

# Comparative Genomics of Hemiascomycete Yeasts: Genes Involved in DNA Replication, Repair, and Recombination

Guy-Franck Richard, Alix Kerrest, Ingrid Lafontaine, and Bernard Dujon

Unité de Génétique Moléculaire des Levures (URA 2171 CNRS, UFR 927 Université Pierre et Marie Curie), Institut Pasteur, Paris cedex, France

Among genes conserved from bacteria to mammals are those involved in replicating and repairing DNA. Following the complete sequencing of four hemiascomycetous yeast species during the course of the *Génolevures 2* project, we have studied the conservation of 106 genes involved in replication, repair, and recombination in *Candida glabrata*, *Kluyveromyces lactis*, *Debaryomyces hansenii*, and *Yarrowia lipolytica* and compared them with their *Saccharomyces cerevisiae* orthologues. We found that proteins belonging to the replication fork and to the nucleotide excision repair pathway were—on the average—more conserved than proteins involved in the checkpoint response to DNA damage or in meiotic recombination. The meiotic recombination proteins Spo11p and Mre11p-Rad50p, involved in making meiotic double-strand breaks (DSBs), are conserved as is Mus81p, involved in resolving meiotic recombination intermediates. Interestingly, genes found in organisms in which DSB-repair is required for proper synapsis during meiosis are also found in *C. glabrata*, *K. lactis*, and *D. hansenii* but not in *Y. lipolytica*, suggesting that two modes of meiotic recombination have been selected during evolution of the hemiascomycetous yeasts. In addition, we found that *SGS1* and *TOP1*, respectively, a DEAD/DEAH helicase and a type I topoisomerase, are duplicated in *C. glabrata* and that *SRS2*, a helicase involved in homologous recombination, is tandemly duplicated in *K. lactis*. Phylogenetic analyses show that the duplicated *SGS1* gene evolved faster than the original gene, probably leading to a specialization of function of the duplicated copy.

## Introduction

A common concern of all living organisms is how to replicate, maintain, and transfer to the next generation an intact pool of chromosomes. For that reason, they have developed a number of partly redundant machineries in order to ensure proper duplication and repair of their genome content. Proteins involved in these machineries are most of the time conserved during evolution from bacteria to mammals (Cann and Ishino 1999; Lindahl and Wood 1999; Pâques and Haber 1999; Zhou and Elledge 2000; Burgers et al. 2001). However, constraints are clearly different between organisms with small and compact genomes and those with large genomes containing numerous repetitive elements. Constraints are also different between unicellular organisms with short generation times and metazoans. Finally, constraints are different between organisms going through a mode of sexual reproduction followed by meiosis and those whose reproduction mode is asexual. During the course of the *Génolevures 2* project (Dujon et al. 2004), four hemiascomycetous yeast species were fully sequenced. *Candida glabrata* is a pathogenic yeast, the second causative agent of human candidiasis, phylogenetically related to *Saccharomyces cerevisiae* (Bennett, Izumikawa, and Marr 2004). *Kluyveromyces lactis* is also related to *S. cerevisiae* and has been used for genetic studies or industrial applications like the production of  $\beta$ -galactosidase (Bolotin-Fukuhara et al. 2000). *Debaryomyces hansenii* is a marine yeast that can tolerate high salinity levels, phylogenetically close to the pathogenic *Candida albicans* (Lépingle et al. 2000). *Yarrowia lipolytica* is a more distantly related yeast, able to grow as individual yeast cells or as a mycelium (Casarégola et al. 2000). The evolutive distance between *S. cerevisiae* and *Y. lipoly-*

*tica*, measured as the amino acid divergence between proteins, is comparable to the entire phylum of Chordates (Dujon et al. 2004). However, genome sizes and general organization are comparable among the five hemiascomycetes sequenced. Hence, constraints on DNA replication and repair should be similar, and any difference detected should reflect a mechanistic difference between such machineries. In the present work, we have analyzed the genomes of the four newly sequenced hemiascomycetous yeasts to look for the presence of 106 genes known to be involved in replication, repair, and recombination in *S. cerevisiae*. We found that some machineries are very well conserved whereas others have diverged more rapidly. In addition, two genes (*SGS1* and *TOP1*) are duplicated in *C. glabrata* and one (*SRS2*) is duplicated in *K. lactis*.

## Materials and Methods

### Analysis of Gene Families

We started from protein families built from sequence similarities during *Génolevures 2* (Dujon et al. 2004) (<http://cbi.labri.fr/Genolevures>). When a family contained only five members and one gene per sequenced species, we considered that this gene was the correct orthologue of the *S. cerevisiae* gene. This happened in 41 cases out of 106. In the three cases of larger gene families (the RFC, RPA, and MCM families, containing altogether 14 genes), orthologues could not be chosen among paralogues based on sequence similarity but were determined based on synteny conservation. No homologue to a *S. cerevisiae* gene was found in any species in only three cases. Finally, in the remaining 48 cases, more than one gene matched with the *S. cerevisiae* gene in at least one sequenced species. In these cases, we used three different criteria to select the correct putative orthologue. First, we performed global alignments using the Needleman-Wunsch algorithm and rejected all alignments with less than 20% identity, unless a portion of the protein showed a very strong similarity to the *S. cerevisiae* protein. Second, we looked for synteny

Key words: hemiascomycete, comparative genomics, replication, repair, recombination.

E-mail: gfrichar@pasteur.fr.

*Mol. Biol. Evol.* 22(4):1011–1023. 2005

doi:10.1093/molbev/msi083

Advance Access publication January 12, 2005

conservation between the *S. cerevisiae* gene and the corresponding gene in a region covering 10 genes upstream and 10 genes downstream. Synteny was considered conserved if at least three genes (including the query) were conserved in the correct order. Evidence of synteny was found in 171 cases out of 240 (63 cases out of 64 for *C. glabrata*, 54 out of 60 for *K. lactis*, 38 out of 65 for *D. hansenii*, and 16 out of 51 for *Y. lipolytica*). In six cases, synteny conservation was not found with *S. cerevisiae* but with at least another species. Third, a possible conserved motif was searched using the Conserved Domain Database (CDD) (<http://www.ncbi.nlm.nih.gov/80/structure/cdd/cdd.shtml>). If none of these three approaches gave a significant result, the corresponding gene was discarded. In the three cases of species-specific gene duplication (*SGS1*, *TOP1*, and *SRS2*), multiple alignments of the homologues in the five species were performed using ClustalW. We also performed tBlastn searches using the orthologue sequence in the closest species as the query sequence. This approach detected five novel orthologues not found by previous methods. In addition, we used PSI-Blast (on the National Center for Biotechnology Information [NCBI] server, with E value threshold = 0.1, restricted to fungal genomes) to look for *S. cerevisiae* genes absent in *C. glabrata* and *K. lactis*, but we did not find any sequence similar to them. Finally, in 10 specific cases (*RFA3*, *RAD28*, *LIF1*, *LIF2/NEJ1*, *MEI4*, *MER1*, *REC104*, *SAE3*, *SPO13*, and *TAM1/NDJ1*), when no orthologue was found by the above methods in one or more species, we tried to detect degenerate homologues using the possible synteny conservation with *S. cerevisiae*. For example, *RFA3* (YJL173c) is located between YJL172w and YJL174w in *S. cerevisiae*. We examined the DNA sequence between these two genes in *D. hansenii* in order to find a possible degenerate orthologue. When a gene was found at the expected position, we tried to align its sequence with the *S. cerevisiae* gene. Using this method, we found a possible orthologue in only one case, *TAM1/NDJ1* in *K. lactis*. All the results are summarized in table 1 and the supplementary table.

#### Calculation of the Mean Conservation of *S. cerevisiae* Gene Products and Their Orthologues in the Four Other Genomes

A predicted gene from a given genome (*Cagl*, *Klla*, *Deha*, or *Yali*) was considered orthologous to a *S. cerevisiae* gene if it was found in a region of conserved synteny between the two genomes. Using this criterion, 3,935 orthologues were found with *C. glabrata*, 3,440 with *K. lactis*, 279 with *D. hansenii*, and 107 with *Y. lipolytica*. The percentage of amino acid identity between a given *S. cerevisiae* gene and its corresponding orthologue was obtained from the Smith-Waterman alignment between the two sequences. Means were calculated using these values. For construction of synteny maps and precise parameters used for alignments see Dujon et al. (2004).

#### Phylogenetic Analyses

Multiple alignments of amino acid sequences were performed using T-coffee (Notredame, Higgins, and Heringa

2000). Gaps and poorly aligned sequences were excluded from alignments using Gblocks (Castresana 2000). Tree reconstruction was performed by the maximum likelihood algorithm as implemented in PHYML (Guindon and Gascuel 2003). The substitution process was modeled by the Jones, Taylor, and Thornton (JTT) model, the heterogeneity of substitution rates among sites was modeled by a gamma distribution, with four categories and a parameter estimated from the data set. Tree topology and support of internal branches were inferred by 500 bootstrap calculations. Calculations of the nonsynonymous/synonymous substitution rate ratio ( $\omega = Dn/Ds$ ) were performed with the maximum likelihood method (Goldman and Yang 1994) implemented in the PAML package version 3.14 (Yang 1997).

