
Homocysteine, 168
Horizontal transfer, 18, 63

I
Ind1 assembly factor, 49
Introns, 14–15

L
L-cysteine, 174
Leigh syndrome, 48
Libraries of cDNA, 5
Liquid bead ion desorption (LILBID) mass spectrometry, 43
Livarot cheese, 170
L-methionine, 171, 173, 179, 181
  permease, 169
  transamination, 175
Long chain alcohols, 112
Long terminal repeat (LTR), 59
Lotka-Volterra model, 170

M
Mating, 20–21, 135
Mercapto-pyruvate, 174
Metabolic pathways, 20
Methional, 173, 175
Mitochondrial DNA, 25
Mitochondrial genome, 42–43
  introns, 42
  translation table, 42
Mitochondrial tRNA, 42
mRNA metabolism, 21
MULE family, 71
mus51 gene, 41
Mutator-like elements, 71
Mutyl, 71

N
n-alkane, 113
NiFe hydrogenases, 39
Non-classical secretion, 151
Non-LTR class I element, 65
Nutrient deficiency, 158

O
Opi1p, 117
ORFeomic microarray, 181
Origins of replication, 9
Overview of genome organisation, 5

P
PacC, 123
Pal, 122
PalC, 124
pH sensor, 121, 128, 129, 136, 137
Phylogeny, 80, 104
Plasma membrane patches, 150
PLP enzymes, 177
Pol, 59
Pol III genes, 82
  snoRNA genes, 82, 103
  5S rRNA genes, 82, 96–98
  tRNA genes, 86, 87, 89, 94, 97, 99
  U6 RNA genes, 82, 100
Pol III transcription factors
  TFIIIA, 82, 96, 98, 99
  TFIIIB, 82, 85
  TFIIIC, 82, 85, 87, 104
Pol II RNA polymerase, 82
Pol I RNA polymerase, 82
Polycistronic Pol III genes, 86
  dicistronic tRNA genes, 89
  dicistronic tRNA/5S-rRNA genes, 97–99
  dicistronic tRNA-Trp/RUF70 genes, 99
Protease
  signalling, 130
Proteasome, 127, 130, 137
Pseudogenes, 16

Q
Quinone binding pocket, 50

R
Redox balance, 182
Redox-Bohr effect, 50
Resistance to high levels of Ca+, Na+ and Li+, 135
Retrotransposons, 59
Retroviral
  budding, 137
  Gag proteins, 121, 137
Reverse transcriptase (RT), 60, 61, 65–68
Ribosomal DNA, 11
RIM101, 125, 132, 133
Rim8p, 129, 136
  Rim8p/PalF, 121, 124, 127, 129, 132, 136
Rim9p, 128
  Rim9p/PalI, 121, 128, 136, 138
Rim13p, 130, 133
  Rim13p/PalB, 121, 125, 130–131, 136
Rim20p
  Rim20p/PalA, 121, 124, 130, 131, 136
Rim21p, 128, 138
  Rim21p/PalH, 121, 127, 129, 136–138
Rim23p, 124
  Rim23p/PalC, 121, 125, 131, 137
Rim101p, 123, 126
  binding, 127
  processing, 126
Rim101p/PacC, 121, 123, 126–127, 130, 131, 136
RNA polymerase III, 60
RT. *See* Reverse transcriptase (RT)

S
S-adenosylmethionine, 169
Sclerotinia sclerotiorum, 126, 136
Semiquinone, 51
Signalling protease, 121, 130, 136
Single transmembrane domain (STMD) subunits, 45
S-methylmethionine, 173
S-methylthioesters, 173
SNF7p, 124, 126, 130–132, 136–138
Sporulation, 135
Staphylococcus xylosus, 170
Strain E150, 4
Strain W29, 4
Sulfurtransferase, 46
Sulphate activation, 168
Sulphate transporter, 167
Sulphur metabolism, 167
Supercomplexes, 43

T
Target site duplication (TSD), 68
Telomeric repeats, 10
Thiols, 171
  H_2S, 171
  MTL, 171
Transcriptome analysis, 176, 179
Transporters, 22
Transposable elements, 58
Transposons, 23
tRNA, 12
TSD. *See* Target site duplication (TSD)
Ty elements, 59
Tyl6, 62–64

U
Ustilago maydis, 122, 135, 136

V
Vectorial acylation, 112
Virulence, 122, 135, 136
Vps23p, 121, 124, 129, 132, 136

X
XPR2, 122, 123, 125, 127, 132, 133

Y
Yarrowia clade, 3
Yarrowia lipolytica, 32
  complex I, 35–40
  respiratory chain, 33–35
Yarrowia specific genes, 15–16
Yas1p and Yas2p, 116
YlHFD1, 125, 133
Ylli, 66
YlPHR1, 125, 132, 133, 135
YlPHR2, 125, 133, 135
Ylrim9, 133, 134
Ylrim13, 125, 132, 133
Ylrim23, 131
Ylt1, 60–62
YlVPS23, 124, 132, 133
YlVPS28, 124, 132, 133
YNAYA motif, 158

Z
Zorros, 66