## Results

In order to study the evolution of DNA replication, repair and recombination pathways among five completely sequenced yeast genomes, we studied the conservation of 106 different *S. cerevisiae* genes, selected on their known function as deduced from genetics or biochemistry. Out of 106 genes, 101 have orthologues in *C. glabrata*, 100 in *K. lactis*, 85 in *D. hansenii*, and only 70 in *Y. lipolytica* (table 1 and supplementary table). The only five genes not detected in *C. glabrata* are involved in meiotic recombination and nonhomologous end joining (NHEJ). In *Y. lipolytica*, one or more gene is not detected in each pathway, except in the nucleotide excision repair (NER) pathway. When a gene is not found, it means either that the gene is not present in the considered organism or that the sequence is too diverged to be recognized using our criteria (see *Materials and Methods*). Out of 29 *S. cerevisiae* essential genes, only three (*DNA2*, *RFA3*, and *DDC2*) are not perfectly conserved in all four hemiascomycetes studied (table 1). This suggests that most of the essential genes in *S. cerevisiae* also encode products that are essential (or at least important enough to be selected for) in the four other species. Seven genes in our list of 106 *S. cerevisiae* genes are split by an intron. Only two of these contain an intron in at least one of the four other hemiascomycetes. Similarly, five introns are predicted by the sequence in the four newly sequenced hemiascomycetes, but none of them is found in the same gene in at least two species. This suggests that introns are differentially lost and acquired during evolution, in accordance with a former study on 13 partially sequenced hemiascomycetous yeasts (Bon et al. 2003).

#### High Conservation of Genes Involved in S-Phase Replication

It is not surprising that almost all the proteins playing a role in chromosome replication (Burgers et al. 2001) are conserved throughout the hemiascomycete evolution. One notable exception however, is Rfa3p, one of the tripartite components of the yeast single-strand binding protein complex, which is not detected in *D. hansenii* and *Y. lipolytica*, the two other members (Rfa1p and Rfa2p) of the same heterotrimeric complex being found. No *RFA3* gene relic was found in these two species. In addition, the gene encoding Dna2p, involved in processing Okazaki fragments, contains

**Table 1**  
**Conservation of Replication, Repair, and Recombination Genes of *Saccharomyces cerevisiae* in the Four Completely Sequenced Hemiascomycetous Yeasts**

Gene <sup>a</sup>	ORF	Human Homologue	Function <sup>b</sup>	Cg	Kl	Dh	Yl
Replication and DNA topology							
<i>CDC9</i>	YDL164c	Lig I	Ligase	1	1	1	1
<i>DNA2</i>	YHR164c	Dna2	Nuclease	1	1	1	1 <sup>c</sup>
<i>MCM2</i>	YBL023c	Mcm2	Helicase	1	1	1	1
<i>MCM3</i>	YEL032w	Mcm3	Helicase	1	1	1	1
<i>MCM4</i>	YPR019w	Mcm4	Helicase	1	1	1	1
<i>MCM5</i>	YLR274w	Mcm5	Helicase	1	1	1	1
<i>MCM6</i>	YGL201c	Mcm6	Helicase	1	1	1	1
<i>MCM7</i>	YBR202w	Mcm7	Helicase	1	1	1	1
<i>POL1</i>	YNL102w	Pol $\alpha$	Polymerase	1	1	1	1
<i>POL12</i>	YBL035c	Pol $\alpha$ (subunit)	Polymerase	1	1	1	1
<i>POL2</i>	YNL262w	Pol $\epsilon$	Polymerase	1	1	1	1
<i>POL3</i>	YDL102w	Pol $\delta$	Polymerase	1	1	1	1
<i>POL30</i>	YBR088c	PCNA	Clamp	1	1	1	1
<i>POL5</i>	YEL055c	—	Polymerase	1	1	1	1
<i>PRII</i>	YIR008c	p48	Polymerase	1	1	1	1
<i>PRI2</i>	YKL045w	p58	Polymerase	1	1	1	1
<i>RAD27</i>	YKL113c	Fen1	Nuclease	1	1	1	1
<i>RFA1</i>	YAR007c	RPA1	SSB	1	1	1	1
<i>RFA2</i> (i)	YNL312w	RPA2	SSB	1	1	1	1
<i>RFA3</i>	YJL173c	RPA3	SSB	1	1	1 <sup>d</sup>	1
<i>RFC1</i>	YOR217w	Rfc1	Clamp loader	1	1	1	1
<i>RFC2</i>	YJR068w	Rfc2	Clamp loader	1	1	1	1
<i>RFC3</i>	YNL290w	Rfc3	Clamp loader	1	1	1	1
<i>RFC4</i>	YOL094c	Rfc4	Clamp loader	1	1	1	1
<i>RFC5</i>	YBR087w	Rfc5	Clamp loader	1	1	1	1
<i>TOP1</i>	YOL006c	Topo I	Topoisomerase	2	1	1	1
<i>TOP2</i>	YNL088w	Topo II	Topoisomerase	1	1	1	1
<i>TOP3</i>	YLR234w	Topo III	Topoisomerase	1	1	1	1
<i>SGS1</i>	YMR190c	<sup>e</sup>	Helicase	2	1	1	1
NER							
<i>RAD1</i>	YPL022w	ERCC4 (XPF)	Nuclease	1	1	1	1
<i>RAD10</i>	YML095c	ERCC1	Nuclease	1	1	1	1
<i>RAD14</i> (i)	YMR201c	XPA	DNA binding	1	1	1	1
<i>RAD16</i>	YBR114w	—	DNA binding	1	1	1 <sup>f</sup>	1
<i>RAD7</i>	YJR052w	—	DNA binding	1	1	1	1
<i>RAD2</i>	YGR258c	ERCC5 (XPG)	Helicase	1	1	1	1
<i>RAD23</i>	YEL037c	HR23A, HR23B	DNA binding	1	1	1	1
<i>RAD4</i>	YER162c	XPC	DNA binding	1	1	1	1
<i>RAD25</i>	YIL143c	ERCC3 (XPB)	Helicase	1	1	1	1
<i>RAD3</i>	YER171w	ERCC2 (XPD)	Helicase	1	1	1	1
Repair							
<i>RAD26</i>	YJR035w	ERCC6 (CSB)	TCR	1	1	1	1
<i>RAD28</i>	YDR030c	CSA	TCR	1	1 <sup>d</sup>	1	1
<i>REV3</i>	YPL167c	Pol $\zeta$	Polymerase	1	1	1	1
<i>RAD6</i>	YGL058w	HR6B	Ubiq.-conjug.	1	1	1(2i)	1
<i>RAD18</i>	YCR066w	Rad18	DNA binding	1	1	1	1
<i>RAD5</i>	YLR032w	—	—	1	1	1	1
<i>POLA</i>	YCR014c	Pol $\lambda$	Polymerase	1	1	—	1
MMR							
<i>MLH1</i>	YMR167w	Mlh1	DNA binding	1	1	1	1
<i>PMS1</i>	YNL082w	Pms2	DNA binding	1	1	1 <sup>f</sup>	1
<i>MLH2</i>	YLR035c	Pms1	—	1	1	—	—
<i>MLH3</i>	YPL164c	-	—	1	1	1	—
<i>MSH1</i>	YHR120w	-	—	1	1	1	1
<i>MSH2</i>	YOL090w	Msh2	DNA binding	1	1	1 <sup>f</sup>	1
<i>MSH3</i>	YCR092c	Msh3	DNA binding	1	1	1	1
<i>MSH6</i>	YDR097c	GTBP	DNA binding	1	1	1	1
NHEJ							
<i>DNL4</i>	YOR005c	Lig IV	Ligase	1	1	1	1
<i>LIF1</i>	YGL090w	Xrcc4	—	1	1 <sup>d</sup>	—	—
<i>LIF2/NEJ1</i>	YLR265c	—	—	1 <sup>d</sup>	1	—	—
<i>HDF1</i>	YMR284w	Ku70	DNA binding	1	1	1	1
<i>HDF2</i>	YMR106c	Ku80	DNA binding	1	1	1	1
<i>MRE11</i>	YMR224c	Mre11	Nuclease	1	1	1	1
<i>RAD50</i>	YNL250w	Rad50	DNA binding	1	1	1 <sup>f</sup>	1
<i>XRS2</i>	YDR369c	Nbs1	—	1	1	—	—
HIR							
<i>EXO1</i>	YOR033c	—	Nuclease	1	1	1	1

**Table 1**  
**Continued**

Gene <sup>a</sup>	ORF	Human Homologue	Function <sup>b</sup>	Cg	Kl	Dh	Yl
<i>RAD51</i>	YER095w	<sup>g</sup>	Strand exch.	1	1	1	1
<i>RAD52</i>	YML032c	<sup>h</sup>	DNA binding	1	1	1	1
<i>RAD54</i>	YGL163c	Rad54		1	1	1 <sup>f</sup>	1
<i>RAD55</i>	YDR076w	<sup>g</sup>		1	1		
<i>RAD57</i>	YDR004w	<sup>g</sup>		1	1	1	1
<i>RAD59</i>	YDL059c	—		1	1		
<i>RDH54/TID1</i>	YBR073w	Rad54B		1	1	1	1
<i>SRS2</i>	YJL092w	—	Helicase	1	2	1	1
Meiotic recombination							
<i>DMC1</i> (i)	YER179w	Dmc1	Strand exch.	1	1	1	
<i>HOP1</i>	YIL072w	—	Synap. comp.	1	1	1	
<i>HOP2</i> (i)	YGL033w	GT198		1	1	1	
<i>MEI4</i>	YER044c-a	—		1	<sup>d</sup>	<sup>d</sup>	
<i>MEI5</i>	YPL121c	—		1	1	1	
<i>MEK1</i>	YOR351c	Chk2	Kinase	1	1	1(i)	
<i>MER1</i>	YNL210w	—	Splicing	<sup>d</sup>	1		
<i>MER2</i> (i)	YJR021c	—		1(i)	1		
<i>MER3</i> (i)	YGL251c	—	Helicase	1	1	1	
<i>MRE2/NAM8</i>	YHR086w	PABP-1	Splicing	1	1	1	1
<i>MSH4</i>	YFL003c	—	DNA binding	1	1		
<i>MSH5</i>	YDL154w	—	DNA binding	1	1		
<i>MUS81</i>	YDR386w	Mus81	Nuclease	1	1	1	1
<i>MMS4</i>	YBR098w	—	Nuclease	1	1	1	
<i>REC102</i>	YLR329w	—		1	1		
<i>REC104</i>	YHR157w	—		<sup>d</sup>	<sup>d</sup>		
<i>REC114</i> (i)	YMR133w	—		1	1(i)		
<i>RED1</i>	YLR263w	—		1	1		
<i>SAE2</i>	YGL175c	—		1	1	1	
<i>SAE3</i>	YHR079c-b	—		<sup>d</sup>	<sup>d</sup>	<sup>d</sup>	
<i>SPO11</i>	YHL022c	Spo11	Topoisomerase	1	1	1	1
<i>SPO13</i>	YHR014w	—		<sup>d</sup>	<sup>d</sup>	<sup>d</sup>	
<i>TAM1/NDJ1</i>	YOL104c	—	DNA binding	1	1		
<i>ZIP1</i>	YDR285w	—	Synap. comp.	1 <sup>c</sup>	1	1	
<i>ZIP2</i>	YGL249w	—		1	1		
Checkpoints							
<i>CHK1</i>	YBR274w	Chk1	Kinase	1	1	1	
<i>RAD24</i>	YER173w	Rad17	Clamp loader	1	1	1	1
<i>DDC1</i>	YPL194w	Rad9	Clamp	1	1	1	
<i>MEC3</i>	YLR288c	Hus1	Clamp	1	1	1	
<i>RAD17</i>	YOR368w	Rad1	Clamp	1	1	1	1
<i>DDC2</i>	YDR499w	ATRIP	DNA binding	1	1	1	
<i>MEC1</i>	YBR136w	ATR	Kinase	1	1	1	1
<i>RAD53</i>	YPL153c	Chk2	Kinase	1	1	1	1
<i>RAD9</i>	YDR217c	—		1	1		
<i>TEL1</i>	YBL088c	ATM	Kinase	1	1	1	1

NOTE.—(i): intron(s). *RAD6* contains two predicted introns.

<sup>a</sup> Number of orthologues to *Saccharomyces cerevisiae* genes (left column) for each yeast species are listed in the last four columns. Genes are classified according to the pathway they belong to (see supplementary table for gene names in the four new hemiascomycetes). Essential genes are underlined.

<sup>b</sup> Functions are indicated when known. Clamp and Clamp loader refer to the processivity factor (PCNA) and its associated loading complex (RFC) during S-phase replication or to the alternate complexes associated to checkpoint activity (Ddc1p-Mec3p-Rad17p and RFC-Rad24p). Strand exch.: Rad51p and Dmc1p exhibit RecA-like strand exchange activity during homologous recombination. Synap. comp.: protein is part of the meiotic synaptonemal complex. SSB: single-strand DNA-binding complex. TCR: protein involved in transcription-coupled repair. Ubiq.-conjug.: ubiquitin-conjugating enzyme (E2). Precise catalytic activities of nucleases, helicases, topoisomerases, polymerases, kinases, and DNA-binding proteins were not detailed to save space.

<sup>c</sup> Probable pseudogene or intron (high-quality sequence containing in-frame stop codons or frameshifts).

<sup>d</sup> Cases in which the synteny with *S. cerevisiae* was used to try to find a putative orthologue (see *Materials and Methods*).

<sup>e</sup> Blm, Wrm, RecQ4, RecQL, RecQ5.

<sup>f</sup> Possible pseudogene or intron (low-quality sequence containing in-frame stop codons or frameshifts).

<sup>g</sup> Rad51B, Rad51C, Rad51D, Xrcc2, Xrcc3.

<sup>h</sup> Rad52, Rad52B.

three in-frame stop codons in *Y. lipolytica* and thus is most probably not properly translated. Dna2p function is at least partly redundant with Rad27p, suggesting that Rad27p is necessary and sufficient to process Okazaki fragments in *Y. lipolytica*, or that the translated N-terminal part of the Dna2 protein is sufficient to carry out its essential function.

*TOP1* and *SGS1* are specifically duplicated in *C. glabrata* (table 1). The two copies of *C. glabrata* Top1p (CAGL0E02431g and CAGL0J11660g, supplementary table) are almost perfectly aligned with ScTop1p, except in the N-terminal part of the protein. Synteny shows that CAGL0E02431g is the correct orthologue, the other copy

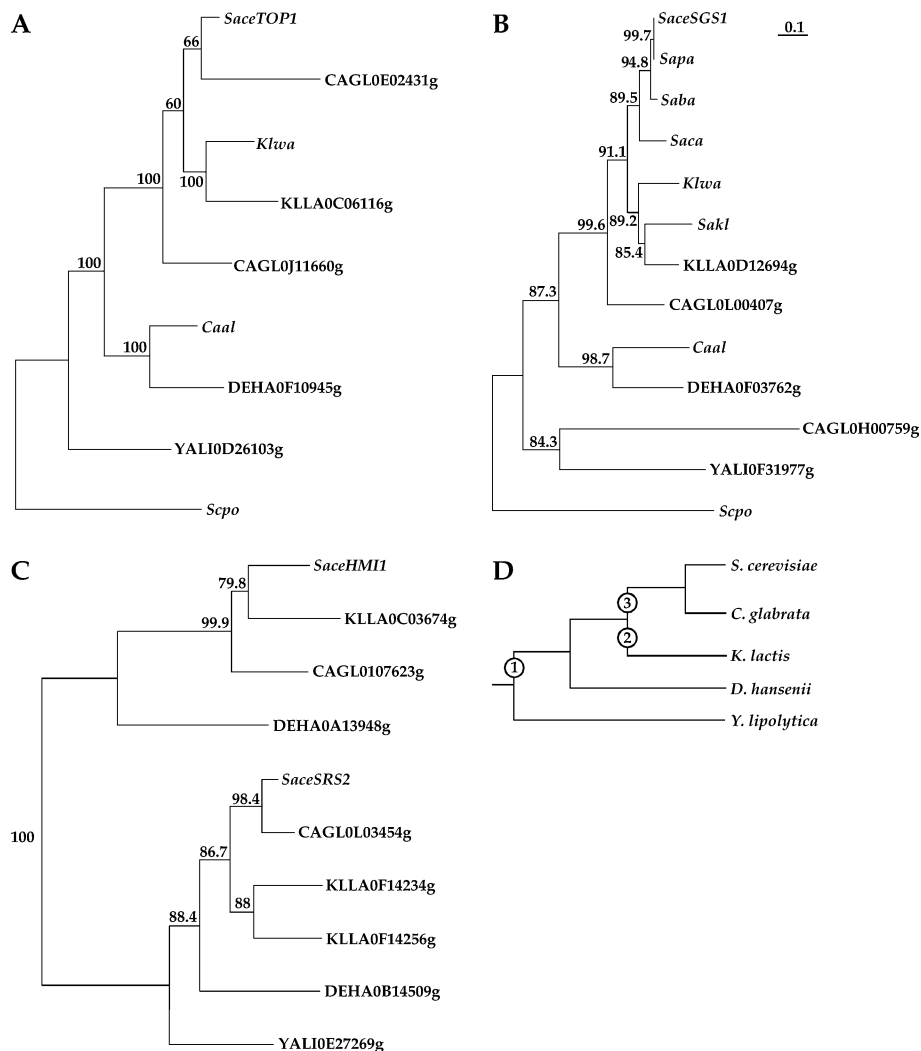


FIG. 1.—Phylogenetic tree of *TOP1*, *SGS1*, and *SRS2* families. Trees were obtained by the maximum likelihood method. Branch lengths are estimated under the model of amino acids substitution JTT (Jones, Taylor, and Thornton 1992). Percentages of bootstrap values for internal nodes are indicated on each branch (see *Materials and Methods* for details). When the number of homologous sequences among the *Génolevures* species was less than 10, additional homologous sequences were retrieved among the other available fungal genomes. (A) phylogeny of *TOP1*. (B) Phylogeny of *SGS1*. (C) Phylogeny of *SRS2*, which is a subset of a larger family of 10 members, also containing *HMI1*, a mitochondrial helicase and its three orthologues. (D) phylogenetic tree of the five yeast species studied here (Dujon et al. 2004). Numbers refer to the branch in which gene duplications occurred. 1: *HMI1/SRS2* duplication. 2: *SRS2* tandem duplication. 3: *TOP1* and *SGS1* duplications. *Sace*: *S. cerevisiae*, *Sapa*: *S. paradoxus*, *Saba*: *S. bayanus*, *Saca*: *S. castellii*, *Sakl*: *S. kluyveri*, *Saku*: *S. kudriavzevii*, *Cagl*: *C. glabrata*, *Klwa*: *K. waltii*, *Klla*: *K. lactis*, *Caal*: *C. albicans*, *Deha*: *D. hansenii*, *Yali*: *Y. lipolytica*, *Scpo*: *Schizosaccharomyces pombe*.

being present in a duplicated chromosomal block present in both *S. cerevisiae* and *C. glabrata* (G. Fischer and B. Dujon, unpublished data). The duplicated copy was conserved in *C. glabrata* but not in *S. cerevisiae* in which no trace of a pseudogene or a relic could be found in the duplicated block (I. Lafontaine and B. Dujon, unpublished data). Consistent with that, the phylogenetic tree shows that CAGL0E02431g is the closest homologue of ScTop1 (fig. 1A). Calculation of synonymous (Ds) and nonsynonymous (Dn) substitutions show that Ds values are very high (Ds > 5). Because synonymous sites are saturated, Dn/Ds ratios are not a reliable measure of evolutionary rates. Hence, we took in consideration only Dn values. They are low and similar for both paralogues (Dn<sub>CAGL0E02431g</sub> = 0.22; Dn<sub>CAGL0J11660g</sub> = 0.18, as compared to *S. cerevisiae*). This suggests that both genes have evolved at a similar rate

and have both probably retained their catalytic activity. Two copies of Sgs1p were found in *C. glabrata* (CAGL0L00407g and CAGL0H00759g). According to synteny results, CAGL0L00407g is the correct orthologue, and CAGL0H00759g is found in a duplicated block in *S. cerevisiae* and *C. glabrata* only (G. Fischer and B. Dujon, unpublished data). Like previously, the copy in the duplicated block has been erased, and no trace of a pseudogene or relic can be detected in *S. cerevisiae*. (I. Lafontaine and B. Dujon, unpublished data) The phylogenetic tree shows that the closest homologue of ScSgs1p is CAGL0L00407g, the other copy being more diverged (fig. 1B). Interestingly, the duplicated copy is shorter than the orthologue. It is deleted for the N-terminal part containing the Top3-binding domain and the C-terminal part containing the DNA-binding domain of the Sgs1 protein (fig. 2). Again, synonymous sites are

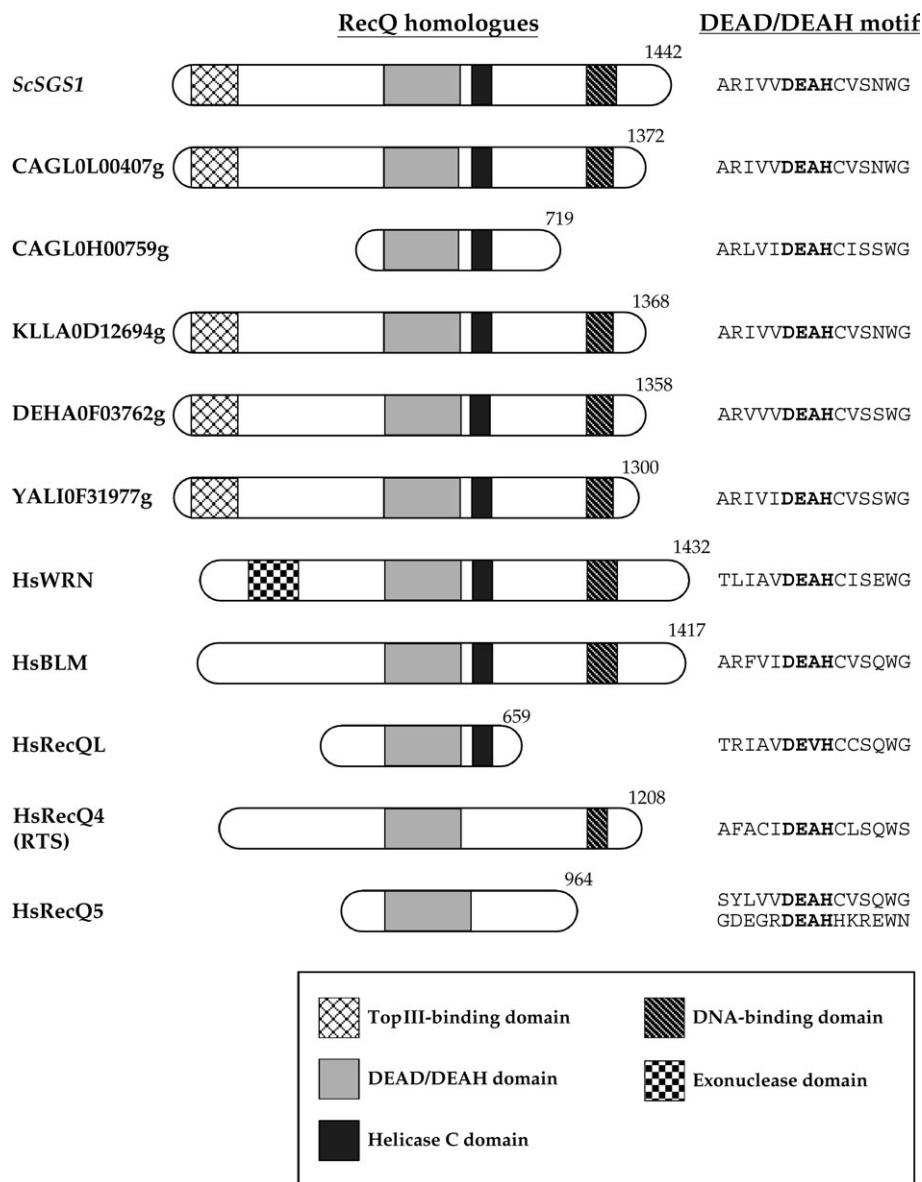


FIG. 2.—Functional domains of yeast Sgs1 proteins and human RecQ homologues. Domains were defined according to the CDD, except the TopIII-binding domain defined as in Mullen, Kaliraman, and Brill (2000). For each protein the number of amino acids (according to the NCBI genome annotation, in the case of the five human orthologues) is indicated above the C-terminus. Right: amino acids surrounding the DEAH motif (bold) are shown. In RecQ5, two DEAH motifs are present in the protein, separated by 340 amino acids.

saturated, but Dn values are low. However, the Dn value of the duplicated copy is higher than that of the orthologue ( $Dn_{CAGL0L00407g} = 0.2$ ;  $Dn_{CAGL0H00759g} = 0.46$ , as compared to *S. cerevisiae*), meaning that not only the copy lost two important parts of the protein (still retaining the helicase motif) but also the remaining part diverged more rapidly.

#### Genes Involved in DNA Repair

NER is the main mechanism used to remove pyrimidine dimers induced by UV cross-linking or chemical damage such as those caused by benzopyrene, aflatoxin, and cisplatin (Lindahl and Wood 1999). In humans, NER-defective individuals are affected by xeroderma pigmentosum (XP) a disorder associated with hypersensitivity to

sunlight and a 1,000-fold increase in the occurrence of skin cancer as compared to normal individuals. The *S. cerevisiae* homologues of XP genes are conserved in other yeast species (table 1). The transcription-coupled repair pathway is not completely conserved in *K. lactis* and *Y. lipolytica* because they both lack *RAD28*, the yeast homologue of human CSA involved in Cockayne syndrome. Genes involved in the mutational repair pathway (*RAD18*, *RAD6*) and its dedicated error-prone DNA polymerase *REV3* (Pol  $\zeta$ ) are well conserved in all species, as well as the postreplicational repair helicase *RAD5*. Interestingly, the *RAD6* gene, which is the most highly conserved of all 106 genes among the five species (fig. 3A), has two introns in *D. hansenii* and none in the four other yeast species. *POL4*, encoding the orthologue of Pol  $\lambda$ , a

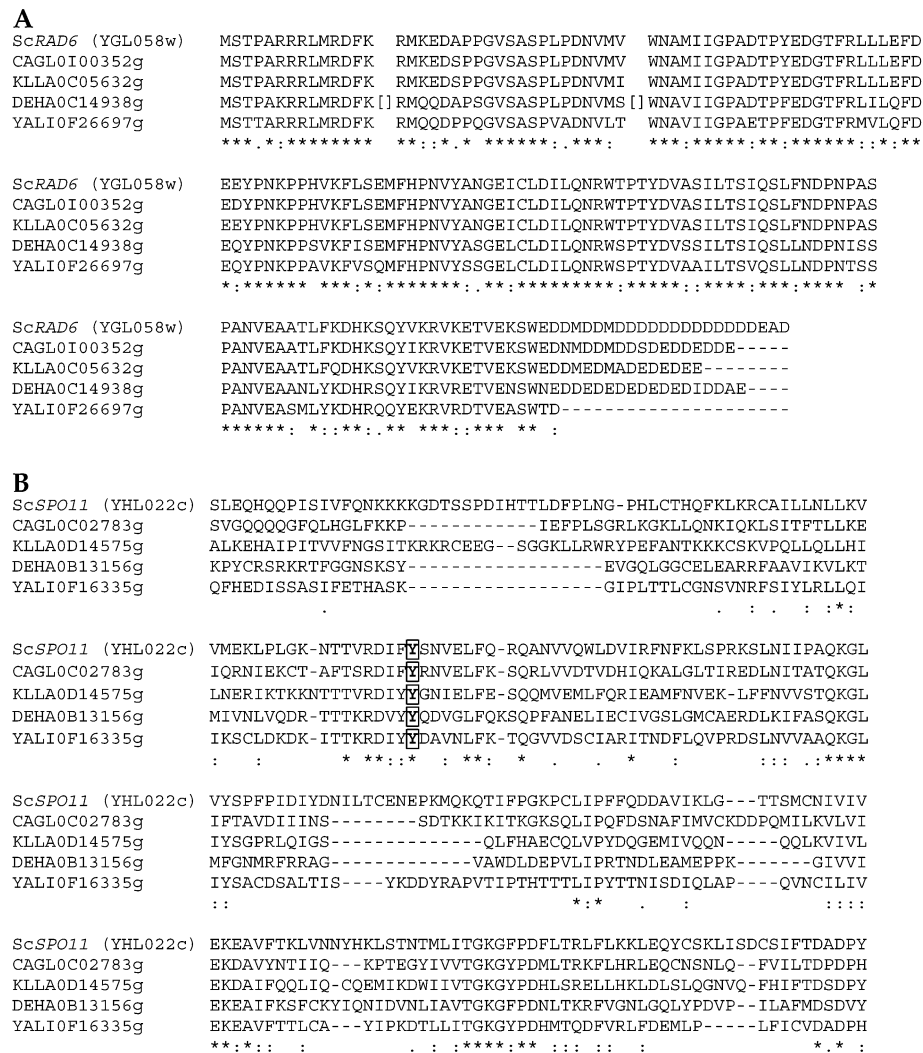


FIG. 3.—(A) Multiple alignment of the *Saccharomyces cerevisiae* RAD6 gene product with its four orthologues using the ClustalW software (Thompson, Higgins, and Gibson 1994). Identical amino acids are shown by a star, amino acids belonging to a ‘strong’ group are indicated by a column, and amino acids belonging to a ‘weaker’ group are shown by a single dot. The brackets show the location of the two introns in the *Debaryomyces hansenii* orthologue. Note the missing acidic tail in *Yarrowia lipolytica*. (B) Multiple alignment of the *S. cerevisiae* SPO11 gene product with its four orthologues using ClustalW. The conserved catalytic Tyr135 residue is boxed. Only the central well-conserved part of the protein is shown.

$\beta$ -like DNA polymerase involved in base excision repair is not detected in *D. hansenii*. Finally, genes involved in the mismatch repair (MMR) pathway are completely conserved in *C. glabrata* and *K. lactis*. For the two more distant species, only the core MutS and MutL homologues, *MLH1*, *MSH2*, *MSH3*, *MSH6*, and *PMS1* (Kolodner 1996), are found (table 1). Interestingly, *MSH1* whose role is essential for maintenance of mitochondrial DNA is conserved in all four species.

#### Double-Strand Break Repair Genes

Genes in this category have been subdivided in two subclasses (table 1), those involved in NHEJ and those involved in homologous recombination (HR). DNA end-joining is a conserved process through evolution, from yeast to man. In human B and T lymphocytes, V(D)J recombination of immunoglobulin chains and of T-cell receptors is achieved by NHEJ, involving a number of

genes including RAG1 and RAG2 (Grawunder, West, and Lieber 1998). In *S. cerevisiae*, there is no Rag protein but the end-joining machinery is very well conserved, and NHEJ has been mostly studied in this organism using HO and I-SceI-induced DSBs (for review see Haber 1995). It was recently shown that a V(D)J recombination substrate was correctly and precisely processed in yeast when the human RAG1 and RAG2 genes are coexpressed, showing that the whole yeast end-joining machinery is proficient to form signal joints (Clatworthy et al. 2003). In hemiascomycetes, the Ku complex is conserved, along with the Ligase IV orthologue (table 1). In the MRX complex, *MRE11* and *RAD50* are found in all species, whereas *XRS2* is only detected in *C. glabrata* and *K. lactis*. *XRS2* is the less well-conserved gene of the MRX complex, having no structural but a functional homologue in humans, called NBS1 (Carney et al. 1998). Tblastn search on *D. hansenii* and *Y. lipolytica*, using NBS1 as the query sequence did not reveal any homologue either. Surprisingly, the Ligase

IV-interacting factor, *LIF1*, is only found in *C. glabrata*, whereas *LIF2/NEJ1* is only found in *K. lactis*. *LIF2/NEJ1* is a haploid-specific gene that regulates the efficiency of NHEJ in yeast cells, depending on whether they express only one set of mating-type proteins ( $\alpha$  or  $\alpha$  proteins) or both sets ( $\alpha$  and  $\alpha$  proteins) (Frank-Vaillant and Marcand 2001; Valencia et al. 2001). Because Lif2p was found to specifically interact with Lif1p in a two-hybrid screen (Frank-Vaillant and Marcand 2001), it is therefore surprising that *K. lactis* does not have a recognizable *LIF1* gene. The Sir1 protein, involved in silencing *HML $\alpha$*  and *HMRA* loci in *S. cerevisiae*, was not detected in *C. glabrata* (Fabre et al. 2005). This species has lost its ability to mate, supposedly because both  $\alpha$  and silent mating-type cassettes are now expressed. Therefore, it is possible that selection pressure to keep the haploid regulator gene *LIF2* in this species was abolished, eventually leading to the loss of this gene. Among genes involved in HR, the *RAD52* gene product essential to this process is found in all four species along with the RecA homologue, *RAD51*, whose product catalyzes strand invasion and strand exchange (for review see Pâques and Haber 1999). Rad54p, Rdh54p, and Exo1p are also well conserved, whereas other accessory proteins that facilitate HR, like Rad55p and Rad59p, are not found in the most distant species. Finally, two copies of Srs2p were found to be duplicated in tandem in *K. lactis* (KLLA0F14256g and KLLA0F14234g, supplementary table). Multiple alignments show that both copies are very well conserved in their N-terminal part, in which the UvrD helicase domain is found (data not shown). Phylogeny demonstrates that both copies are at the same distance from ScSrs2p (fig. 1C) and that Dn values are almost identical ( $Dn_{KLLA0F14256g} = 0.5$ ;  $Dn_{KLLA0F14234g} = 0.42$ , as compared to *S. cerevisiae*). This suggests that either both copies have evolved at the same rate, that the duplication is fairly recent in the evolution of this yeast, or that there is a high level of gene conversion between tandemly duplicated genes. Tandem paralogues are often more conserved than dispersed paralogues; this is a general trend of tandem duplications (Dujon et al. 2004).

#### Weak Conservation of Genes Involved in Meiotic Recombination

HR during meiosis is a highly regulated process by which genetic information is reshuffled between homologous chromosomes (for review see Zickler and Kleckner 1998). During this process, DSBs are generated by the Spo11p topoisomerase and then processed by the meiotic recombination machinery involving the Mre11p-Rad50p-Xrs2p complex. Spo11p is homologous to the A subunit of type VI topoisomerases, such as those found in archaeobacteria (Bergerat et al. 1997). Homologues to *SPO11* are found in all four yeast species, despite extensive sequence divergence. The Tyr135 residue essential for its catalytic activity is conserved, strongly suggesting that the four orthologues are functional in vivo (fig. 3B). The occurrence of crossovers is also regulated during meiosis, although little is known about the precise mechanism by which a recombination intermediate is resolved as a crossover or as a noncrossover, in vivo. It involves—at least—two

different pathways: the Msh4-Msh5 pathway and the Mus81-Mms4 pathway. Msh4 and Msh5 proteins function as heterodimers in *S. cerevisiae*, and the corresponding mutants show a reduced frequency of meiotic crossovers as compared to wild-type strains (Pochart, Woltering, and Hollingsworth 1997). The *msh5* mutant is profoundly affected at an early stage during meiotic recombination, showing a decreased level of early recombination intermediates leading to crossovers (Börner, Kleckner, and Hunter 2004). Their simultaneous absence in *D. hansenii* and *Y. lipolytica* might reflect a different mechanism to control crossovers in these species. The *S. cerevisiae* Mus81-Mms4 complex is able to process branched structures arising during mitotic or meiotic replication/recombination that are not canonical Holliday junctions (Fricke, Bastin-Shanower, and Brill 2005). Mus81p is conserved in all species, whereas Mms4p was not found in *Y. lipolytica*. However, although Mus81p is known to be conserved throughout evolution, its partner is poorly conserved (Ögrünç and Sancar 2003). It is therefore possible that a functional homologue of Mms4p is also present in *Y. lipolytica* but not detected.

The other genes involved in the meiotic recombination pathway are most of the time poorly conserved in *D. hansenii* and absent in *Y. lipolytica*, with the exception of *MRE2*, whose product is involved in the splicing of *MER2* and *MER3* messenger RNAs in *S. cerevisiae*. *MER2* is predicted to contain an intron only in *C. glabrata*. Mre2p belongs to the U1 snRNP in *S. cerevisiae* and therefore splices many transcripts other than those involved in meiosis. It is thus probable that the *MRE2* gene does not play a role anymore in meiotic recombination in *K. lactis* or *D. hansenii*. In conclusion, the only genes that are found in all five yeast species are genes involved in initiating recombination by making and processing DSBs (*SPO11*, *MRE11*, *RAD50*) in resolving recombination intermediates (*MUS81*) or the general splicing factor *MRE2*.

#### Checkpoint Proteins

Signaling DNA damage during the cell cycle is regulated by a series of proteins that activate the so-called “checkpoints” (for review see Zhou and Elledge 2000). Most of them are conserved except in *Y. lipolytica*. *RAD9* is the only gene that is missing in *D. hansenii* in addition to *Y. lipolytica*. Most probably, *DDC2*, which has a human functional homologue (ATRIP, table 1), is also conserved in *Y. lipolytica* but is too diverged to be recognized.

#### Conservation of DNA Maintenance Pathways During Evolution

Given that some of the pathways are very well conserved in the five hemiascomycetous yeasts (e.g., replication or NER proteins) and others are missing several components, we wanted to know if amino acid conservation was the same among the different pathways. We performed pairwise Smith-Waterman alignments between each *S. cerevisiae* protein and its putative orthologues. Percentages of identity are shown in figure 4 for each



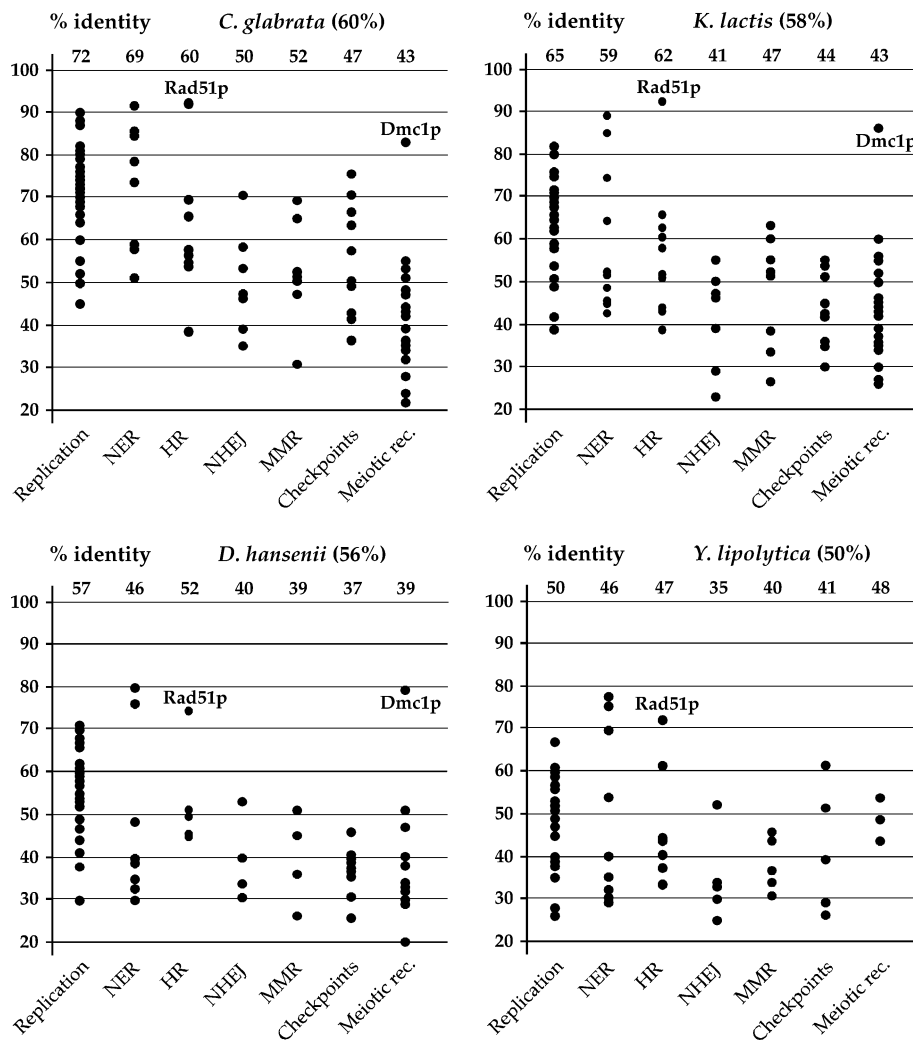


FIG. 4.—Values of Smith-Waterman identity scores for conserved proteins in each pathway for each species. Each dot corresponds to a pairwise alignment between a protein and its *Saccharomyces cerevisiae* orthologue. Rad51p and Dmc1p are indicated just below or above the corresponding dot (see text). Average identity of each pathway is indicated above the 100% line. Average identity of all orthologous proteins for each species is indicated in parentheses following the species name.

species in each pathway. The average identity for each pathway was also calculated. Note that proteins have been classified in a pathway according to one of their functions, although some of them act in several distinct pathways. The best example is the *MRE11-RAD50-XRS2* complex, classified in the NHEJ pathway, but which is known to be involved in formation and processing of meiotic DSBs, S-phase checkpoint activation, and HR (for review see Haber 1998). Nevertheless, when amino acid conservations of each pathway are compared, they generally follow the phylogenetic tree, i.e., *C. glabrata* is the closest to *S. cerevisiae* and *Y. lipolytica* is the farthest (fig. 4). The only exception is the meiotic recombination pathway, in which the only three genes to be conserved in *Y. lipolytica* (*MRE2*, *MUS81*, and *SPO11*) show a higher identity to *S. cerevisiae* orthologues than the corresponding *D. hansenii*, *K. lactis*, and *C. glabrata* genes. In order to determine if evolutionary rates were similar in the five species for these three genes, we calculated the Dn and Ds rates of nonsynonymous and synonymous substitutions. Because informative sites are saturated ( $D_s > 5$ ), we took

into consideration only Dn values. Using this criterion, we confirmed that *SPO11* and *MRE2* (but not *MUS81*) evolved slower in *Y. lipolytica* ( $D_{nSPO11} = 0.76$ ;  $D_{nMRE2} = 0.69$ , as compared to *S. cerevisiae*) than in *D. hansenii* ( $D_{nSPO11} = 0.98$ ;  $D_{nMRE2} = 0.73$ , as compared to *S. cerevisiae*). As a control, we also determined the average level of amino acid conservation between all *S. cerevisiae* proteins and their orthologues in each of the four species and used it as a baseline (see *Materials and Methods*). As expected, conservation follows the phylogenetic tree, i.e., *C. glabrata* proteins share a higher percentage of identity with *S. cerevisiae* proteins (60%) than *Y. lipolytica* proteins (50%). Hence, the only pathway in which proteins reach the amino acid identity baseline in each species is the replication pathway; most of the others (and all of them in *Y. lipolytica*) are below the baseline. This suggests that proteins involved in pathways whose average amino acid conservation is under the baseline diverge more rapidly than the average orthologous proteome, perhaps reflecting more flexibility in proteins involved in repair and recombination than in proteins involved in replication.

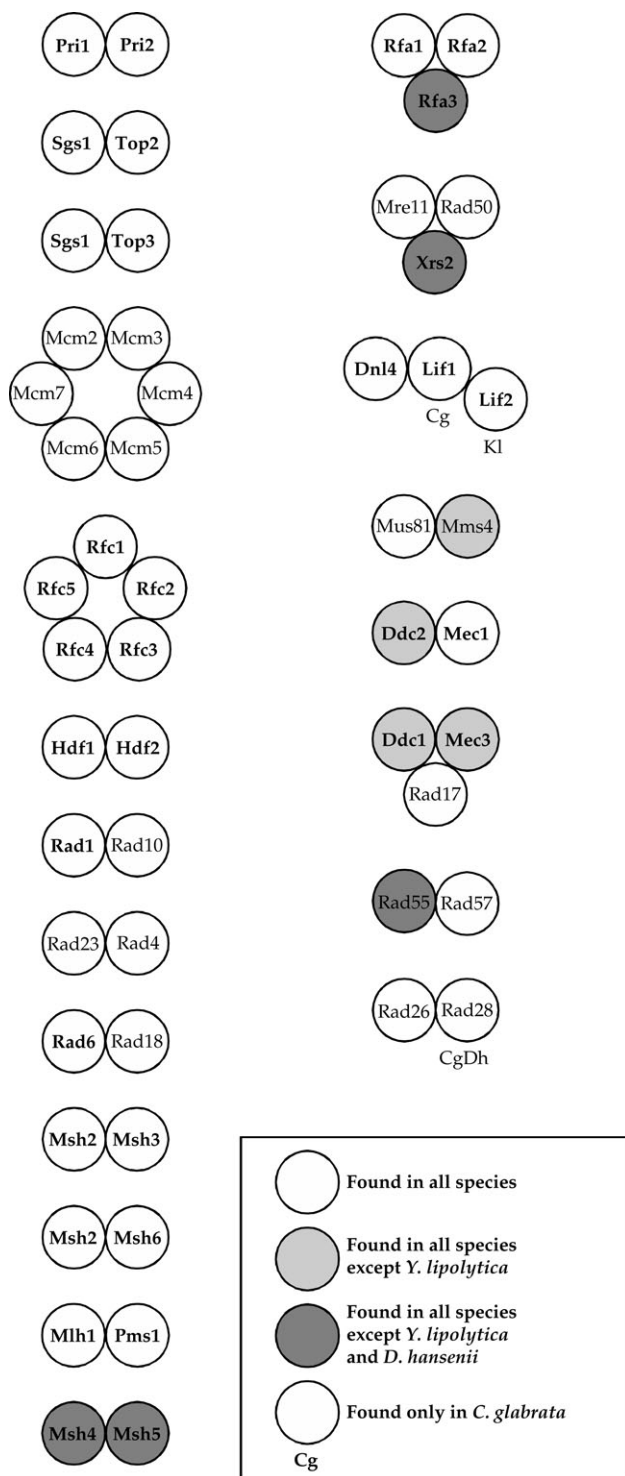


FIG. 5.—Conservation of known protein complexes in hemiascomycetous yeasts. Left: protein complexes for which each species contains at least one member. Right: complexes in which at least one member is present in at least a species and absent in at least another.

### Protein Complexes

Among the 106 genes we have studied, many were known to encode products belonging to multiprotein complexes. One might expect that selection pressure would be the same for all members of a protein complex because if

one of the members accumulates mutations faster than the other members, interaction between the different members could be rapidly lost and complex functionality disrupted. Therefore, one expects that in some cases all the members of a given complex are absent (they all evolved faster and are hence not recognizable anymore, Snel and Huynen 2004). Out of 21 known complexes, 12 are found in all organisms and 1 (*Msh4p-Msh5p*) is found only in *C. glabrata* and *K. lactis*, suggesting that either *MSH4* and *MSH5* genes evolved faster and are not detected anymore in the two more distant yeast species or that they appeared in the common ancestor of *S. cerevisiae* and *K. lactis*. The last eight complexes contain one or two members that are not conserved in each species (fig. 5).

### Discussion

In the present study, we have analyzed the content of four fully sequenced hemiascomycetous yeast genomes to find orthologues of 106 *S. cerevisiae* genes involved in replication, repair, and recombination pathways. The aim of this work was (1) to identify orthologous pathways in other yeast species and to investigate the conservation of these pathways; (2) get insights into the evolution of genes involved in such pathways, particularly the frequency with which gene duplication/loss occurred; and (3) try to draw conclusions about the biological properties of these hemiascomycetous yeast species based on their gene content.

### Conservation of Pathways

Pathways have been defined arbitrarily because many proteins belong to several pathways and therefore all pathways are interconnected with each other. However, despite such interconnection, some pathways such as meiotic recombination and checkpoints are less conserved than others, such as replication and NER. There are two independent criteria that may be used to estimate the conservation of pathways. The first—the presence/absence criterion—is used to determine the ratio of genes that are found in each species over the total number of genes in this pathway (table 1). The second—the conservation criterion—is used to calculate the average conservation in amino acid of proteins belonging to a given pathway for each species (fig. 4). Not surprisingly, the replication machinery comes first using both criteria, and almost all genes are present in each species and exhibit a high level of similarity with *S. cerevisiae* genes (table 1 and fig. 4). The NER pathway is very well conserved (all the genes are found in each species), but amino acid conservation is lower in *K. lactis*, *D. hansenii*, and *Y. lipolytica* than for proteins belonging to the HR pathway, in which many accessory proteins are not found in the more distant species (table 1 and fig. 4). In terms of presence/absence, the meiotic recombination machinery is missing several members, even in species related to *S. cerevisiae*. Most of the genes that are not found in *C. glabrata* and *K. lactis* belong to this pathway (table 1). We found that, in general, proteins interacting with DNA are more conserved than structural proteins, proteins that are part of a scaffold and other cofactors. It is striking that Rad51p and Dmc1p catalyzing strand exchange reactions are the most conserved of their respective pathways. Simi-

larly, the Mre11p-Rad50p complex and Spo11p, necessary to make and process meiotic DSBs, are conserved along with Mus81p, involved in resolving recombination intermediates. All these proteins interact directly with DNA and are much more conserved than proteins involved in making the synaptonemal complex or other structural proteins and cofactors.

### Gene Duplications During Evolution

Paralogous sets of genes play a key role in defining functional biological systems. For example, the MutS family of proteins contains six members in *S. cerevisiae* (*MSH1-6*, table 1), with distinct functions and specializations. Another example is the replicative helicase, formed by assembly of six distinct subunits, encoded by six different genes (*MCM2-7*), arising from successive gene duplications during evolution. In the present work, we found that both *SGS1* and *TOP1* were duplicated in *C. glabrata* and that *SRS2* is tandemly duplicated in *K. lactis*. *SGS1* encodes a DEAD/DEAH helicase of the RecQ/BLM/WRN family and has been shown to interact genetically with Top3p (Gangloff et al. 1994) and Top1p (Tong et al. 2001) and physically with Top2p (Watt et al. 1995). The duplicated Sgs1p and the duplicated Top1p both arose from duplication events prior to the *S. cerevisiae*-*C. glabrata* speciation, and both duplicated genes have been conserved in *C. glabrata* and lost in *S. cerevisiae* (fig. 1D). Given that Dn values are rather low for both genes, it is probable that both duplicated proteins are under selection pressure in *C. glabrata*. This could imply being part of an alternative complex involved in replication and/or repair or being part of a Sgs1-containing complex that would be specific to the life cycle of this pathogenic yeast. Interestingly, the duplicated copy of Sgs1p lacks its N-terminal and C-terminal parts (fig. 2) but retains the central helicase domain. It is therefore possible that it lost its DNA-binding activity but is still active as a helicase, maybe as part of a multicomponent complex. In humans, there are five homologues of Sgs1p, and two of them (RecQ5 and RecQL, fig. 2) are shorter versions, lacking either the N-terminal part (RecQL) or both the N- and C-terminal parts (RecQ5) but retaining their helicase domain. It is interesting that in *C. glabrata*, a short copy of Sgs1p was also found. We performed local and global alignments between the Sgs1p copy and the five human orthologues and concluded that although being a shortened version of Sgs1p, the *C. glabrata* copy is closer to WRN, BLM, and RecQ4 (RTS) than to the human RecQ5 and RecQL helicases. We therefore concluded that evolution of this protein family in *C. glabrata* and man was different. The Top1p duplication is interesting because this gene is duplicated in vertebrates but not in *S. cerevisiae*, *Schizosaccharomyces pombe*, or plants (Zhang et al. 2004). In vertebrates, one gene product is addressed to the nucleus and the other to mitochondria (Zhang et al. 2001). In *C. glabrata*, the duplicated copy (CAGLOJ11660g) is predicted to encode a nuclear product, but no obvious nuclear nor mitochondrial addressing signal could be found in the original gene (CAGLOE02431g; Y. Pommier, personal communication). How-

ever, we know that the *S. cerevisiae* orthologue functions in the nucleus. Hence, this suggests that both gene products in *C. glabrata* are nuclear, and therefore that the evolution of this protein family in *C. glabrata* and in vertebrates was also different. It was previously shown that the very conserved lysine residue (K41) in Srs2p was essential for the adenosine triphosphatase (ATPase) activity (Krejci et al. 2004). This residue is present among the five species in the center of the completely conserved motif  $^{35}\text{G}_{36}\text{P}_{37}\text{G}_{38}\text{T}_{39}\text{G}_{40}\text{K}_{41}\text{T}_{42}\text{K}_{43}$ . In addition, in the duplicated copy of *SRS2* in *K. lactis*, this motif is also completely conserved, suggesting that all the Srs2p orthologues are functional in the four other hemiascomycetes. The Hmi1p helicase is a paralogue of Srs2p in *S. cerevisiae*. Hmi1p is a mitochondrial protein and is essential for maintenance of mitochondrial DNA (Sedman et al. 2000). It was found in all species except *Y. lipolytica*, and the phylogenetic tree shows that the duplication of the *SRS2/HMI1* gene ancestor occurred in the common ancestor to *S. cerevisiae* and *D. hansenii* (fig. 1C and D). Consistent with this observation, the conserved ATPase motif in Hmi1p only differs by one amino acid (Thr<sub>39</sub> → Ser<sub>39</sub>) from Srs2p. This conservative mutation is found in all four species in which a *HMI1* orthologue is detected, strengthening the idea that the formation of paralogues occurred before speciation of our yeasts. Finally, it was shown that the C-terminal part of the Hmi1 protein contains the mitochondrial targeting signal (Lee et al. 1999). Alignments of Srs2p and Hmi1p orthologues show that both proteins are very well conserved in the N-terminal part, containing the ATPase motif, but conservation of the C-terminal part is weak. Therefore, Srs2p/Hmi1p is probably a case of gene duplication before speciation, leading after alteration of the C-terminal part of one of the duplicated copies to a specialization of function, with both proteins being DNA helicases but one addressed to the mitochondria and the other to the nucleus. Our results strongly suggest that in each case of gene duplication, both copies are probably functional and have retained their catalytic activity, although they might be active in different cell compartments and/or on different substrates (subfunctionalization) (Lynch and Conery 2000).

### Pathway Conservation, Evolution, and Yeast Biological Properties

We showed that genes belonging to the meiotic recombination machinery are poorly conserved in hemiascomycete species (in terms of presence/absence). However, *K. lactis*, *D. hansenii*, and *Y. lipolytica* undergo meiosis (Herman and Roman 1966; Kreger-van Rij and Veenhuis 1975; Casarégola et al. 2000). This means that although most of the genes necessary to go through meiotic recombination in *S. cerevisiae* are not detected in other yeasts, they must have functional orthologues able to carry out similar functions. Interestingly, the most highly conserved protein of the HR pathway is Rad51p in each species, and the most highly conserved protein of the meiotic recombination pathway is Dmc1p in the three species in which it is detected (fig. 4). Because *DMC1* and *RAD51* presumably come from the duplication of a common ancestor, our results suggest that this duplication occurred after

the divergence between *Y. lipolytica* and the four other yeast species. It has also been proposed that organisms undergoing meiosis can be classified in two different groups (Stahl et al. 2004). In group I, organisms do not depend on meiotic DSB-repair functions to achieve synapsis (*Drosophila melanogaster*, *Caenorhabditis elegans*, *Neurospora crassa*), whereas in group II organisms, synapsis may only occur if DSB-repair is functional (e.g., *S. cerevisiae*). In group II organisms, the *DMC1*, *HOP2*, and *MND1* genes are found, whereas they are apparently absent in group I organisms. This would suggest that group I organisms have lost these three genes or that they have been independently acquired during evolution. Therefore, *Y. lipolytica* would be classified as a group I organism because none of these three genes is found, whereas the four other species all contain these three genes (table 1 and data not shown for *MND1*). In addition, the Msh4-Msh5 protein complex involved in crossover control is also missing in this yeast. Taken together, these data suggest that although *Y. lipolytica* undergoes meiotic recombination (Wickerham, Kurtzman, and Herman 1970; Gaillardin, Charoy, and Heslot 1973), its properties are most probably very different from the four other hemiascomycetous yeasts.

In order to determine if the differences observed among the different yeast species for the NHEJ pathway could reflect a difference in the efficiency of DSB-repair mechanisms, we irradiated haploid cells with a source of  $\gamma$ -radiation.  $\gamma$ -Rays are known to induce single- and double-strand breaks in chromosomes, and resistance to ionizing radiations is a measure of how efficient the DSB-repair systems are in a given organism (Esposito and Wagstaff 1981). At low energy (50 Gys), the four hemiascomycetes are slightly more resistant than a haploid *S. cerevisiae* strain to ionizing radiations (supplementary figure). At higher doses (300 Gys), all five yeast species show the same sensitivity to  $\gamma$ -rays. We concluded that, most probably, no gene dramatically affecting the efficiency of DSB-repair was missing in the four species. This suggests again that the NHEJ pathway is functional, despite the apparent absence of some of its members. The higher resistance at low doses may be hypothesized by the existence of a more efficient pathway, for example HR with the sister chromatid, that would occur more often in those species as compared to *S. cerevisiae*, perhaps because of a longer S-G2 phase of the cell cycle. Further experimentation will be required to determine if cell cycles are the same in these five hemiascomycetes.

### Supplementary Materials

One supplementary table: list of homologues in each yeast species.

One supplementary figure: comparison of survival to  $\gamma$ -irradiation between the five hemiascomycetes.

SUPPLEMENTARY FIGURE. Comparison of survival to  $\gamma$ -irradiation between the five hemiascomycetes. For each haploid strain, approximately 300 cells were plated on YPGlu plates and irradiated at different doses (0, 50, 100, and 300 Gy) using a  $^{137}\text{Cs}$  source, at a dose rate of 4 Gy/min. After 3 days of incubation at 30°C, survival

was determined as the number of colony forming units (CFU) at each dose divided by the number of CFU at 0 Gy. The average of two independent experiments is shown for each species.

### Acknowledgments

We thank our colleagues for fruitful discussions, particularly G. Fischer for careful reading of the manuscript, J. Haber for many suggestions, and A. Thierry for her expertise with in silico intron designing. This work was supported by the Consortium National de Recherche en Génomique (to Génoscope and to Institut Pasteur Génomique), the CNRS (GDR2354, Génolevures sequencing consortium), the Ministère de la Jeunesse, de l'Éducation et de la Recherche (ACI IMPBio no. IMPB114 "Génolevures en ligne"), and the "Conseil Régional d'Aquitaine" ("Génotypage et Génomique Comparée"). A.K. is the recipient of a doctoral fellowship from the "Ministère de l'Éducation Nationale, de l'Enseignement Supérieur et de la Recherche." B.D. is a member of the Institut Universitaire de France.

### Literature Cited

- Bennett, J. E., K. Izumikawa, and K. A. Marr. 2004. Mechanism of increased fluconazole resistance in *Candida glabrata* during prophylaxis. *Antimicrob. Agents Chemother.* **48**:1773–1777.
- Bergerat, A., B. de Massy, D. Gabelle, P. C. Varoutas, A. Nicolas, and P. Forterre. 1997. An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature* **386**:414–417.
- Bolotin-Fukuhara, M., C. Toffano-Nioche, F. Artiguenave et al. (11 co-authors). 2000. Genomic exploration of the hemiascomycetous yeasts: 11. *Kluyveromyces lactis*. *FEBS Lett.* **487**:66–70.
- Bon, E., S. Casaregola, G. Blandin et al. (11 co-authors). 2003. Molecular evolution of eukaryotic genomes: hemiascomycetous yeast spliceosomal introns. *Nucleic Acids Res.* **31**:1121–1135.
- Börner, G. V., N. Kleckner, and N. Hunter. 2004. Crossover/non-crossover differentiation, synaptonemal complex formation, and regulatory surveillance at the leptotene/zygotene transition of meiosis. *Cell* **117**:29–45.
- Burgers, P. M., E. V. Koonin, E. Bruford et al. (21 co-authors). 2001. Eukaryotic DNA polymerases: proposal for a revised nomenclature. *J. Biol. Chem.* **276**:43487–43490.
- Cann, I. K. O., and Y. Ishino. 1999. Archaeal DNA replication: identifying the pieces to solve a puzzle. *Genetics* **152**:1249–1267.
- Carney, J. P., R. S. Maser, H. Olivares, E. M. Davis, M. Le Beau, J. R. Yates III, L. Hays, W. F. Morgan, and J. H. J. Petrini. 1998. The hMre11/hRad50 protein complex and Nijmegen breakage syndrome: linkage of double-strand break repair to the cellular DNA damage response. *Cell* **93**:477–486.
- Casarégola, S., C. Neueglise, A. Lepingle, E. Bon, C. Feynerol, F. Artiguenave, P. Wincker, and C. Gaillardin. 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.
- Clatworthy, A. E., M. A. Valencia, H. J. H., and M. A. Oettinger. 2003. V(D)J recombination and RAG-mediated transposition in yeast. *Cell* **12**:489–499.
- Dujon, B., D. Sherman, G. Fischer et al. (67 co-authors). 2004. Genome evolution in yeasts. *Nature* **430**:35–44.

- Esposito, M. S., and J. E. Wagstaff. 1981. Mechanisms of mitotic recombination. Pp. 341–370 in J. N. Strathern, E. W. Jones, and J. R. Broach, eds. *The molecular biology of the yeast Saccharomyces—life cycle and inheritance*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Fabre, E., H. Muller, P. Therizols, I. Lafontaine, B. Dujon, and C. Fairhead. 2005. Comparative genomics in hemiascomycete yeasts: evolution of sex, silencing and subtelomeres. *Mol. Biol. Evol.* (in press).
- Frank-Vaillant, M., and S. Marcand. 2001. NHEJ regulation by mating type is exercised through a novel protein, Lif2p, essential to the Ligase IV pathway. *Genes Dev.* **15**:3005–3012.
- Fricke, W. M., S. A. Bastin-Shanower, and S. J. Brill. 2005. Substrate specificity of the *Saccharomyces cerevisiae* Mus81-Mms4 endonuclease. *DNA Repair* **4**:243–251.
- Gaillardin, C., V. Charoy, and H. Heslot. 1973. A study of copulation, sporulation and meiotic segregation in *Candida lipolytica*. *Arch. Mikrobiol.* **92**:69–83.
- Gangloff, S., J. P. McDonald, C. Bendixen, L. Arthur, and R. Rothstein. 1994. The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase. *Mol. Cell. Biol.* **14**:8391–8398.
- Goldman, N., and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**:725–736.
- Grawunder, U., R. B. West, and M. R. Lieber. 1998. Antigen receptor gene rearrangement. *Curr. Opin. Immunol.* **10**:172–180.
- Guindon, S., and O. Gascuel. 2003. A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**:696–704.
- Haber, J. E. 1995. *In vivo* biochemistry: physical monitoring of recombination induced by site-specific endonucleases. *BioEssays* **17**:609–620.
- . 1998. The many interfaces of Mre11. *Cell* **95**:583–586.
- Herman, A., and H. Roman. 1966. Allele specific determinants of homothallism in *Saccharomyces lactis*. *Genetics* **53**:727–740.
- Jones, D. T., W. R. Taylor, and Thornton J.M. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**:275–282.
- Kolodner, R. 1996. Biochemistry and genetics of eukaryotic mismatch repair. *Genes Dev.* **10**:1433–1442.
- Kreger-van Rij, N. J. W., and M. Veenhuis. 1975. Electron microscopy of ascus formation in the yeast *Debaryomyces hansenii*. *J. Gen. Microbiol.* **89**:256–264.
- Krejci, L., M. Macris, Y. Li, S. Van Komen, J. Villemain, T. Ellenberger, H. Klein, and P. Sung. 2004. Role of ATP hydrolysis in the anti-recombinase function of *Saccharomyces cerevisiae* Srs2 protein. *J. Biol. Chem.* **279**:23193–23199.
- Lee, C. M., J. Sedman, W. Neupert, and R. A. Stuart. 1999. The DNA helicase, Hmi1p, is transported into mitochondria by a C-terminal cleavable targeting signal. *J. Biol. Chem.* **274**:20937–20942.
- Lépingle, A., S. Casaregola, C. Neuveglise, E. Bon, H. Nguyen, F. Artiguenave, P. Wincker, and C. Gaillardin. 2000. Genomic exploration of the hemiascomycetous yeasts: 14. *Debaryomyces hansenii* var. *hansenii*. *FEBS Lett.* **487**:82–86.
- Lindahl, T., and R. D. Wood. 1999. Quality control by DNA repair. *Science* **286**:1897–1905.
- Lynch, M., and J. S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290**:1151–1155.
- Mullen, J. R., V. Kaliraman, and S. J. Brill. 2000. Bipartite structure of the SGS1 DNA helicase in *Saccharomyces cerevisiae*. *Genetics* **154**:1101–1114.
- Notredame, C., D. Higgins, and J. Heringa. 2000. A novel method for multiple sequence alignments. *J. Mol. Biol.* **302**:205–217.
- Ögrünc, M., and A. Sancar. 2003. Identification and characterization of human MUS81-MMS4 structure-specific endonuclease. *J. Biol. Chem.* **278**:21715–21721.
- Pâques, F., and J. E. Haber. 1999. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **63**:349–404.
- Pochart, P., D. Woltering, and N. M. Hollingsworth. 1997. Conserved properties between functionally distinct MutS homologs in yeast. *J. Biol. Chem.* **272**:30345–30349.
- Sedman, T., S. Kuusk, S. Kivi, and J. Sedman. 2000. A DNA helicase required for maintenance of the functional mitochondrial genome in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **20**:1816–1824.
- Snel, B., and M. A. Huynen. 2004. Quantifying modularity in the evolution of biomolecular systems. *Genome Res.* **3**:391–397.
- Stahl, F. W., H. M. Foss, L. S. Young, R. H. Borts, M. F. F. Abdullah, and G. P. Copenhaver. 2004. Does crossover interference count in *Saccharomyces cerevisiae*? *Genetics* **168**:35–48.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
- Tong, A. H., M. Evangelista, A. B. Parsons et al. (13 co-authors). 2001. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**:2364–2368.
- Valencia, M., M. Bentele, M. B. Vaze, G. Herrmann, E. Kraus, S.-E. Lee, P. Schär, and J. E. Haber. 2001. *NEJ1* controls non-homologous end-joining in *Saccharomyces cerevisiae*. *Nature* **414**:666–669.
- Watt, P. M., E. J. Louis, R. H. Borts, and I. D. Hickson. 1995. Sgs1: a eukaryotic homolog of *E. coli* RecQ that interacts with topoisomerase II *in vivo* and is required for faithful chromosome segregation. *Cell* **81**:253–260.
- Wickerham, L. J., C. P. Kurtzman, and A. I. Herman. 1970. Sexual reproduction in *Candida lipolytica*. *Science* **167**:1141.
- Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**:555–556.
- Zhang, H., J. M. Barcelo, B. Lee, G. Kohlhagen, D. B. Zimonjic, N. C. Popescu, and Y. Pommier. 2001. Human mitochondrial topoisomerase I. *Proc. Natl. Acad. Sci. USA* **98**:10608–10613.
- Zhang, H., L. H. Meng, D. B. Zimonjic, N. C. Popescu, and Y. Pommier. 2004. Thirteen-exon-motif signature for vertebrate nuclear and mitochondrial type IB topoisomerases. *Nucleic Acids Res.* **32**:2087–2092.
- Zhou, B.-B. S., and S. J. Elledge. 2000. The DNA damage response: putting checkpoints in perspective. *Nature* **408**:433–439.
- Zickler, D., and N. Kleckner. 1998. The leptotene-zygotene transition of meiosis. Pp. 619–697 in A. Campbell, W. W. Anderson, and E. W. Jones, eds. *Palo Alto, CA: Annual Review of Genetics*.

Franz Lang, Associate Editor

Accepted January 4, 2005